



# Kent Academic Repository

Mbioh, Will (2026) *Social media addiction and the Digital Services Act: mitigation measures or missed opportunities? An intra-active critique.* European Law Open . ISSN 2752-6135.

## Downloaded from

<https://kar.kent.ac.uk/113676/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1017/elo.2026.10081>

## This document version

Publisher pdf

## DOI for this version

## Licence for this version

CC BY-NC (Attribution-NonCommercial)

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).


### Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



CORE ANALYSIS

# Social media addiction and the Digital Services Act: mitigation measures or missed opportunities? An intra-active critique

Will R. Mbioh 

Kent Law School, University of Kent, UK  
Email: [w.r.mbioh@kent.ac.uk](mailto:w.r.mbioh@kent.ac.uk)

(Received 16 June 2025; revised 3 November 2025; accepted 6 March 2026)

## Abstract

This paper examines how the European Union's Digital Services Act (DSA) understands platform addiction and, specifically, how it conceptualises its causative phenomenology: namely, the processes through which addiction is thought to arise. It begins by analysing the Act's architecture and contends that the DSA operates with a largely mechanical, linear view: addiction is treated as the outcome of particular interface features that trigger or drive harmful behaviour. The paper then sets out four dominant models for understanding platform addiction and, drawing on Karen Barad's concepts of intra-action and diffraction, argues that addiction is better approached as an emergent phenomenon produced through entangled human, social, and technical relations. On that basis, the paper's central claim is not simply that the DSA simplifies; it is that the way it simplifies does not stand outside the phenomenon it regulates. By fixing predicates, thresholds, and evidential tests, the Act makes agential cuts that shape what counts as harm, when intervention is permitted, and how risk is distributed. These cuts are not neutral. They have distributional effects and could be drawn differently (through alternative predicates or thresholds) to produce other evidential routes and earlier intervention points, with different consequences for how risks and harms are allocated and managed.

**Keywords:** Digital Services Act (DSA); social media addiction; systemic risk; intra-action

## 1. Introduction

This paper asks how the European Union's Digital Services Act (DSA, 2022) conceptualises 'platform addiction' and the causative phenomenology of that addiction. In other words, I am interested in how the Act understands what platform addiction is and how it comes about. My first claim is that the DSA adopts a largely mechanical, almost linear, model of causation: addiction is treated as something produced by a specific interface feature that 'causes' or 'stimulates' addictive behaviour in a way that culminates in serious (near-clinical) harm. However, this framing sits uneasily with the four dominant approaches to platform addiction, which do not reduce the matter to a single triggering feature. Rather, they present a complex picture that spans neurology, mental health, psychosocial conditions, behaviourist conditioning, and the expanding role of large-scale machine-learning systems.

Building on that broader account, and drawing on Karen Barad's concepts of intra-action and diffraction,<sup>1</sup> I argue that platform addiction is better understood as a *relatant*: an emergent phenomenon arising from the entanglement of these factors. Consequently, my core critique is not that the DSA simplifies (law must simplify to act), but that it simplifies in a particular way. Even simplifications are never neutral.<sup>2</sup> In Barad's terms, the Act makes *agential cuts*, boundary-drawing decisions within the regulatory apparatus that determine what becomes visible, measurable, and actionable:<sup>3</sup> by fixing predicates, thresholds, tests, and evidential standards for when addiction 'happens' and when intervention is permitted, it actively shapes the ontology of harm and the terms of causation. Those cuts carry ethical and distributional consequences in deciding which harms are tolerated, which are actionable, and whose risks are left in place, and, more significantly, whether the behavioural-modification logics and incentive structures of the attention economy (what Shoshana Zuboff identifies as the core machinery of surveillance capitalism)<sup>4</sup> are preserved intact.

Accordingly, I suggest that other cuts are available, without prescribing a single legislative formula. For example, moving away from a narrow causative vocabulary ('cause', 'stimulate') towards contribution and foreseeability could support different evidential standards and, in turn, open distinct moments and modes for regulatory action. In practical terms, this would shift the focus from isolating one proximate trigger to recognising cumulative, patterned effects across bundles of features and user cohorts.

To be clear, I do not claim that the DSA is the cause of addiction, nor that it could or should, on its own, solve it. Yet it does not stand outside the phenomenon. By setting thresholds and procedures, the Act can enable and stabilise elements of the attention economy and what Shoshana Zuboff calls the 'behavioural modification'<sup>5</sup> techniques of surveillance capitalism. Put simply, the argument is not only that the DSA simplifies the complexity of addiction, but that the way it does so produces particular distributions of risk and specific ontologies of harm. Those outcomes are not inevitable. Even within the Act's existing scope, alternative cuts (different thresholds, predicates, and standards) could justifiably be made, with different consequences for what becomes visible, measurable, and actionable.

To develop the argument, the paper proceeds in seven sections after the introduction. Sections 2 and 3 examine the DSA's legislative architecture, its ontology of harm, and its causal predicates and thresholds. Section 4 advances the critique of legal causality. Section 5 presents alternative explanatory models. Section 6 details recommender-system dynamics. Section 7 reframes addiction as what I term a Baradian *relatant*. Section 8 draws out the core critique of the DSA.

## 2. Beyond content regulation: how the DSA targets platform design and interface features to address addiction

The EU's Digital Services Act (DSA, 2022) stands out from comparable laws in the UK (ie, the Online Safety Act (OSA) 2023) in addressing online safety because it extends its focus beyond regulating content.<sup>6</sup> That is, as with these other laws, the Act covers the removal of illegal content (eg, hate speech, terrorism-related material)<sup>7</sup> and the protection of children from inappropriate or

<sup>1</sup>K Barad, *Meeting the Universe Halfway* (Duke University Press 2007).

<sup>2</sup>D Harris, *K Barad's Feminist Materialism: Intra-Action and Diffraction* (Cambridge Scholars Publishing 2021).

<sup>3</sup>Barad, *Meeting the Universe Halfway* (n 1).

<sup>4</sup>S Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019).

<sup>5</sup>Zuboff (n 4).

<sup>6</sup>W Mbioh, 'Beyond Echo Chambers and Rabbit Holes: Algorithmic Drifts and the Limits of the Online Safety Act, Digital Services Act, and AI Act' 33 (2024) Griffith Law Review 189.

<sup>7</sup>Digital Services Act, recital 22, Arts 3(t), 8–9, 14, 16.

harmful material<sup>8</sup> (eg, explicit content).<sup>9</sup> However, it also recognises that harm can arise independently of content legality, originating instead from the very design, architecture, and functionality of digital platforms. Unlike the UK's OSA (which regulates algorithms and features only insofar as they amplify harmful content),<sup>10</sup> the DSA identifies design-level risks as potential sources of what it terms 'systemic risks'.<sup>11</sup> These include, but are not limited to, deceptive or manipulative interface practices (ie, dark patterns),<sup>12</sup> recommender systems that operate without user choice,<sup>13</sup> discriminatory targeting in advertising,<sup>14</sup> and certain compulsive engagement mechanisms. It is within this latter category that the risk of platform or social media addiction is addressed, framed as a distinct systemic risk in both Recitals 81 and 83 and within the obligations imposed on very large online platforms (VLOPs), such as social media platforms, under Articles 34 and 35(a).

The DSA obliges such platforms to undertake risk assessments<sup>15</sup> and adopt 'reasonable, proportionate, and effective' mitigation measures<sup>16</sup> for systemic risks, including those linked to addictive platform features. These measures may relate to the modification of algorithmic systems (eg, recommender systems), interface features (eg, infinite scroll, autoplay), and notification systems, provided a demonstrable connection between the feature and addictive behaviour can be identified. However, while the DSA requires a 'systemic' approach to risk governance (by requiring platforms to evaluate and mitigate risks across design, operation, and moderation systems), what counts as systemic is not uniform across domains. Each identified systemic risk (eg, disinformation, gender-based violence, addictive behaviours, targeting of minors, discriminatory advertising) is associated with distinct criteria, thresholds, and regulatory expectations. For instance, the bar for intervention in cases of platform addiction is 'serious negative consequences to a person's physical or mental well-being',<sup>17</sup> whereas the threshold for intervention regarding risks to public health is a 'foreseeable negative effect'. Accordingly, each systemic risk should be read on its own terms, with careful attention to the legal thresholds, linguistic formulations, and evidentiary burdens that delimit the obligations attached to it. Moreover, the DSA's risk-based architecture implicitly permits certain forms of harm to persist, provided they do not exceed the threshold of legal seriousness, thereby embedding a logic of harm management rather than harm prevention.<sup>18</sup> This permissive posture is further suggested by the indeterminacy of 'systemic risk' itself, which scholars note is not fixed in meaning but will be gradually defined through negotiated practice between the Commission and covered platforms.<sup>19</sup>

<sup>8</sup>*Ibid.*, Art 28.

<sup>9</sup>R Griffin, 'The Politics of Algorithmic Censorship: Automated Moderation and Its Regulation' in J Garrat (ed), *Music and the Politics of Censorship: From the Fascist Era to the Digital Age* (Brepols 2025).

<sup>10</sup>Online Safety Act 2023, ss 9(5)(2), 10(3)–(4), 11(6)(e).

<sup>11</sup>M Husovec, *Principles of the Digital Services Act* (Oxford University Press 2024).

<sup>12</sup>Digital Services Act, Art 25.

<sup>13</sup>*Ibid.*, Art 27.

<sup>14</sup>*Ibid.*, Art 26(3).

<sup>15</sup>*Ibid.*, Art 34.

<sup>16</sup>*Ibid.*, Art 35(1).

<sup>17</sup>*Ibid.*, recital 83.

<sup>18</sup>G De Gregorio and P Dunn, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' 59 (2022) *Common Market Law Review* 473; A Palmieri, K Kollnig and A Tamò-Larrieux, 'Systemic Risks of Dominant Online Platforms: A Scoping Review' (2024) SSRN Working Paper <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4839060](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4839060)> accessed 9 March 2026.

<sup>19</sup>N Helberger, 'The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power' 8 (2020) *Digital Journalism* 842; N Helberger et al, 'Regulation of News Recommenders in the Digital Services Act: Empowering David against the Very Large Online Goliath' 26 (2021) *Internet Policy Review*.

### 3. Addiction as a systemic risk: regulatory thresholds, causal language, and design-focused governance

For example, the Act does not provide a definition of what ‘addiction’ is. However, it does specifically provide that the obligation to mitigate it as a risk is restricted to ‘interfaces’<sup>20</sup> that ‘cause addictive behaviour’<sup>21</sup> or ‘stimulate addictive behaviours’<sup>22</sup>. In other words, the Act recognises addiction as a systemic risk only insofar as it arises from design, use, or functioning of interface elements: ie, structural and interactive features embedded in the platform’s design that shape user behaviour and engagement patterns. These *may* include, as the Act does not provide a prescriptive or exhaustive list, infinite scrolling, autoplay, or loot boxes; features understood to cause or stimulate addictive behaviours. Here, to give both terms their ordinary meaning, to stimulate is to provoke, incite, or trigger an action or response, typically by serving as an external input that activates or intensifies a behavioural pattern. This is consistent with the Oxford English Dictionary definition, where *to stimulate* is defined as ‘to rouse to action’, ‘to impart additional energy to an activity’, or ‘to act as a stimulus to’, and *stimulus* as ‘an agency or influence that quickens an activity or process’.<sup>23</sup> In this sense, stimulation denotes more than a background condition or marginal influence, but a proximate and activating role in the behavioural process. Accordingly, when the DSA refers to interface features that ‘stimulate behavioural addictions’, it should be understood as invoking a directional, initiating relationship between the feature and the compulsive behaviour, rather than a diffuse or merely contributory correlation.

Likewise, to cause, in its ordinary usage, denotes a direct or principal agency that brings something about. Thus, while the terms may differ in nuance (*cause* implying origination and *stimulate* implying provocation or intensification), their placement within Recitals 81 and 83, without gradated qualifiers, suggests that both are employed to capture primary, rather than peripheral, relationships between design features and addictive outcomes. The absence of intermediate formulations such as ‘contribute to’, ‘facilitate’, or ‘exacerbate’ reinforces this interpretation. As such, both terms function to delimit the DSA’s obligations to instances where there is a *reasonably demonstrable* and *proximate link* between the platform feature and the emergence or reinforcement of addictive behaviour.

This does not deny the Recitals’ modal ‘may’<sup>24</sup>, which speaks to contingency (the risk may arise) rather than to the strength of the causal predicate. That is to say, we need to distinguish between two different things that the Recitals are doing in the phrase ‘may cause/stimulate’.

First, the modal ‘may’ signals contingency: the risk is possible, not inevitable. It tells us when the obligation applies (ie, in situations where such a risk *could* arise and is evidenced), not how strong the *causal link* must be once we are assessing it. That is, ‘may’ operates as a trigger for inquiry and risk assessment under Articles 34–5 in contexts where the risk *could* materialise, but it does not dilute *the causal threshold* that must then be met. In that way, the modal ‘may’ determines *when* inquiry is triggered, while the strong predicates ‘cause’/‘stimulate’ prescribe *what* must ultimately be proven, so contingency and strict attribution operate sequentially rather than in conflict. In the sense that, the modal ‘may’ merely opens the door to inquiry (is there a risk worth assessing?), and only thereafter do the predicates ‘cause’/‘stimulate’ set the evidential bar for attribution (has this feature been shown to be a proximate driver?), so they mark successive stages of one analysis rather than competing thresholds. Secondly, and as such, the causal predicates that follow, ‘cause’ or ‘stimulate’, set *the strength of the relationship* the law is concerned with.

<sup>20</sup>Digital Services Act, recitals 81 and 83.

<sup>21</sup>*Ibid.*, recital 81.

<sup>22</sup>*Ibid.*, recital 83.

<sup>23</sup>Stimulus, n. Meanings, Etymology and More | Oxford English Dictionary’ <[https://www.oed.com/dictionary/stimulus\\_n](https://www.oed.com/dictionary/stimulus_n)> accessed 27 March 2025.

<sup>24</sup>Digital Services Act, recital 81: ‘such risks may arise . . . in relation to the design of online interfaces . . . which may cause addictive behaviour’; recital 83: ‘online interface design that may stimulate behavioural addictions’.

These are strong predicates. If the drafters had wanted weaker, gradated relationships, they could have written ‘may contribute to’, ‘may facilitate’, or ‘may exacerbate’, but they did not.

This means that the ‘may’ gate, the preliminary trigger in Recitals 81/83 that opens an inquiry whenever a risk could arise, is, by implication, colonised by the later predicates ‘cause/stimulate’. In that, once the door is open, regulators and platforms assemble and evaluate evidence with those stronger predicates front of mind, so the initial possibility screen is applied as if proximate causation must already be demonstrated. The sequence is therefore inverted: the downstream attribution test shapes what is even allowed upstream into the inquiry. Ontologically, the object of concern is recast from a distributed, cohort-level pattern into a specific, feature-linked event; temporally, contingency is compressed into attribution at the threshold. The effect is a measurability bias: diffuse, cumulative harms struggle to enter the frame, while only risks that can be narrated as feature-specific drivers pass through.

### A. Causal predicates and the ‘serious harm’ threshold: doctrinal effects and limits

In addition, the Act sets a high threshold for intervention by imposing obligations on platforms only when addictive behaviours resulting from specific platform features lead to ‘serious negative consequences to a person’s physical and mental well-being’<sup>25</sup> which is a *much higher standard* when compared to harms to public health and minors, where the standard is ‘actual or foreseeable negative effect’.<sup>26</sup> By setting the bar for intervention at ‘serious’ harm (as opposed to other terms such as ‘negative effects’, ‘harmful consequences’, or ‘detrimental effects’, which could imply a lower threshold for regulatory intervention or platform liability), the DSA suggests a high regulatory tolerance for a wide range of compulsive behaviours stimulated by platform features that do not amount to addictive behaviours; this is because, in the Court of Justice of the European Union’s (CJEU) jurisprudence, particularly in cases related to data protection and enforcement, the term ‘serious’ has been consistently interpreted as signifying a higher threshold for enforcement action or liability.<sup>27</sup> In that, a negative effect implies any adverse or detrimental outcome, whether mild, moderate, or temporary, that could foreseeably result from platform use, while *seriousness*, comparatively, denotes a demonstrable and substantial impact that significantly disrupts a person’s physical or mental well-being.

By framing the issue around ‘addiction’, the DSA, I contend, narrows its focus to extreme or severe cases that meet clinical or near-clinical and pathological definitions, which then, in effect, excludes a wide spectrum of behaviours and harms that might still be negative, disruptive, problematic, damaging, detrimental, or even distressing, but fall short of the implied diagnostic weight, severity, or *seriousness* of addiction (eg, compulsively checking notifications to alleviate anxiety, excessive scrolling or video-watching that disrupts sleep patterns, constant comparison with peers leading to lowered self-esteem, or overuse of gaming features causing irritability and reduced productivity).

In light of the foregoing, the logical implication is that the DSA’s goal is not to ban or prohibit features that lead to *problematic, negative*, or detrimental compulsive behaviour, even where such behaviour is foreseeable or frequent. Rather, the Act aims only to mitigate *the more extreme manifestations of such behaviour* when they meet the criteria of serious harm amounting to addiction. If the Act were interpreted to cover platform features that generate broad behavioural

<sup>25</sup>*Ibid.*, recital 83.

<sup>26</sup>*Ibid.*, recital 83.

<sup>27</sup>For an analysis of the meaning of ‘seriousness’ in CJEU jurisprudence see: WR Mbioh, ‘Post-Och Telestyrelsen and Watson and the Investigatory Powers Act 2016’ 3 (2017) *European Data Protection Law Review* 273; WR Mbioh, ‘The TFTP Agreement, Schrems Rights, and the Saugmandsgaard Requirements’ 20 (2016) *Journal of Internet Law* 30; V Janeček and C Teixeira Santos, ‘The Autonomous Concept of “Damage” According to the GDPR and Its Unfortunate Implications: Österreichische Post’ 61 (2024) *Common Market Law Review* 531.

modification,<sup>28</sup> even without causing or stimulating clinical or near-clinical addiction, it would significantly expand its regulatory scope. Platforms would then be required to assess and mitigate a wide range of compulsive or habit-forming design practices, even where these do not result in serious negative harm. This would shift the DSA's focus from narrowly preventing pathological outcomes to more broadly regulating behavioural influence and persuasive design, which would go against the clear intention of the drafters of the Act. This would, in effect, broaden the DSA's application from mitigating *serious/severe outcomes* (eg, addiction) caused or stimulated by a specific interface feature to regulating *behavioural design* more generally to reduce compulsive use.

This suggests that the Act: (1) accepts that platforms may use habit-forming features (eg, scrolling loops, notifications, algorithmic reinforcements) as part of normal user engagement, so long as these do not result in serious mental or physical harm; (2) positions user autonomy and digital literacy tools (eg, usage analytics, toggles, break reminders) as the preferred means for managing low- to moderate-level compulsive use, without imposing mandatory redesign obligations on platforms; and (3) draws an *implicit* boundary between *problematic but permissible* behavioural nudging (what Shoshana Zubov terms behavioural modification) and *pathological, regulated compulsion* requiring legal intervention.

As such (and unlike categorical prohibitions such as those triggered under Article 39(1) DSA, as in the *LinkedIn case*,<sup>29</sup> which ban certain practices (namely, the use of special category personal data for targeted advertising), regardless of the threshold of harm), there is no requirement, for the purposes of mitigating the systemic risk of behavioural addiction, that platforms must prioritise non-personalised content or make it the default. Instead, as suggested by Articles 25 and 27 of the DSA, the Act's approach appears to feature user choice and informational transparency as the preferred method of addressing problematic but permissible behavioural nudging. That is, rather than obligating design changes to limit engagement, the Act promotes mechanisms through which users are informed about design parameters and enabled, at least in principle, to modify their interaction with recommender systems or opt out of certain nudges. Article 25 prohibits manipulative or deceptive design but stops short of prescribing specific interface redesign. Instead, it encourages the Commission to issue guidelines on practices such as prominence of choices, repetitive prompts, or difficulties in terminating a service. Similarly, Article 27 requires platforms to explain how recommender systems work, and importantly, to offer users the ability to modify or choose between options, with such functionality being 'directly and easily accessible' from the interface (ie. features and toggles) as part of a transparency-and autonomy-based approach to managing persuasive or habit-forming design that privileges informed choice over mandatory constraints.

### **B. The TikTok Lite lesson: inducement loops versus attention architecture**

Although it is open to regulators, especially the Commission, to adopt a broader construction of 'addiction' and 'serious' (for example, by treating escalation and chronicity as probative of seriousness) its enforcement practice since the entry into force of the DSA has *tended to confirm the narrower reading*. The main DSA action against TikTok and other social media platforms to date has, for example, centred on child safety (principally the ban on personalised advertising to minors, and duties around age assurance and moderation of access to age-inappropriate content), rather than on engagement features *per se*.<sup>30</sup> In the TikTok Lite case, for example, the Commission alleged that TikTok failed to submit, prior to launch, the Article 34 risk assessment and to set out

<sup>28</sup>Zuboff (n 4).

<sup>29</sup>European Commission, 'Statement by Commissioner Breton on Steps Announced by LinkedIn to Comply with DSA Provisions on Targeted Advertising' <[https://ec.europa.eu/commission/presscorner/detail/en/statement\\_24\\_3172](https://ec.europa.eu/commission/presscorner/detail/en/statement_24_3172)>.

<sup>30</sup>European Commission, 'Commission Opens Formal Proceedings against TikTok under the Digital Services Act' (European Commission 2024) Press release.

proportionate mitigation measures under Article 35. In other words, the primary concern was *procedural* non-compliance. It was not, by itself, a determination that the scheme in question was *substantively* unlawful; *in principle*, a similar scheme could be run lawfully if the platform had completed the assessment, put appropriate mitigations in place (for example, age gating, reward caps, slower accrual, clearer disclosures), and could evidence that residual risk stays below the ‘serious’ threshold.<sup>31</sup>

Launched in France and Spain in April 2024, TikTok Lite was a ‘Task and Reward’ scheme that awarded points for in-app actions (eg, watching videos, liking content, following creators, inviting friends) which users could redeem for monetary vouchers or similar benefits.<sup>32</sup> Put differently, the scheme is notable because, until now, most attention regarding compulsive use had focused on infinite scrolling, autoplay, and like/follower counters; here is something that is distinctive and, in many ways, fundamentally different from these mechanisms. Infinite scroll, autoplay, and notification cadence are built into the grain of the product: they recruit attention through ambient, always-on prompts, variable rewards, and low-friction transitions that feel endogenous to the experience.<sup>33</sup> TikTok Lite’s ‘tasks’ invert that logic. They are exogenous, overt, and transactional: join a programme, perform specified actions, accrue points, redeem value. The motivational structure shifts from hedonic pull to instrumental purpose; what sustains engagement is not the felt tug of the next clip but the prospect of a payout. Phenomenologically, it reads less like being drawn along by an interface and more like doing piece-rate crowdwork inside a social app, closer to a loyalty scheme, bounty programme, or referral/affiliate drive than to the subtle hooks of frictionless feeds. It also repositions the user: not simply a viewer nudged by design, but a micro-contractor executing growth tasks that manufacture engagement (watch X minutes, follow Y accounts, invite Z friends), with obvious user-acquisition and metrics-inflation effects.<sup>34</sup> For that reason, it appears unlike ‘compulsion’ in the ordinary sense: the behaviour is purposeful, time-for-value, and ends when the inducement is withdrawn. Hence, the scheme grafted a pay-per-action layer onto the platform that turned attention into remunerated labour and recruitment in a way that is categorically different from the background mechanics that keep people scrolling when no one is paying them to do so.

If that reading holds, the safest inference is that enforcement will gravitate towards engagement devices that are both high-amplitude (that is, producing large, short-run, statistically salient shifts in engagement metrics such as session starts, time-on-task, or conversions) and clearly attributable to a particular ‘treatment’. In practice, that means schemes that convert specific user actions into redeemable value or immediate prizes; referral bounties and ‘invite-to-earn’ programmes; watch-time or task-completion rewards; loyalty points with cash-equivalent redemption; and gambling-adjacent mechanics such as spins, loot-box-style draws, or time-limited challenges with tangible rewards. Consequently, features that lack those inducement mechanics’ core properties (ie, explicit, extrinsic rewards; specific, event-based triggers with immediate or cash-equivalent payoffs; time-bounded scarcity or uncertainty; and precise, loggable attribution to a specific ‘treatment’) are likely to attract less near-term enforcement priority. For such features, mechanisms such as autoplay, infinite scroll, and public like/follower counters are seen as more background, default-on elements of the attention architecture: endogenous to the interface, continuous rather than episodic, driven by low-friction flow and socially mediated validation, and yielding *diffuse, cumulative effects that resist point attribution*.

<sup>31</sup>European Commission, ‘Commission Opens Formal Proceedings against TikTok under the Digital Services Act’ (n 15).

<sup>32</sup>L O’Carroll, ‘TikTok Reward-to-Watch Feature Suspended after EU Threats to Block It’ *The Guardian* (24 April 2024) <<https://www.theguardian.com/technology/2024/apr/24/tiktok-reward-to-watch-feature-suspended-after-eu-threats-to-block-it>> accessed 31 October 2025; FY Chee, ‘TikTok Suspends New App’s Reward Programme amid EU Concerns’ *Reuters* (24 April 2024) <<https://www.reuters.com/technology/tiktok-suspend-tiktok-lites-reward-programme-amid-eu-concerns-2024-04-24/>> accessed 31 October 2025.

<sup>33</sup>O’Carroll (n 32); Chee (n 32).

<sup>34</sup>O’Carroll (n 32); Chee (n 32).

On current practice, such background mechanics seem to be channelled toward procedural duties and *user-option mitigations* rather than direct prohibitions. By ‘user-option mitigations’ I mean controls that leave the attention feature in place but shift the choice to the user (or guardian) to limit or disable it. In practice, this covers things such as an autoplay on/off toggle, granular notification settings and quiet hours, ‘take-a-break’ prompts and session timers, time-spent dashboards, and switches to a chronological or non-personalised feed. These measures are presented as options the user must select, rather than as default design changes, and they allow a platform to document ‘reasonable, proportionate’ mitigation under Articles 34–5 while keeping the underlying engagement mechanics, or tools of ‘behavioural modification’,<sup>35</sup> unchanged by default.

Treating retention mechanics as default while offering opt-outs recasts addiction risk as a matter of personal configuration rather than a property of the system. Ontologically, the object of regulation becomes the user’s choice (the toggle, the dashboard), so ‘harm’ is recognised only when an individual elects to curtail it; the architecture that produces compulsive engagement remains background. Ethically, responsibility shifts from designers to users (or parents), inviting a narrative of self-management and blame when harms persist. Distributionally, this arrangement loads risk (as I will illustrate in Section 5 below) onto groups least able to manage settings or sustain vigilance: younger teens; users with attention, mood, or executive-function difficulties; those with lower digital literacy or limited language support; time-poor households, and communities for whom connectivity and platforms double as social lifelines. The accepted consequences are cumulative, cohort-level harms that fall below a ‘serious’ threshold yet compound over time: sleep displacement from late-night autoplay, escalation of checking driven by notification cadence, mood volatility from social comparison, and displacement of offline study, rest, and relationships. Compliance artefacts (dashboards, prompts, toggles) document ‘reasonable’ mitigation while leaving the incentive structure untouched and, in so doing, externalising costs to families, schools, and public health. By such means, *the policy choice normalises everyday attention capture as an acceptable by-product of platform design*, treating its costs as user-borne unless and until a high, easily attributable threshold is crossed.

This is a concern the European Parliament has recognised. For example, in its resolution of 12 December 2023 on addictive design, it urged the Commission to table a legislative package, including a ‘digital right not to be disturbed’ and default-off attention-seeking features, precisely because the DSA does not address (because of its restrictive reading of addiction) adequately the everyday mechanics of attention capture.<sup>36</sup> The Commission’s Digital Fairness Fitness Check (3 October 2024)<sup>37</sup> has reached a similar position, noting that current EU consumer law, including the DSA, leaves manipulative/addictive design only partially addressed and, in practice, largely relegated to user-option measures (opt-outs, toggles, disclosures, and transparency notices), rather than mandatory redesign or default changes.<sup>38</sup> That is, both the Parliament and the Commission have, in effect, acknowledged that the DSA’s current framing (requiring a proximate, feature-specific causal link and a ‘serious outcome’) and its preference for user-option remedies (ie, opt-outs, toggles, time-spent dashboards, ‘take-a-break’ prompts, transparency notices) leave the everyday attention mechanics largely outside the statute’s practical reach. Put differently, the Act

<sup>35</sup>Zuboff (n 4).

<sup>36</sup>European Parliament, ‘Addictive Design of Online Services and Consumer Protection in the EU Single Market’ (2023).

<sup>37</sup>European Commission, ‘Commission Staff Working Document: Fitness Check of EU Consumer Law on Digital Fairness’ (European Commission 2024) Staff Working Document <[https://commission.europa.eu/document/download/707d7404-78e5-4aef-acfa-82b4cf639f55\\_en?filename=Commission+Staff+Working+Document+Fitness+Check+on+EU+consumer+law+on+digital+fairness.pdf](https://commission.europa.eu/document/download/707d7404-78e5-4aef-acfa-82b4cf639f55_en?filename=Commission+Staff+Working+Document+Fitness+Check+on+EU+consumer+law+on+digital+fairness.pdf)> accessed 17 October 2025.

<sup>38</sup>European Commission, ‘Digital Fairness Act’ (*Better Regulation Portal – Have your say (Initiative 14622)*, 17 July 2025) <[https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14622-Digital-Fairness-Act\\_en](https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14622-Digital-Fairness-Act_en)> accessed 17 October 2025.

tends to treat engagement risks as matters for disclosure and optional control rather than for mandatory redesign or default-off configurations.

#### 4. The limits of legal causality: addiction, persuasive design, and the mechanistic imaginary of regulation

I contend that one core factor behind this misalignment is the DSA's systemic risk model, which conceptualises platform addiction as isolated, traceable outcomes produced by identifiable features within digital environments. Addiction, in this framework, is not understood as a complex, context-dependent process but as an effect that can be causally linked to specific design elements or functionalities, such as infinite scroll, autoplay, or variable notification systems. Under this model, digital platforms are treated as machines whose harmful outputs can be controlled by regulating their internal components. Accordingly, platform features are regarded as separable and isolatable mechanisms that can be individually evaluated for their risk potential. If a feature 'causes' or 'stimulates' addiction (as narrowly defined by its capacity to produce serious physical or mental consequences), then regulatory obligations are triggered. If no such direct or proximate link can be established, or if the resulting harm does not meet the threshold of seriousness, then the system is presumed to be functioning within acceptable parameters.

This approach mirrors a conventional input–output logic: if Feature A produces Harm B, and Harm B crosses a defined threshold, then Mitigation C must follow. It assumes that the platform environment can be broken down into bounded, modular parts, and that interventions targeted at one part (eg, a recommender system or interface loop) can correct or neutralise the harm without, as I will explain below, needing to address the broader ecosystem of behavioural, psychological, and social dynamics in which that harm takes shape.

At the centre of this framework, therefore, lies an ontology of separation: harms are treated as external to the system; users are treated as autonomous actors whose vulnerabilities are largely background conditions; and platforms are treated as bounded entities whose features can be modified independently of the socio-technical milieu in which they are used. Regulation is positioned as an external force, intervening from outside to correct malfunctioning parts, rather than as a participant in shaping the very conditions through which harm is recognised, managed, and enacted.

Furthermore, by requiring a demonstrable causal connection between a feature and a serious outcome, the Act implicitly positions risk as an exception rather than a structural condition. It frames addiction as a deviance from normal platform use rather than as a possibility that is latent, if not structurally incentivised, within the ordinary functioning of persuasive design. This reinforces a default assumption of platform neutrality: unless serious harm is proven, the system is presumed safe.

What this produces, I contend, is a narrow, individuated conception of harm. Regulatory obligations focus on whether a platform feature leads to severe outcomes for individuals, rather than whether the platform environment systematically cultivates conditions of compulsive engagement. Features are treated as if they operate in isolation, user harms are defined by diagnostic thresholds, and responsibility is framed in terms of compliance with enumerated duties, rather than as part of a shared, embodied, socio-technical field of influence and co-production. Or, to put it more simply, the DSA's approach to platform addiction rests on a linear, mechanistic model of causality, in which harms are viewed as traceable to isolated sources, remediable through proportionate design modifications, and exceptional rather than endemic. It is an approach that preserves the intelligibility of law and governance through containment: defining the problem in terms that regulation can manage, even if that means failing to account for the full complexity of how such harms actually emerge as I will now explain.

## 5. Four models of platform addiction: internal dispositions, external triggers, social contexts, algorithmic forces

In contrast to the DSA's mechanistic and feature-specific treatment of addiction, academic literature identifies a more expansive and differentiated understanding of how compulsive platform use arises. This section outlines four distinct but overlapping models through which platform addiction has been conceptualised: the clinical/psychiatric model, which foregrounds internal vulnerabilities and mental health comorbidities; the neurological model, which focuses on brain development, neurotransmitter function, and genetic predisposition; the Skinnerian model, which emphasises behavioural reinforcement and environmental conditioning; and the psychosocial model, which situates addiction within cultural, relational, and structural contexts. Each model highlights different vectors of risk (whether internal, behavioural, social, or algorithmic) and shows how platform use becomes compulsive and/or addictive through complex and intersecting mechanisms.

### A. The clinical and psychiatric model: addiction as individual vulnerability

The first is the clinical/psychiatric model, which emphasises internal characteristics and individual attributes. This model considers addiction to be primarily rooted in personality traits and psychological vulnerabilities: ie, enduring patterns of thought, emotion, and behaviour that shape how individuals perceive, respond to, and manage their internal and external environments.<sup>39</sup>

For example, research in this area suggests a strong connection between internet/platform addiction and mental health conditions, which may either pre-exist or manifest as comorbidities alongside addiction, such as depression, anxiety, attention-deficit/hyperactivity disorder (ADHD), and obsessive-compulsive disorder (OCD).<sup>40</sup> Additionally, platform addiction has been linked to certain personality traits, including impulsivity, neuroticism, low self-esteem, and other contributing factors, such as psychological inflexibility (ie, difficulty coping with stress and challenging emotions), experiential avoidance (the tendency to avoid negative experiences or emotions), and sensation-seeking behaviour, or, in other words, a preference for novel and stimulating experiences.<sup>41</sup> For individuals with these internal traits, addiction emerges not only because they are predisposed but also because they use platforms as coping mechanisms to manage negative emotions, such as loneliness and isolation, boredom, stress, or *dysphoric moods*<sup>42</sup> and *intrusive thoughts*.<sup>43</sup> Platforms are used, in this context, as an immediate and accessible source of distraction, stimulation, or validation.

<sup>39</sup>S Bernardi and S Pallanti, 'Internet Addiction: A Descriptive Clinical Study Focusing on Comorbidities and Dissociative Symptoms' 50 (2009) *Comprehensive Psychiatry* 510; A Weinstein et al, 'Internet Addiction Disorder: Overview and Controversies' in KP Rosenberg and LC Feder (eds), *Behavioral Addictions* (Academic Press 2014) 99; Y Olsen, 'What Is Addiction? History, Terminology, and Core Concepts' 106 (2022) *Medical Clinics* 1.

<sup>40</sup>Bernardi and Pallanti (n 39); LT Lam, 'Risk Factors of Internet Addiction and the Health Effect of Internet Addiction on Adolescents: A Systematic Review of Longitudinal and Prospective Studies' 16 (2014) *Current psychiatry reports* 1; HM Pontes, DJ Kuss and MD Griffiths, 'Clinical Psychology of Internet Addiction: A Review of Its Conceptualization, Prevalence, Neuronal Processes, and Implications for Treatment' [2015] *Neuroscience and Neuroeconomics* 11; DJ Kuss and O Lopez-Fernandez, 'Internet Addiction and Problematic Internet Use: A Systematic Review of Clinical Research' 6 (2016) *World Journal of Psychiatry* 143.

<sup>41</sup>D J Kuss et al, 'Internet Addiction: A Systematic Review of Epidemiological Research for the Last Decade' 20 (2014) *Current pharmaceutical design* 4026; Kuss and Lopez-Fernandez (n 25); RE Pezoa-Jares, IL Espinoza-Luna, and JA Vasquez-Medina, 'Internet Addiction: A Review' 6 (2012) *Journal of Addiction Research and Therapy* S.

<sup>42</sup>KS Kurniasanti et al, 'Internet Addiction: A New Addiction?' 28 (2019) *Medical Journal of Indonesia* 82; K Young, 'The Evolution of Internet Addiction Disorder' in C Montag and M Reuter (eds), *Internet Addiction: Neuroscientific Approaches and Therapeutic Implications Including Smartphone Addiction* (Springer 2017) 3.

<sup>43</sup>M Mihajlov and L Vejmelka, 'Internet Addiction: A Review of the First Twenty Years' 29 (2017) *Psychiatria Danubina* 260; H Cash et al, 'Internet Addiction: A Brief Summary of Research and Practice' 8 (2012) *Current Psychiatry Reviews* 292.

This dynamic is particularly pronounced in individuals with mental health conditions such as eating disorders and body dysmorphia, where traits such as low self-esteem, self-objectification, perfectionism, and social comparison amplify vulnerability.<sup>44</sup> Such individuals are often predisposed to engage with platforms where curated, idealised, and aspirational content dominates, seeking validation or inspiration that may ultimately reinforce their vulnerabilities. However, the more time they spend on these platforms, the more they are exposed to content that intensifies their insecurities and reinforces harmful behaviours. This exposure increases their emotional distress and feelings of inadequacy, which then drives them to seek further distraction, validation, or solutions from the same platforms. In doing so, they become more entrenched in a pattern of compulsive engagement, where the platform both worsens their underlying condition and becomes their primary means of attempting to cope with it.<sup>45</sup> This pattern has also been observed in individuals with other mental health conditions, including anxiety disorders, depression, and obsessive-compulsive disorder (OCD), where similar dynamics of seeking validation, distraction, or reassurance through platform engagement exacerbate their symptoms.<sup>46</sup> Moreover, comorbidities, such as eating disorders with anxiety or depression, amplify vulnerability and reinforce maladaptive coping.<sup>47</sup>

### **B. The neurological model: addiction in the brain**

The neural biological model for understanding risk factors in platform addiction, particularly in teenagers, combines insights from brain development, neurotransmitter function, genetic predisposition, and psychological vulnerabilities.<sup>48</sup> It suggests that teenagers are especially at risk because of the way their brains develop during this period.<sup>49</sup> Specifically, the prefrontal cortex, responsible for regulating impulse control, planning, and decision-making, matures more slowly than the limbic system, which drives emotions and reward-seeking behaviours.<sup>50</sup> This imbalance means that teenagers are more likely to engage in impulsive behaviours like excessive platform use, making them more vulnerable to addiction.<sup>51</sup>

Dopamine, a key neurotransmitter involved in the brain's reward system, plays a central role in platform addiction. In teenagers predisposed to this behaviour, research shows increased dopamine activity in reward-related brain regions can, with overstimulation, reduce the availability of receptors (ie, proteins in brain cells that bind to dopamine, enabling its effects on mood, motivation, and pleasure) over time.<sup>52</sup> This means that typical rewards (such as social

<sup>44</sup>EP Meier and J Gray, 'Facebook Photo Activity Associated with Body Image Disturbance in Adolescent Girls' 17 (2014) *Cyberpsychology, Behavior, and Social Networking* 199; D Carlson Jones, 'Social Comparison and Body Image: Attractiveness Comparisons to Models and Peers among Adolescent Girls and Boys' 45 (2001) *Sex Roles* 645; M Gupta, A Jassi and G Krebs, 'The Association between Social Media Use and Body Dysmorphic Symptoms in Young People' 14 (2023) *Frontiers in Psychology* 1231801; L Cerniglia et al, 'Internet Addiction in Adolescence: Neurobiological, Psychosocial and Clinical Issues' 76 (2017) *Neuroscience & Biobehavioral Reviews* 174; G Haddad, 'The Impact of Social Media on Body Dissatisfaction and Dysmorphia in Adolescents in the US' (PhD Thesis, University of Pittsburgh 2022).

<sup>45</sup>Haddad (n 44); Gupta, Jassi and Krebs (n 44).

<sup>46</sup>Bernardi and Pallanti (n 24); DS Bickham, 'Current Research and Viewpoints on Internet Addiction in Adolescents' 9 (2021) *Current Pediatrics Reports* 1.

<sup>47</sup>Mihajlov and Vejmelka (n 43); Kuss and Lopez-Fernandez (n 40).

<sup>48</sup>Pezoa-Jares, Espinoza-Luna et al (n 41); Young (n 42).

<sup>49</sup>Cerniglia et al, (n 44); S Tereshchenko and E Kasparov, 'Neurobiological Risk Factors for the Development of Internet Addiction in Adolescents' 9 (2019) *Behavioral Sciences* 62; C-Y Wee et al, 'Disrupted Brain Functional Network in Internet Addiction Disorder: A Resting-State Functional Magnetic Resonance Imaging Study' 9 (2014) *PloS one* e107306.

<sup>50</sup>Cerniglia et al, (n 44); M Brand, KS Young and C Laier, 'Prefrontal Control and Internet Addiction: A Theoretical Model and Review of Neuropsychological and Neuroimaging Findings' 8 (2014) *Frontiers in Human Neuroscience* 375.

<sup>51</sup>Tereshchenko and Kasparov (n 49).

<sup>52</sup>Brand, Young and Laier (n 50); T Love et al, 'Neuroscience of Internet Pornography Addiction: A Review and Update' 5 (2015) *Behavioral Sciences* 388; Pontes, Kuss and Griffiths (n 25).

interaction or physical activities) feel less stimulating because the brain has adapted to the constant, high levels of stimulation provided by platforms. This reduced sensitivity to natural rewards leads users to seek out platform-based activities more frequently, as these activities continue to provide the intense levels of stimulation their brains now crave.<sup>53</sup> This can create a dependency on platform-based activities to achieve the same level of stimulation, much like what is observed in substance addiction, where the brain adapts to repeated exposure to addictive stimuli by altering its neurochemical balance; that is, by reducing the number of dopamine receptors or their sensitivity, the brain becomes less responsive to rewards outside the addictive stimulus and so perpetuates the cycle of dependence.<sup>54</sup>

Neuroimaging studies have also provided some evidence of structural and functional changes in the brains of individuals with platform addiction. For instance, reductions in grey matter density (ie, the volume of brain tissue containing neuronal cell bodies, which is essential for processing information, decision-making, and regulating behaviour) have been observed in areas such as the prefrontal cortex, orbitofrontal cortex, and supplementary motor areas, regions responsible for executive functions such as decision-making, self-control, planning, emotional regulation, and habit formation.<sup>55</sup> This suggests that platform addiction may impair the brain's capacity to regulate impulsive behaviours and make thoughtful, deliberate decisions. Moreover, functional imaging techniques, such as fMRI (functional magnetic resonance imaging), also show that the brains of those suffering from platform addiction exhibit heightened activity in reward-related areas, particularly the orbitofrontal cortex and anterior cingulate, when exposed to platform-related stimuli, indicating that these individuals have an exaggerated neural response to platform cues.<sup>56</sup> This heightened sensitivity to stimuli reinforces compulsive engagement by making platform-related activities feel disproportionately rewarding, even when the overall experience may be detrimental.<sup>57</sup>

In addition, there is growing evidence suggesting a genetic predisposition to platform addiction, rooted in the neurobiological mechanisms associated with reward processing and emotional regulation.<sup>58</sup> Research indicates that individuals with certain genetic variations may have a reduced number of dopamine receptors or impaired serotonin and dopamine functioning that make it more difficult for them to experience typical levels of pleasure from everyday activities.<sup>59</sup> This deficiency can drive individuals to seek out activities that provide heightened dopamine stimulation, such as excessive platform use, to compensate for the lack of reward sensitivity. Studies of twins across various populations (including Chinese, Dutch, Australian, and German groups) have shown that genetics play a significant role in internet addiction, accounting for 21–66% of differences depending on the group and study. Specific genetic variations, such as changes in the dopamine D2 receptor gene (linked to lower dopamine availability), the COMT gene (which affects how dopamine is processed), and the serotonin transporter gene (influencing mood and emotional regulation), have all been associated with a heightened risk of developing platform addiction.<sup>60</sup>

---

<sup>53</sup>A Griffin, 'Adolescent Neurological Development and Implications for Health and Well-Being', *Healthcare* (MDPI 2017); Brand et al (n 50); Pontes et al (n 40).

<sup>54</sup>Brand et al (n 50); Pontes et al (n 40).

<sup>55</sup>Brand et al (n 50); Y Zhu, H Zhang, and M Tian, 'Molecular and Functional Imaging of Internet Addiction' 2015 (2015) *BioMed research international* 378675; Cerniglia et al, (n 44).

<sup>56</sup>G Sepede et al, 'Functional Magnetic Resonance Imaging of Internet Addiction in Young Adults' 8 (2016) *World Journal of Radiology* 210; Brand et al (n 50).

<sup>57</sup>*Ibid.* (n 50).

<sup>58</sup>Cash et al, (n 43); Tereshchenko and Kasparov (n 49).

<sup>59</sup>Cash et al, (n 43).

<sup>60</sup>Tereshchenko and Kasparov (n 49); Pezoa-Jares et al (n 41).

### C. The Skinnerian model: addiction as conditioned behaviour

The Skinnerian model of platform addiction, rooted in the behavioural theories of B.F. Skinner, provides a complementary yet distinct perspective from the neurobiological model. Whereas the neurobiological model focuses on the physical and chemical changes within the brain, the Skinnerian model emphasises the environmental and behavioural mechanisms that drive addiction, particularly the role of reinforcement and operant conditioning.<sup>61</sup> Reinforcement refers to processes that increase the likelihood of a behaviour being repeated, while operant conditioning is the method by which behaviours are shaped through the application of rewards (positive reinforcement) or the removal of unpleasant states (negative reinforcement).<sup>62</sup> Thus, on platforms, positive reinforcement occurs when social validation (likes, comments, shares) increases the likelihood of further posting or engagement. Negative reinforcement arises when use reduces aversive states such as anxiety, boredom, or loneliness. Within a behaviourist account, these internal states are not treated as explanatory mechanisms; they are observable antecedents or consequences correlated with behaviour, while conditioning is defined by the contingencies between actions and their reinforcing outcomes. These forms of reinforcement, it is contended, are central to how platforms capture and retain user attention.<sup>63</sup>

It is also suggested that platforms often use variable-ratio reinforcement schedules, a particularly powerful tool in operant conditioning where rewards (eg, new notifications, unpredictable content in feeds, or in-game achievements) are delivered at unpredictable intervals in ways that mimic the mechanisms of slot machines.<sup>64</sup> As a result, users are drawn to check and recheck the platform, not knowing when the next reward will arrive but anticipating its possibility, which keeps them engaged in a cycle of compulsive use. This unpredictability makes the rewards more psychologically impactful and fosters a sense of dependency, as users are conditioned to keep engaging in the hope of obtaining further gratification, which, drawing from the neurobiological model, manifests as heightened dopamine release in the brain's reward centres (eg, the nucleus accumbens and ventral striatum) when rewards are received. This, in turn, reinforces compulsive behaviour and increases sensitivity to platform-related cues, making disengagement more difficult.<sup>65</sup>

Given teenagers' heightened reward sensitivity and underdeveloped prefrontal cortex (responsible for impulse control and decision-making), but more dominant activity in their limbic system, especially the amygdala, which processes emotions and rewards;<sup>66</sup> this means they are especially vulnerable to this. Their brains are more reactive to rewards and emotionally charged stimuli but less capable of self-regulation and delayed gratification.<sup>67</sup> Meaning that, for them, variable-ratio reinforcement is especially effective at maintaining engagement and

<sup>61</sup>Y Sun and Y Zhang, 'A Review of Theories and Models Applied in Studies of Social Media Addiction and Implications for Future Research' 114 (2021) *Addictive behaviors* 106699; CS Andreassen, 'Online Social Network Site Addiction: A Comprehensive Review' 2 (2015) *Current addiction reports* 175; A Alter, *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked* (Penguin Press 2017).

<sup>62</sup>BF Skinner, 'The Evolution of Behaviour 1' in *Behaviour Analysis and Contemporary Psychology* (Routledge 2022); R Groom et al, 'BF Skinner: Environmental Reinforcement in Coaching' in L Nelson, R Groom and P Potrac (eds), *Learning in Sports Coaching* (Routledge 2016).

<sup>63</sup>Alter (n 46); BJ Fogg, 'A Behavior Model for Persuasive Design', *Proceedings of the 4th International Conference on Persuasive Technology* (Association for Computing Machinery 2009) <<https://doi.org/10.1145/1541948.1541999>> accessed 18 February 2023; Zuboff (n 4).

<sup>64</sup>Alter (n 61).

<sup>65</sup>L Clark and M Zack, 'Engineered Highs: Reward Variability and Frequency as Potential Prerequisites of Behavioural Addiction' 140 (2023) *Addictive Behaviors* 107626; V Medrano, 'The Effects of Social Media on Anxiety, Reward Sensitivity, and Risk-Taking on Emerging Adults'; N Ihssen and M Wadsley, 'A Reward and Incentive-Sensitization Perspective on Compulsive Use of Social Networking Sites—Wanting but Not Liking Predicts Checking Frequency and Problematic Use Behavior' 116 (2021) *Addictive Behaviors* 106808.

<sup>66</sup>Cerniglia et al, (n 44); Brand et al (n 50).

<sup>67</sup>Cerniglia et al, (n 44); Brand et al (n 50).

reinforcing compulsive platform use. The combination of heightened reward response and limited self-regulatory capacity creates a perfect storm, where the unpredictable rewards offered by platforms exploit their developmental vulnerabilities and can, potentially, lock them into cycles of use.

#### **D. The psychosocial model: addiction in a cultural and relational context**

The psychosocial model of platform addiction offers a lens through which to understand how societal norms, interpersonal relationships, cultural dynamics, and individual vulnerabilities intersect with platform design to generate compulsive or addictive behaviours.<sup>68</sup> Unlike models that focus predominantly on individual traits (eg, psychiatric or neurological models) or platform mechanisms (eg, Skinnerian models), the psychosocial approach centres on the complex relationship between these elements. That is, the psychosocial model recognises the role of external stabilising or intensifying factors, such as family, social support systems, and broader environmental influences in shaping an individual's relationship with platforms.<sup>69</sup>

For example, teenagers in supportive environments, with strong family bonds and open communication, are better equipped to manage the pressures of social comparison and platform use.<sup>70</sup> Supportive relationships foster resilience by providing alternative sources of validation, emotional reassurance, and practical guidance on navigating digital spaces. Conversely, when these stabilising factors are absent or compromised, external circumstances can intensify a young person's (or, for that matter, any user's) reliance on platforms. For instance, family dynamics characterised by emotional neglect, unresolved conflict, or financial strain may leave teenagers feeling isolated and lacking a sense of security.<sup>71</sup> These circumstances often make platforms a substitute for the emotional connection and reassurance they lack at home.<sup>72</sup>

Moreover, social rejection or bullying, whether experienced in school, online, or both, can further intensify these feelings, as platforms become a refuge for distraction or a way to monitor and evaluate their social standing.<sup>73</sup> However, such use frequently deepens feelings of inadequacy, especially when they encounter idealised portrayals of peers or influencers who appear to lead more successful, fulfilling lives; a dynamic particularly significant in the context of *upward social comparison*, where the perceived proximity between users and these individuals (whether through shared age, similar life circumstances, or overlapping social networks) creates heightened expectations or pressures to achieve comparable success.<sup>74</sup> This is because the perceived proximity

<sup>68</sup>Cerniglia et al, (n 44); M Peris et al, 'Psychological Risk Factors That Predict Social Networking and Internet Addiction in Adolescents' 17 (2020) *International Journal of Environmental Research and Public Health* 4598; W Zhang et al, 'Demographic Characteristics, Family Environment and Psychosocial Factors Affecting Internet Addiction in Chinese Adolescents' 315 (2022) *Journal of Affective Disorders* 130; D Sarti et al, 'Tell Me a Story: Socio-Emotional Functioning, Well-Being and Problematic Smartphone Use in Adolescents with Specific Learning Disabilities' 10 (2019) *Frontiers in Psychology* 482087.

<sup>69</sup>E Esen and DMM Siyez, 'An Investigation of Psycho-Social Variables in Predicting Internet Addiction among Adolescents' 4 (2011) *Turkish Psychological Counseling and Guidance Journal* 127; Peris et al, (n 53).

<sup>70</sup>Pontes et al (n 40); A Fumero et al, 'Personal and Social Factors Involved in Internet Addiction among Adolescents: A Meta-Analysis' 86 (2018) *Computers in Human Behavior* 387.

<sup>71</sup>PE Nathan, M Conrad and AH Skinstad, 'History of the Concept of Addiction' 12 (2016) *Annual Review of Clinical Psychology* 29.

<sup>72</sup>MK Sharma and TS Palanichamy, 'Psychosocial Interventions for Technological Addictions' 60 (2018) *Indian Journal of Psychiatry* S541; Peris et al, (n 68).

<sup>73</sup>P Best, R Manktelow and B Taylor, 'Online Communication, Social Media and Adolescent Wellbeing: A Systematic Narrative Review' 41 (2014) *Children and Youth Services Review* 27; Michelle O'reilly et al, 'Is Social Media Bad for Mental Health and Wellbeing? Exploring the Perspectives of Adolescents' 23 (2018) *Clinical Child Psychology and Psychiatry* 601.

<sup>74</sup>M Tiggemann and J Polivy, 'Upward and Downward: Social Comparison Processing of Thin Idealized Media Images' 34 (2010) *Psychology of Women Quarterly* 356; Y Alfasi, 'The Grass Is Always Greener on My Friends' Profiles: The Effect of Facebook Social Comparison on State Self-Esteem and Depression' 147 (2019) *Personality and Individual Differences* 111;

magnifies the sense that the achievements, appearance, or lifestyles of others are realistically attainable or should be within reach. For instance, seeing peers of the same age or from similar social backgrounds achieving significant milestones, gaining popularity, or showcasing material success can lead individuals to internalise these as benchmarks they are expected to meet. This proximity reduces the psychological distance between the observer and the observed, which makes the comparisons feel more personal and the perceived inadequacies more acute. The sense of ‘if they can do it, why can’t I?’ fosters a belief that failing to achieve similar outcomes reflects a personal deficiency rather than broader contextual differences, and so increases the pressure to perform, conform, or strive for comparable success, even when the idealised images they are comparing themselves to may be unattainable or unrealistic.<sup>75</sup>

In addition, the absence of accessible coping mechanisms or activities outside the digital sphere (whether due to financial constraints, lack of local opportunities, or barriers to mental health support such as insufficient state investment, cuts to services, or the predominance of costly private options) reduces alternatives for emotional relief and self-expression.<sup>76</sup> Teenagers in these situations may turn to platforms as a means of distraction or temporary escape from stress, anxiety, or feelings of worthlessness.<sup>77</sup> Yet, platforms can reinforce these emotions through exposure to curated content that idealises unattainable lifestyles or amplifies social hierarchies that further isolate the individual.<sup>78</sup>

The role of peer relationships also emerges as an important factor.<sup>79</sup> While supportive friendships can buffer the effects of online comparisons, teenagers with strained or superficial peer connections often turn to platforms to compensate for their sense of disconnection. Here, the constant visibility of peers’ activities and their social validation metrics, such as likes or comments, reinforces feelings of exclusion or inadequacy, especially for those who already struggle with self-esteem.<sup>80</sup> Over time, this can lead to a feedback loop where the platform simultaneously serves as a source of comfort (ie, providing distraction, a sense of connection, or a temporary escape from negative emotions) and a reinforcement of negative self-perceptions, that is, amplifying feelings of inadequacy, exclusion, or low self-worth through repeated exposure to curated content, social validation metrics, and comparisons with peers or influencers who appear more successful or fulfilled.<sup>81</sup>

---

H Appel, AL Gerlach and J Crusius, ‘The Interplay between Facebook Use, Social Comparison, Envy, and Depression’ 9 (2016) *Current Opinion in Psychology* 44.

<sup>75</sup>D Schmuck et al, ‘“Looking Up and Feeling Down”: The Influence of Mobile Social Networking Site Use on Upward Social Comparison, Self-Esteem, and Well-Being of Adult Smartphone Users’ 42 (2019) *Telematics and informatics* 101240; R Aubry, A Quiamzade and LL Meier, ‘Depressive Symptoms and Upward Social Comparisons during Instagram Use: A Vicious Circle’ 217 (2024) *Personality and Individual Differences* 112458.

<sup>76</sup>B Brossard and A Chandler, *Explaining Mental Illness: Sociological Perspectives* (Policy Press 2022); ID Sigfusdottir et al, ‘Stress and Adolescent Well-Being: The Need for an Interdisciplinary Framework’ 32 (2017) *Health Promotion International* 1081.

<sup>77</sup>C Cheng, P Sun and K-K Mak, ‘Internet Addiction and Psychosocial Maladjustment: Avoidant Coping and Coping Inflexibility as Psychological Mechanisms’ 18 (2015) *Cyberpsychology, Behavior, and Social Networking* 539; ML McNicol and EB Thorsteinsson, ‘Internet Addiction, Psychological Distress, and Coping Responses among Adolescents and Adults’ 20 (2017) *Cyberpsychology, Behavior, and Social Networking* 296.

<sup>78</sup>Alfasi (n 74).

<sup>79</sup>GC Huang et al, ‘Peer Influences: The Impact of Online and Offline Friendship Networks on Adolescent Smoking and Alcohol Use’ 54 (2014) *Journal of Adolescent Health* 508; DJ Kuss and MD Griffiths, ‘Online Social Networking and Addiction: a Review of the Psychological Literature’ 8 (2011) *International Journal of Environmental Research and Public Health* 3528; I Savolainen et al, ‘Addictive Behaviors and Psychological Distress among Adolescents and Emerging Adults: A Mediating Role of Peer Group Identification’ 7 (2018) *Addictive Behaviors Reports* 75.

<sup>80</sup>EA Vogel et al, ‘Social Comparison, Social Media, and Self-Esteem’. 3 (2014) *Psychology of Popular Media Culture* 206; Schmuck et al, (n 75).

<sup>81</sup>Schmuck et al, (n 75).

## 6. The machine-learning model: how recommendation systems drive platform addiction

However, one of the key contributory factors of platform addiction that is often missed or inadequately addressed is the role of machine-learning, especially in recommendation systems, in making platforms highly addictive: ie, the issue of algorithmic addiction. In that, recommendation systems on platforms such as YouTube, TikTok, or Instagram are engineered to operate with extraordinary predictive accuracy by relying on a range of advanced algorithmic techniques.<sup>82</sup> At a foundational level, they function in two primary phases: an offline training phase and a live interaction phase, each of which contributes uniquely to their ability to predict and influence user behaviour.<sup>83</sup>

In the offline phase, platforms collect and process massive historical datasets that capture hundreds of billions of user interactions that include both explicit data (such as likes, shares, comments, or subscriptions) and implicit data (such as watch time, scrolling behaviour, or video completion rates) and contextual information, such as users' device type, location, and time of activity, and inferred demographic information.<sup>84</sup> This data is then used to train a deep neural network (DNN), a computational architecture composed of multiple interconnected layers, that can identify latent and multi-dimensional relationships between different user behaviours and contextual factors to refine and optimise its ability to predict future actions.<sup>85</sup>

In addition, because of their size, the larger recommendation systems also benefit from two statistical principles: the law of large numbers and scaling laws. The law of large numbers holds that if we observe or track random probabilistic behaviour on a large enough scale, we will converge on the actual likelihood or probability of an event occurring.<sup>86</sup> For example, if we toss a coin only 10 or 100 times, there might still be significant noise or variation in the results, making it difficult to determine the true underlying probability of getting heads or tails. However, if we repeat the toss 1,000 or 100,000 times, the results will converge towards the true probability of the event: 50% for heads and 50% for tails. This convergence means that when we make decisions based on the likelihood of a coin toss, we can reliably trust that there is an equal probability of heads or tails.<sup>87</sup>

Similarly, large-scale deep neural networks, which are increasingly used in recommendation systems, benefit from this principle. Systems like YouTube's or TikTok's recommendation algorithms are trained on hundreds of billions of historical user interactions and behaviours.<sup>88</sup> These vast datasets allow the system to make probabilistic predictions about what content is most likely to keep a user engaged. Since these probabilities are derived from an immense number of data points, analogous to metaphorically "tossing coins" millions or billions of times, the predictions become highly reliable. For instance, when a recommendation system predicts that a specific piece of content or sequence of content is likely to keep a user on the platform longer, or to

---

<sup>82</sup>P Covington, J Adams and E Sargin, 'Deep Neural Networks for Youtube Recommendations', *Proceedings of the 10th ACM conference on recommender systems* (2016); B Smith, 'How TikTok Reads Your Mind' *The New York Times* (6 December 2021) <<https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html>> accessed 4 January 2025; D Guo et al, 'Deep Bayesian Bandits: Exploring in Online Personalized Recommendations', *Proceedings of the 14th ACM Conference on Recommender Systems* (2020).

<sup>83</sup>Z Chen et al, 'Data Scarcity in Recommendation Systems: A Survey' [2024] *ACM Transactions on Recommender Systems* <<https://dl.acm.org/doi/10.1145/3639063>> accessed 13 December 2024; Covington et al (n 82).

<sup>84</sup>*Ibid.* (n 82); Chen et al, (n 83); G Zhang et al, 'Scaling Law of Large Sequential Recommendation Models', *Proceedings of the 18th ACM Conference on Recommender Systems* (2024).

<sup>85</sup>OA Montesinos López, A Montesinos López and J Crossa, 'Fundamentals of Artificial Neural Networks and Deep Learning' in *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (Springer 2022).

<sup>86</sup>P Révész, *The Laws of Large Numbers*, vol 4 (Academic Press 2014); Zhang et al, (n 84); Covington et al (n 82).

<sup>87</sup>Révész (n 86).

<sup>88</sup>Covington et al (n 82).

click, subscribe, or share, these predictions carry very high confidence and small error margins.<sup>89</sup> Thus, if such a system determines that there is an 80% or 90% probability that users will engage longer when shown a specific video or series of videos, the observed user behaviour will likely align closely with this prediction. Further, since the system is designed to recommend only the content with the highest probability of engagement,<sup>90</sup> its predictions and users' actual behaviours tend to closely mirror each other. In practice, this means users end up staying on the platform longer, clicking, or engaging in ways that reflect the system's probabilistic calculations.

Scaling laws, on the other hand, describe the predictable improvement in model performance as both the model size (measured by the number of parameters in the DNN adjusted during training to minimise error and optimise predictions) and the dataset size (measured by the number of training examples) increase.<sup>91</sup> Unlike the law of large numbers, which focuses on the stabilisation of aggregated data trends, scaling laws capture the efficiency gains achieved by increasing the model's complexity and the dataset's scope. Larger models can process more complex, multi-dimensional relationships in the data and identify subtle, latent, contextual correlations that simpler models would miss. At the same time, larger datasets provide the model with a more diverse and representative set of examples, that is, examples that adequately reflect the diversity and variability of the real-world scenarios the model is expected to encounter, enabling it to generalise its predictions across broader contexts.

Here too, the same dynamic shapes what the recommendation systems behind platforms such as YouTube, TikTok, and Instagram can actually do. Because these platforms operate with models trained on billions of parameters and datasets of comparable magnitude,<sup>92</sup> they are able to identify subtle, cross-contextual patterns that smaller systems simply cannot detect: sequence-level viewing habits, fine-grained variations in dwell time, and latent similarities between users who appear unrelated on surface demographics.<sup>93</sup> The consequence is a qualitative rather than a merely quantitative shift. Scaling enables the system to place even a sparsely observed user within a probabilistic neighbourhood of millions of analogues and to draw reliable engagement inferences from that placement.<sup>94</sup> Accordingly, scaling laws do not simply make these systems larger; they alter what the systems can perceive and how confidently they can predict, even in cases where the data on any given user is thin.<sup>95</sup> In that sense, scale is one of the constitutive conditions of the compulsive character of platform engagement.

This is because when such systems 'learn', their learning process differs from following explicit, predefined instructions (eg, 'if A happens, then do B'). Instead, these systems, benefiting from principles like the law of large numbers, develop the capacity to identify generalisable patterns, logics, or rules that are sufficiently robust and reliable.<sup>96</sup> These learned principles enable the system to make accurate predictions even when it encounters new contexts, events, or user behaviours that were not explicitly represented in its training data.<sup>97</sup> This is akin to an experienced driver adapting to new roads, not by memorisation but by learned orientation.

When combined with the law of large numbers, the system's capabilities become even more refined. By training on vast datasets containing millions or billions of user interactions, the system reduces noise and random variations in its predictions and achieves a level of precision akin to the

<sup>89</sup>Guo et al, (n 82); Covington et al (n 82); Révész (n 86).

<sup>90</sup>Guo et al, (n 82); Covington et al (n 82); Smith (n 82).

<sup>91</sup>Zhang et al, (n 84); P Zivic, H Vazquez and J Sánchez, 'Scaling Sequential Recommendation Models with Transformers', *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2024).

<sup>92</sup>Covington et al (n 82); Chen et al, (n 83).

<sup>93</sup>Zivic et al (n 91); Zhang et al, (n 84).

<sup>94</sup>Zhang et al, (n 84).

<sup>95</sup>J-Y Song and B Suh, 'Data Augmentation Strategies for Improving Sequential Recommender Systems' [2022] arXiv preprint arXiv:2203.14037.

<sup>96</sup>S Vieira et al, 'Deep Neural Networks' in A Mechelli and S Vieira (eds), *Machine Learning* (Elsevier 2020).

<sup>97</sup>Zivic et al (n 91); Zhang et al, (n 84); Montesinos López et al (n 85).

skill set of a driver with hundreds of thousands of hours of experience. This allows the system to confidently apply its learned patterns and generalisations to new and unpredictable contexts so that its probabilistic predictions match closely real-world behaviours.

Significantly, it is important to note that such systems can achieve this level of accuracy even when data about individual users is sparse or incomplete because of advanced techniques designed to address missing information.<sup>98</sup> Synthetic data generation is one such technique, where artificial data points are created to mimic real-world patterns observed in existing datasets.<sup>99</sup> For example, if a new user has little interaction history, the system can simulate behaviours based on similar users' data to generate predictions. Similarly, if a new piece of content has no prior engagement, synthetic data helps the system infer how it might perform based on analogous content. Another important potential tool is the use of knowledge graphs, which link related data points to infer likely behaviours.<sup>100</sup> For example, if a user has watched videos in a particular genre but has not interacted with similar content, the system can connect these gaps by linking their preferences to broader categories or related attributes. This enables the algorithm to make reliable inferences about their preferences, even with minimal direct input. Finally, transfer learning allows systems to generalise insights gained from one context or group of users to another.<sup>101</sup> For instance, data from users with similar behaviours or demographics can inform predictions for a new user who shares similar characteristics.<sup>102</sup>

#### **A. Multi-armed bandits and algorithmic addiction: the feedback loops driving user engagement**

Once the model has been trained, it is deployed to generate recommendations in real time during user interactions. Here, recommendation systems use reinforcement learning, a technique where the system learns from the outcomes of its actions. For instance, if a user engages positively with recommended content (by watching, liking, or commenting in ways that meet probabilistic scores of the model), the system interprets this as a 'reward' and adjusts its future recommendations accordingly.<sup>103</sup> Conversely, if the user skips or ignores the content, this acts as a 'punishment', prompting the system to recalibrate its probabilities and engagement strategies. A specific approach used in this process is called the multi-armed bandit model. This idea originates from probability theory and is named after the analogy of a gambler faced with multiple slot machines (referred to as 'one-armed bandits'), each with an unknown probability of payout.<sup>104</sup> The gambler's challenge is to balance two competing goals: *exploiting* the slot machine that seems to give the best rewards based on past experience and *exploring* other machines that might yield even better rewards. This trade-off between exploitation (capitalising on known rewards) and exploration (searching for potentially better options) forms the core of the multi-armed bandit problem.<sup>105</sup>

In recommendation systems, content or recommendations effectively take the place of the slot machines, while the algorithm acts as the gambler attempting to learn which content maximizes user engagement.<sup>106</sup> Users provide indirect feedback (such as clicking on, liking, or ignoring content), which serves as the rewards or punishments that guide the system's learning. Each time a

<sup>98</sup>Chen et al, (n 83).

<sup>99</sup>*Ibid.*

<sup>100</sup>*Ibid.*

<sup>101</sup>Z Zeng et al, 'Knowledge Transfer via Pre-Training for Recommendation: A Review and Prospect' 4 (2021) *Frontiers in Big Data* 602071.

<sup>102</sup>Chen et al, (n 83).

<sup>103</sup>N Silva et al, 'Multi-Armed Bandits in Recommendation Systems: A Survey of the State-of-the-Art and Future Directions' 197 (2022) *Expert Systems with Applications* 116669.

<sup>104</sup>M Feng, 'A Analytical and Practical Insights into Multi-Armed Bandit Problems in Recommendation Systems' 94 (2024) *Highlights in Science, Engineering and Technology* 211.

<sup>105</sup>HL Kwa, JL Kit and R Bouffanais, 'Balancing Collective Exploration and Exploitation in Multi-Agent and Multi-Robot Systems: A Review' 8 (2022) *Frontiers in Robotics and AI* 771520.

<sup>106</sup>Silva et al, (n 103).

user interacts with the platform, the system experiments with different types of content to maximise engagement. For example, it might show the user a piece of content similar to what they have previously liked (exploitation) or test something new to gauge their reaction (exploration). Accordingly, exploration and exploitation are not separate because they are fundamentally part of the same decision-making process in recommendation systems. They do not choose to explore or exploit independently; they integrate both into every recommendation they make.<sup>107</sup> In that, every recommendation the system makes (whether it is exploiting known preferences or exploring new ones) provides feedback that informs future decisions: If the system exploits (eg, shows a user a favourite genre) and gets positive feedback, it reinforces its confidence in similar recommendations. If the system explores (eg, suggests a new genre) and gets positive feedback, it expands its understanding of the user's preferences, which informs future exploitation.<sup>108</sup> In this way, exploration feeds exploitation by providing new data, and exploitation strengthens confidence in existing knowledge. They are not separate actions but parts of a continuous learning loop.<sup>109</sup>

Accordingly, by responding to feedback, the system is effectively learning from the user what types of content are most likely to keep them engaged and, on the platform, longer. However, the system does not learn in isolation; it aggregates and generalises insights from many users.<sup>110</sup> This means that when one user engages with the platform, they are not just teaching the system how to engage or 'hook them'<sup>111</sup> individually; they are also contributing to the system's understanding of how to engage/hook users with similar patterns or characteristics. For example, if a user repeatedly engages with short, funny videos, the system might generalise this to prioritise such content for other users with similar behaviour or demographics. Because of this, users are simultaneously teaching the system and being influenced by it and the boundary between engagement (a behavioural outcome) and machine learning (the computational process driving it) becomes invisible because user behaviour is both the input (teaching the system) and the target outcome (increased engagement).<sup>112</sup>

### **B. From traits to patterns: the ontology of algorithmic addiction**

Thus, while the sense of addiction or compulsion users experience during platform engagement is shaped by a constellation of individual traits, mental health conditions, neurological responses, and broader social contexts, it is increasingly patterned and intensified by the structural logic of large-scale algorithmic systems, particularly those governed by scaling laws, the law of large numbers, reinforcement learning, and multi-armed bandits. These factors, while relevant in broader discussions of behavioural tendencies or susceptibility, are treated as statistical 'noise' within the computational logic of recommendation systems. This is because the systems rely on immense datasets and mathematical calculations to identify and act upon patterns that go beyond individual variability as such, focusing instead on aggregate trends that can be generalised across large populations. As a result, the compulsion to engage (whether to keep watching, clicking, subscribing, or interacting) is not rooted *solely* in the idiosyncrasies of a single user but emerges, also, from the aggregate weight of highly sophisticated data models and their ability to predict and influence behaviour at scale. These systems, through multi-armed bandits, relentlessly probe, test, and experiment with every interaction, pushing boundaries, pulling users into loops of engagement, blurring the lines between behaviour and system feedback, and creating a dynamic,

<sup>107</sup>Silva et al, (n 103).

<sup>108</sup>Silva et al, (n 103).

<sup>109</sup>Silva et al, (n 103).

<sup>110</sup>Silva et al, (n 103); Guo et al, (n 82); Y Su and M Chen, 'Nonlinear Bandits Exploration for Recommendations', *Proceedings of the 17th ACM Conference on Recommender Systems* (2023).

<sup>111</sup>Alter (n 61).

<sup>112</sup>Silva et al, (n 103).

ever-adapting force that draws users deeper into the platform's gravitational pull, all while continuously refining its strategies to keep them there.<sup>113</sup>

In effect, the feeling users describe as being 'pulled in' or 'hooked'<sup>114</sup> by platforms is a product, in part, of this gravitational force created by vast datasets, complex relationships between users and algorithms, and historical training data and reinforcement learning. As such, these systems do not explicitly target or, to use the language of the EU's Artificial Intelligence Act (AIA, 2024), 'exploit' individuals 'due to' their unique traits or vulnerability because their of age or disability (AIA, Article 5); rather, they use generalised yet highly accurate predictions to create the sensation of personal relevance. This phenomenon has been likened, as the *New York Times* once described it, to a system 'reading the user's mind'.<sup>115</sup> It manifests as a strange, hard-to-describe sensation where the user feels an internal sense of agency (believing they are making autonomous choices by clicking, swiping, or watching), yet this is paired with the unsettling experience of the platform seemingly anticipating their actions, almost as if it knows what they want before they do.<sup>116</sup> As a result, the addictive pull of these systems is less about the specifics of an individual user and more about the immense computational power that aggregates, analyses, and applies patterns across billions of interactions. This enables platforms to create an experience where the user's actions seem both voluntary *and* anticipated. It is this combination of precise prediction and perceived autonomy that makes the sensation so directive, even if users are unable to pinpoint exactly why they feel driven to continue engaging.

## 7. Platform addiction as a *relatant*: intra-action, diffraction, and emergent entanglements

However, this does not mean I wish to minimise or negate the role of other variables shaping platform addiction. Instead, my contention is that, drawing from Karen Barad's concepts of intra-action and diffraction,<sup>117</sup> platform addiction can be understood not as a 'thing' located in a specific feature, mechanism, or even within the user, but, rather, as what I call a *relatant*: an emergent effect, unfolding, or becoming that arises through the relational and diffractive processes among the many variables described in the clinical/psychiatric, neurological, Skinnerian, psychosocial, and algorithmic/machine learning models.

What I mean by calling platform addiction a *relatant* is this: it is not a pre-given 'thing' that sits waiting to be detected or triggered by a single feature. It is a phenomenon that comes into being through relations among bodies, brains, families, devices, interfaces, business models, and classifications such as 'risk' or 'harm'. In Barad's vocabulary, these relations are *intra-actions* rather than interactions.<sup>118</sup> The distinction matters. Interaction assumes separate entities that later meet; intra-action holds that the very identities of the entities ('user', 'feature', 'harm', 'vulnerability') are themselves formed through the ongoing process. On this view, the clinical, neurological, Skinnerian, psychosocial, and machine-learning accounts presented not rival explanations of a fixed object; they are parts of the apparatus through which 'addiction' becomes real, measurable, and actionable in particular ways.<sup>119</sup>

<sup>113</sup>W Mbioh, 'Beyond Possessive Agency: TikTok, YouTube, and the Inadequacies of GDPR, OSA, DSA, and AIA' 15 (2025) *Oñati Socio-Legal Series* 1022.

<sup>114</sup>Alter (n 61).

<sup>115</sup>Smith (n 82).

<sup>116</sup>*Ibid.*

<sup>117</sup>Barad, *Meeting the Universe Halfway* (n 1); K Barad, 'Diffracting Diffraction: Cutting Together-Apart' 20 (2014) *Parallax* 168.

<sup>118</sup>Barad, *Meeting the Universe Halfway* (n 1).

<sup>119</sup>NS Mauthner, 'K Barad's Posthumanist Relational Ontology: An Intra-Active Approach to Theorising and Studying Family Practices' 10 (2021) *Families, Relationships and Societies* 33; M Dillen, 'Time to (Intra-) Act'; Ștefania Chihaia, 'New Materialism in Sociological Thinking: Barad's Quantum View On The World's Matters' 32 (2021) *Revista Romana de*

Two further Baradian ideas help to make this concrete. First, *positionality*: there is no view from nowhere. Researchers, designers, clinicians, parents, teachers, and regulators stand somewhere in the arrangement, and their methods, measures, and expectations help to bring certain patterns to the fore while pushing others to the edges. Second, *agential cuts*: to analyse or to regulate, we have to draw boundaries: between ‘normal’ and ‘compulsive’, ‘feature’ and ‘context’, ‘minor’ and ‘adult’, ‘serious’ and ‘not yet serious’. Those cuts are not mere labels applied to a finished object; they help produce the very object they describe. They also carry ethical weight because they decide what counts as harm, who is seen as at risk, and which forms of evidence are admitted.<sup>120</sup>

Thus, much like socio-cultural accounts of addiction in the work of, for example, Nancy D. Campbell, Elizabeth Ettorre, and, more recently, Jaeyoon Park, what is recognised as ‘addiction’, and whose conduct, and what forms of conduct, are pathologised or normalised, is produced within social meanings, the political economies of race, gender, class, and place, the organisation of pharmacological markets and practices, and the institutional routines through which care, punishment, and profit are arranged, including clinical taxonomies, welfare eligibility rules, policing priorities, and regulatory design.<sup>121</sup> In other words, ‘addiction’ is not a neutral observation but an outcome of how societies allocate attention, blame, and resources: the same pattern of behaviour can be read as pathology, vice, coping, or commerce depending on where it appears, who exhibits it, and which institutions, law among them, are doing the reading, since legal definitions, thresholds, and evidential standards crystallise these judgements and decide who becomes visible to intervention and on what terms.

*Diffraction* to Barad names a way of reading these arrangements or agential cuts: following differences that make a difference.<sup>122</sup> Rather than asking which single cause ‘really’ explains compulsion, a diffractive approach asks how particular cuts organise attention and consequence: which relations are amplified; which are made peripheral; which groups are centred; which are backgrounded.<sup>123</sup> Consequently, it is a reading of effects and responsibilities: how a boundary fixes what is to be counted, what is to be ignored, who must change practice, who must endure, and how evidence is to be gathered. It is, in Barad’s terms, an *ethico-onto-epistemic stance*: because knowing and doing are entangled with what is, the way we carve the phenomenon *is already a participation in it*.<sup>124</sup> The ethical question, then, is not only whether a cut is accurate but *what it does*: what patterns of harm it stabilises, what forms of care it enables, which bodies it exposes, and which it shelters.

Seen this way, ‘addiction’ is not reducible to a linear trigger–response chain. It is an effect of many forces moving together: adolescent neurodevelopment and mood, family stability or stress, reinforcement schedules and notification cadence, cohort expectations and public metrics, data-driven personalisation, and commercial incentives. Because nothing stands outside these relations, law does not either. Without taking a position on any statute yet, the simple point is this: whenever

---

*Sociologie* 105; MI Bilbao-Nieva and A Meyer, ‘Contributions of Barad’s New Materialism to Well-Being Research’. 29 (2024) Qualitative Report.

<sup>120</sup>K Barad, ‘Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter’ 28 (2003) *Signs: Journal of Women in Culture and Society* 801; G Hollin et al, ‘(Dis) Entangling Barad: Materialisms and Ethics’ 47 (2017) *Social Studies of Science* 918; Mauthner (n 119).

<sup>121</sup>J Park, *Addiction Becomes Normal: On the Late-Modern American Subject* (University of Chicago Press 2024); ND Campbell and E Ettorre, *Gendering Addiction: The Politics of Drug Treatment in a Neurochemical World* (Palgrave Macmillan 2011).

<sup>122</sup>Barad, ‘Diffracting Diffraction’ (n 117).

<sup>123</sup>K Merten, *Diffractive Reading: New Materialism, Theory, Critique* (Rowman & Littlefield International 2021); K Murriss and V Bozalek, ‘Diffraction and Response-Able Reading of Texts: The Relational Ontologies of Barad and Deleuze’ 32 (2019) *International Journal of Qualitative Studies in Education* 872; A Schrader, ‘Diffractive Experiments in the Ethics of Mattering’ 28 (2009) *Subjectivity* 349.

<sup>124</sup>Harris (n 2); K Barad, ‘Queer Causation and the Ethics of Mattering’ in N Giffney and MJ Hird (eds), *Queering the Non/Human* (Routledge 2016); M Juelskjær and N Schwennesen, ‘Intra-Active Entanglements: An Interview with K Barad’ 21 (2012) *Kvinder, Koen og Forskning* 10.

law defines terms, sets thresholds, or specifies evidence, it participates in drawing agential cuts. Those cuts will never be neutral. They will decide whose experience is legible, which harms surface as actionable, and how responsibility is shared. That is the frame the next sections build on.

## 8. Core critique of the Digital Services Act (DSA) on platform addiction

What follows from this is not a claim that platform addiction is something a single statute such as the DSA could ‘fix’. The account here has insisted that compulsive use is co-produced by many forces at once as a *relatant*: interface patterns are only one strand alongside neurodevelopment, psychosocial conditions, and the probabilistic pull of large-scale recommender systems. It would be unreasonable, accordingly, to expect the DSA, on its own, to rewire adolescent brain development, remake the socio-economic conditions that heighten vulnerability, or resolve the muddled agency that emerges in interaction with complex algorithms.<sup>125</sup> At the same time, on a Baradian reading, the DSA is not outside that field. Law is a material–discursive apparatus that participates in the very intra-actions it seeks to govern or intervene in. It enacts agential cuts by fixing thresholds, predicates, and categories, deciding, for example, when ‘addiction’ is cognisable, what counts as ‘serious’, and whether a design element is treated as a proximate cause or a background condition. Through those cuts, law reshapes incentives and defaults, configures compliance practices (risk assessments, mitigations, dashboards), and diffracts attention toward certain patterns of harm while rendering others peripheral.

### A. Scope, sufficiency, and constitutive effect: why ‘insufficient to solve’ does not mean ‘causally inert’

These two claims, ie, the proposition that no single statute can ‘fix’ a relational phenomenon and the idea that law nonetheless shapes the field through its agential cuts, are not in tension. The first claim is about *scope and sufficiency*: platform addiction is produced by many intra-actions (developmental, psychosocial, economic, and computational), so no single statutory instrument is sufficient to resolve it. The second claim is about *constitutive effect*; legal rules help make the world they regulate by shaping defaults, incentives, evidentiary standards, and what counts as a cognisable harm. There is no contradiction because ‘insufficient to solve’ does not entail ‘causally inert’. A statute can be unable to remediate the whole while still reconfiguring parts of the apparatus in ways that amplify, dampen, or redirect how the phenomenon appears and is addressed. Put differently: the ontology of the problem exceeds law’s reach, yet law remains one of the material practices through which the problem is continuously organised and made legible. In intra-active terms, there is no external vantage point from which law governs a pre-given object. ‘Platform addiction’ and ‘the DSA’ co-emerge within a shared apparatus; each regulatory move is an agential cut that reconfigures what becomes visible, measurable, and actionable while consigning other relations to the penumbra. On a diffractive reading, the two claims sit together: the phenomenon exceeds any single instrument, yet every intervention alters the interference pattern, redistributing attention, accountability, and risk.

Law is *simultaneously* participant and partitioner: inside as a practice that helps enact the field, and ‘outside’ only in the narrow sense that it names, from within, the boundary it draws for its own operation. In this sense, the apparent paradox (ie, law cannot solve the phenomenon yet still reshapes it) resolves if we think diffractively: change the aperture (the regulatory cut) and the interference pattern shifts. The underlying waves (neurodevelopment, social context, interface flow, recommender pacing) are the same, but different cuts redistribute what shows up as a bright band of actionability and what recedes into darkness. Insufficiency and effect are therefore

<sup>125</sup>Mbioh (n 113).

simultaneous: the DSA cannot remake the whole, yet its cut redirects incentives, evidence, and responsibility, altering who is protected, which harms count, and where residual risk settles.

### **B. Alternative cuts: lower thresholds and the governance of everyday retention mechanics**

Against that backdrop, the specific critique of the DSA is not that it simplifies (for, realist accounts rightly note that law must simplify to act) but that it has adopted a particular simplification or agential cut: it renders harm cognisable only where an interface feature ‘causes’ or ‘stimulates’ addiction and the outcome is ‘serious’. As suggested by the European Commission’s recent enforcement posture, this largely translates to prioritising high-amplitude, specifically attributable inducement loops: cash- or points-for-engagement programmes (watch/follow/invite-to-earn), referral bounties, loyalty schemes with cash-equivalent redemption, and gambling-adjacent mechanics such as spins or loot-box draws, alongside a strong emphasis on procedural compliance under Articles 34–5 (documented risk assessments and mitigation plans), with everyday retention architecture addressed, if at all, via user-option mitigations rather than mandatory redesign. That is *one* possible cut, not the only one.

The European Parliament’s resolution on addictive design suggests there are other ways to draw the boundary:<sup>126</sup> one could move away from the vocabulary of ‘addiction’, lower the intervention threshold, and bring into scope the everyday techniques of behavioural modification (endless feeds, default autoplay, public like counters), even where they do not individually produce clinically severe outcomes.<sup>127</sup> These are policy choices, not inevitabilities. The point is not to demand the impossible from law, but to be explicit about which cut has been chosen, what it renders visible and invisible, and how different cuts, now under active consideration, could better reorient legal duties towards how harm is intra-actively produced.

Enacting those changes would recut the phenomenon. Ontologically, harm would cease to be a rare, case-by-case pathology tied to a single feature and ‘serious’ outcome; it would be treated as a patterned, cumulative process that emerges from configurations of design and pacing at cohort level. Causation would shift from proximate triggers (‘cause/stimulate’) to contribution and foreseeability, so evidence could rest on regularities over time (sleep displacement after late-night autoplay, denser session chains with notification cadence), rather than switch-off counterfactuals for one lever. What becomes visible, measurable, and actionable is therefore the everyday retention architecture itself: recommender tempo, infinite feeds, default autoplay, public counters, and their joint effects under default settings.

Distributionally, the burden would move away from individual self-help (toggles, literacy, parental vigilance) towards platform-side duties to test, redesign, and prove cohort safety under defaults. Residual risk would be reallocated: younger users and other susceptible groups would carry less of it; platforms and, indirectly, advertisers and growth teams would carry more through friction, slower funnels, and outcome-based audits. Creators whose reach depends on uninterrupted flow could face slower audience accumulation; conversely, users would gain time, sleep, and mood stability that are presently externalised. Compliance resources would pivot from paperwork about procedures to measurements of population outcomes, with enforcement keyed to longitudinal metrics and stepped roll-outs rather than one-off feature takedowns. So, the cut would reprice attention: it would reward designs that keep foreseeable detriment below defined thresholds and penalise architectures that rely on diffuse, always-on prompts, altering who is protected, which harms count, and where the costs of prevention sit. In diffractive terms, the aperture changes; the bright band of actionability moves from inducements to defaults, and with it the distribution of accountability and risk.

<sup>126</sup>European Parliament (n 36).

<sup>127</sup>*Ibid.*

### C. Distributional effects of gating: identity demands, exclusion, and displacement

Within that architecture, service-age gating could be a possible structural control, much like the Australian minimum-age model under the Online Safety Amendment (Social Media Minimum Age) Act 2024. By ‘service-gating’ I mean a rule that conditions access to a whole service on age, rather than tweaking particular features: below a statutory minimum (eg, under-16s) platforms must block access or confine users to a restricted mode, typically backed by some form of age assurance and penalties for systemic non-compliance. In the Australian formulation, this includes age checks at sign-up and, where risk indicators warrant, periodic checks during use.

Ontologically, service-gating relocates the object of concern from ‘harmful features’ to ‘illicit presence’. The primary wrong becomes the existence of a minor on a service, not the everyday attention mechanics themselves. Evidence shifts accordingly: from cohort-level detriment under default designs to compliance metrics about age detection, gating efficacy, and leakage rates. More precisely, harm is redrawn as a status condition presumed from cohort membership: the operative predicate becomes that the class of services presents an unacceptable *ex ante* risk for this age group. Causation is recast as prophylactic and time-bound, with vulnerability inferred from neurodevelopment rather than from proving that a particular interface produced ‘serious’ effects in a given child.

Distributionally, the burden moves towards identity, documentation, and gatekeeping infrastructures: families without easy ID face higher friction; false positives can exclude older teens who benefit from supportive communities (ie, LGBTQ+, neurodivergent, disabled, or otherwise marginalised); false negatives can funnel younger users toward less regulated or offshore services.<sup>128</sup> Privacy risk concentrates around age assurance vendors; accountability drifts from redesigning attention flows to policing entry points. Platforms assume front-loaded obligations (building and maintaining checks, handling appeals, and funding privacy safeguards) and may absorb smaller youth audiences; identity-verification vendors gain bargaining power. Advertisers and creators lose access to younger cohorts or must adapt to teen-specific spaces. At the same time, some exposure (and therefore some harms) will fall, because a subset of minors will be blocked from high-intensity feeds; yet other harms are re-routed, appearing as displacement to private messaging, identity borrowing, or migration to dark-pattern-heavy alternatives beyond domestic jurisdiction. So, the cut no longer reprices attention within the service; it redefines who may be an addressee of attention at all and shifts the costs of control from user-side toggles to platform-side eligibility and identity infrastructure. In Baradian terms, the cut changes the interference pattern: the same underlying forces remain, but the regulatory aperture now renders ‘being underage on the service’ brightly visible while pushing the ordinary retention architecture further into the penumbra of adult-facing design.

## 9. Conclusion: where does this leave us?

However, where does this leave us? The core point is simple. The DSA’s text approaches ‘addiction’ through a narrow lens, but the models surveyed (and a Baradian account of intra-action) show that compulsive use is a complex, relational phenomenon. We should begin by acknowledging that complexity. From there, we must hold two ideas at once. First, law necessarily simplifies: it has to draw lines to act. Second, the DSA’s particular line is only one possible choice.

<sup>128</sup>A Cavalcante, ‘Tumbling into Queer Utopias and Vortexes: Experiences of LGBTQ Social Media Users on Tumblr’ in D Romesburg (ed), *LGBTQ Culture* (Routledge 2021); A Bates, T Hobman and BT Bell, ‘“Let Me Do What I Please with It . . . Don’t Decide My Identity for Me”: LGBTQ+ Youth Experiences of Social Media in Narrative Identity Development’ 35 (2020) *Journal of Adolescent Research* 51; T Vouglanis and A Drigas, ‘The Internet Addiction and the Impact on the Cognitive, Psychological and Social Side of People’s Personality with Disabilities’ 35 (2022) *Technium Social Sciences Journal* 93; SI Fayzhia and Y Setiawati, ‘The Interplay of ADHD, Social Media Usage, and Dopamine Receptors in Adolescents: A Literature Review’ 23 (2024) *World Journal of Advanced Research and Reviews* 2698.

Its way of cutting the problem (through ‘cause/stimulate’ plus ‘serious’) is not the sole route available. And while no single statute can resolve the full constellation of psychological, social and algorithmic drivers, the choice of cut still matters. It shapes what is counted as harm, how evidence is gathered, which mitigations are expected, and who bears the residual risk. In other words, the DSA’s framing is part of the ethical entanglement: it has consequences, including distributional ones.

Seen in this light, several things can be true at the same time. The DSA is only one element within a wider apparatus; it cannot ‘solve’ platform addiction. Yet it also participates in producing what becomes visible and actionable, and so can reproduce, dampen, or intensify existing dynamics. Recognising this does not require rejecting the statute. It requires naming its present cut, acknowledging what it leaves out, and remaining open to other, equally lawful cuts that would bring everyday attention mechanics into scope. A diffractive stance lets us hold these tensions coherently: law simplifies and must act; its simplification is contingent and revisable; and different simplifications will steer accountability and risk along different paths. That is the claim I am making, and the invitation to think with the consequences of how we choose to draw the line.

**Competing interests.** The author has no conflicts of interest to declare.