Real-Time Hyper-Personalized Generative AI Should Be Regulated to Prevent the Rise of "Digital Heroin"

Raad Khraishi* **Cristovao Iglesias** **Devesh Batra** **Peter Gostev**

Giulio Pelosio **Ramin Okhrati** **Greig A. Cowan**

Abstract

This position paper argues that real-time generative AI has the potential to become the next wave of addictive digital media, creating a new class of digital content akin to "digital heroin" with severe implications for mental health and youth development. By shortening the content-generation feedback loop to mere seconds, these advanced models will soon be able to hyper-personalize outputs on the fly. When paired with misaligned incentives (e.g., maximizing user engagement), this will fuel unprecedented compulsive consumption patterns with far-reaching consequences for mental health, cognitive development, and social stability. Drawing on interdisciplinary research, from clinical observations of social media addiction to neuroscientific studies of dopamine-driven feedback, we illustrate how real-time tailored content generation may erode user autonomy, foment emotional distress, and disproportionately endanger vulnerable groups, such as adolescents. Due to the rapid advancement of generative AI and its potential to induce severe addiction-like effects, we call for strong government oversight akin to existing controls on addictive substances, particularly for minors. We further urge the machine learning community to act proactively by establishing robust design guidelines, collaborating with public health experts, and supporting targeted policy measures to ensure responsible and ethical deployment, rather than paving the way for another wave of unregulated digital dependence.

1 Introduction

Recent breakthroughs in generative artificial intelligence (AI), from large language models (LLMs) to advanced diffusion-based image and video generators, are transforming digital content creation at a rapid pace [1, 2, 3]. These innovations enable near-instant generation of highly personalized text, images, and videos, shifting the user experience from a largely static, consumption-based interaction to an interactive, on-demand loop of AI-driven creativity [4, 5, 6]. While such progress holds promise for productivity, education, and entertainment, there is an emerging concern that real-time generative AI could intensify addictive usage patterns, especially among vulnerable populations such as adolescents [7, 8, 9].

Previous waves of digital technology have already demonstrated the capacity to induce compulsive behavior and dependency, clinically recognized as behavioral addiction [10, 11]. Social media platforms, for instance, incorporate AI-driven recommendation systems to maximize engagement, but these same strategies have been implicated in excessive screen time, mental distress, and even neurobiological changes comparable to those seen in substance abuse [7, 12, 13]. With the advent of large-scale generative models, this dynamic may escalate further: users may soon receive continuous

*This paper was undertaken in the authors' personal capacities. The views expressed are the authors' own and do not necessarily reflect those of any employer or institution. Correspondence to raad.khraishi@ucl.ac.uk.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Position Paper Track.

streams of personalized content fine-tuned in real time to maintain engagement through techniques such as reinforcement learning [5, 14, 15].

Crucially, when these systems are coupled with misaligned objectives like maximizing user dwell time or ad revenue, they become adept at "hacking" human reward pathways [16, 17, 18]. By systematically exploiting psychological triggers (e.g., intermittent rewards, novelty, and social validation) reinforcement learning algorithms can fuel compulsive overuse, locking users into endless engagement loops with content specifically tailored to their individual vulnerabilities [19, 20]. In essence, the platform's optimization goal for growth or revenue can directly conflict with user well-being, resulting in powerful AI systems capable of producing content so addictive it becomes akin to "digital heroin" – content that can hijack the brain's reward system in a manner comparable to narcotics (as evidenced by dopamine surges and neuroadaptation observed in social media use) [16, 17, 18, 21, 22].

We posit that real-time generative AI platforms pose an imminent risk of unprecedented digital addiction, analogous to substance abuse, and require immediate regulatory interventions to safeguard public mental health.

By placing this discussion within the heart of the machine learning community, we underscore the urgency for researchers and practitioners to collaborate with public health experts, policymakers, and other stakeholders. This paper seeks to catalyze a proactive approach, urging the AI community and technology companies to adopt ethical design principles, advance transparent auditing methods, and champion policy measures that can prevent real-time hyper-personalized generative AI from evolving into an unchecked vector for digital addiction.

2 From Algorithmic Content Recommendations to Real-Time Personalized Content Generation

A novel form of real-time dynamically generated AI content, designed to maximize user engagement, is rapidly emerging. Compared to traditional human-generated content, generative AI will significantly accelerate content production. When integrated with real-time personalization, these methods extend beyond mere recommendations of existing content, enabling on-the-fly generation specifically tuned for maximum engagement. We hypothesize that this personalized, dynamically generated AI content will surpass current addictive patterns observed with conventional short-form social media content and recommendation systems. In this section, we elaborate on each of these core concepts in greater detail.

2.1 Addictive Behavior in Existing Human-Generated Content Recommendations

Platforms such as TikTok (via its "For You" page), YouTube (via Shorts), and Instagram (via Reels) rely on endless streams of short-form content combined with AI-driven recommendation engines to maximize user engagement in real time [23]. This design has proven highly effective in capturing attention and prolonging screen time. For example, the average TikTok user now spends approximately 95 minutes per day on the app, more than any other social network [24]. Modern recommendation algorithms, powered by deep reinforcement learning, continuously personalize the feed to each user, updating recommendations with every swipe or tap to serve up the most engaging next video [19, 25, 26]. These systems preferentially surface novel or emotionally charged content that elicits strong reactions, thereby reinforcing compulsive usage patterns [27, 28].

Excessive social media use has become a prevalent behavioral problem, particularly among adolescents [29], in part due to such engagement-optimizing tactics [30]. Major technology companies have been reported to intentionally cultivate these addictive behaviors through "dark patterns" and personalized recommender systems as a means of maintaining user attention and market dominance [31, 32]. Notably, the introduction of short-form video feeds has supercharged these trends: Instagram Reels, for instance, experience 22% more interaction than standard video posts, illustrating how short, algorithmically curated clips can dramatically spike user engagement [33, 34].

Research has shown that heavy social media use can alter brain structure and function, including reductions in gray matter in areas responsible for impulse control and decision-making [13] and changes in the prefrontal cortex and limbic system [35]. Meta-analytic findings confirm consistent structural differences, especially diminished gray matter in reward and self-control regions, among

individuals with problematic Internet use [36]. Chronic use also disrupts dopamine-based reward pathways, mirroring substance dependence [17], and may weaken executive functions [7]. Adolescents with high usage report elevated depression and anxiety [37], with risk of depression increasing by 13% per additional hour of daily use [37]. Teens exceeding three hours per day have about twice the likelihood of developing depression or anxiety [9].

The addictiveness of social media content is rooted in fundamental psychological reward mechanisms [10, 20, 17, 7]. Platforms often deliver rewards on a variable schedule: as users scroll through an algorithmically curated feed, they encounter an unpredictable mix of mundane posts and highly rewarding content, a pattern known to strongly reinforce repetitive behavior (akin to a slot machine effect) [38]. The continual influx of novel information further stimulates the brain's reward system, studies show that acquiring new, unexpected stimuli can activate neural reward pathways much like receiving tangible rewards [16]. These combined factors trigger surges of dopamine during social media use, conditioning the brain to crave continued engagement in a manner similar to substance addictions [21].

Although already linked to addictive behavior, human-created content remains constrained by production cost, speed, and limited personalization; however, AI-generated content could soon remove these limits, potentially intensifying addictive use.

2.2 Advances in AI Content Generation

In a little more than two years, diffusion-based video-generation models have progressed from rudimentary, low-fidelity clips to outputs that verge on professional animation. Figure 1 juxtaposes Google's Imagen Video 1 (2022) [39] with Veo 2 (2024) [40] for the same prompt with the recent release notably exhibiting major gains in motion coherence, lighting realism, and material detail. Similar advances have been seen from other model providers. For example, OpenAI's Sora text-to-video system, unveiled in December 2024, can render minute-long 1080p clips with spatial–temporal consistency [6]. Runway's Gen-4 model (March 2025) adds browser-based 24 fps generation with reference-driven character consistency [41], while Midjourney's v7 release (April 2025) introduces style-locking, in-painting, and a reduced prompt-to-image latency [42]. More recently, Google released Veo 3 (May 2025) as the first multi-modal diffusion-based model to introduce native audio generation in combination with video, significantly enhancing versatility and viewer appeal by synthesizing natural-sounding speech, immersive background noises, and audio [43].

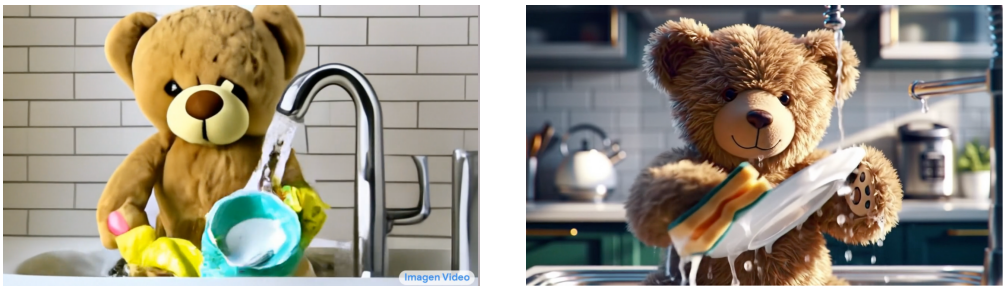

Figure 1: Visual comparison of two models from the same provider illustrating video generation quality improvements over two years, featuring the prompt "A teddy bear washing the dishes," by Google Imagen Video 1 (Oct 2022, left) [39] and Google Veo 2 (Dec 2024, right) [40].

A pivotal milestone for generative AI video is the point at which footage can be produced in real time at negligible cost. Recent hardware advances are already pushing the field toward this threshold. In Table 1, we highlight that using the recent HunyuanVideo [44] model, the step from Nvidia's 2020 A100 GPU to the 2025 GB200 GPU cuts render time by 16x (5.9 min $\rightarrow$ 22 seconds) and lowers per-second cost by 4.5x ($0.27 $\rightarrow$ $0.06). Algorithmic progress will further magnify these gains. For example, SF-V [45] compresses Stable Video Diffusion into a single U-Net pass with a 23x inference speed-up, and T2V-Turbo [46] requires only four sampling steps yet tops the VBench leaderboard [47], delivering a >10x acceleration at equal or higher visual quality than a DDIM baseline.

AI-generated content has already begun to appear heavily on social media platforms. For example, Meta's first disclosure of its "AI Info" label counted over 220 million Reels and 140 million posts

Table 1: Runtime and dollar cost to render a 1-second, 720p, 50-step clip with the HunyuanVideo [44] on three GPU generations assuming an A100 for their single-GPU benchmark results with estimates for H100 and GB200 extrapolated. Hourly cloud rates for GPUs sourced from CoreWeave [48].

GPU (year)	**Cloud rate** [$ / hr]	**Runtime** [s]	**Cost / clip**	**Cost vs A100**
A100 80 GB (2020)	$2.70	354	$0.27	1×
H100 80 GB (2022)	$6.16	55	$0.09	3× cheaper
GB200 NVL2 (2025)	$10.50	22	$0.06	4.5× cheaper

on Instagram carrying the tag in October 2024 attracting more than a trillion views [49]. Similarly, TikTok reported that more than 37 million creators had already used its "AI-generated" label since the feature rolled out in September 2023 [50].

Further improvements in quality, cost, and speed indicate that real-time content generation will soon be within reach. Combined with personalization these advances will soon erase the boundary between recommendation and creation, super-charging the reward loop described in Section 2.1.

2.3 Improvements in Content Personalization

Personalization has long been at the heart of digital platforms, but the focus is rapidly shifting from curating pre-existing content to generating customized content for each user. Early-stage recommender systems typically relied on static matrix factorization techniques that predicted user-item interactions based on sparse historical data [51]. Over the past few years, however, these comparatively simple methods have been eclipsed by deep RL approaches that treat each user action (e.g., clicks, swipes) as states in a Markov Decision Process [52]. With every user interaction, the RL agent updates its policy to optimize multi-step engagement metrics such as dwell time or retention [26, 19, 25]. However, a key inflection point will be the integration of generative models, such as LLMs and diffusion-based architectures, with personalization. While older recommendation systems merely ranked or selected existing content to display, emerging frameworks now create new videos, images, or text snippets aligned with each user's preferences [5, 53].

Illustrating this shift, recent work shows that a lightweight preference embedding distilled from just a few pair-wise comparisons can already steer a diffusion decoder toward the imagery a user consistently "likes" [54]. Complementary approaches apply direct preference optimization to video backbones, blending aesthetic and narrative rewards, so the entire motion sequence follows the viewer's preferred style and pacing [55]. Fine-grained reward models such as VisionReward further decompose human feedback into factors like motion stability and semantic fidelity, giving recommender stacks a structured objective for online tuning of their generative components [56]. Leveraging techniques like Personalized RLHF, platforms can compress recent interactions into compact user embeddings [5]. Rather than fully retraining an LLM or video diffusion model, low-rank adapters may also be used to modulate model parameters on the fly, enabling fine-grained personalization at scale [53, 4, 5].

Recent breakthroughs in text-to-video and image-generation latencies mean that content will soon be served up (or altered) nearly instantaneously [40, 6]. For example, as discussed previously, large video-generation models such as HunyuanVideo (see Table 1) are pushing inference times close to real-time, thereby enabling on-demand, high-fidelity media [44]. Integrating these generative engines with user embeddings creates a tight loop between user feedback signals, such as clicks, swipe velocity, or facial expressions, and subsequent model outputs [57, 58]. Thus, after every micro-gesture or moment of user hesitation, the system refines its internal state representation and regenerates content to elicit stronger engagement. This adaptation is further bolstered by surging context-window capacities in modern LLMs, which have ballooned from a few thousand tokens to over one million tokens in cutting-edge systems [59, 60, 61]. Parallel progress is now visible on the video side as well: long-context tuning and cache-augmented autoregressive approaches let diffusion and transformer generators maintain narrative coherence across longer duration clips [62, 63]. This expansion allows platforms to store and recall extensive user histories on the fly, including nuanced engagement patterns, mood markers, or emotional states. By tapping into such historical data, generative models can tune their outputs to reflect current user interests, thus creating an always-evolving stream of uniquely engaging content.

Platforms are already moving in this direction. ByteDance's Monolith embedding system updates user and item vectors in real time to rank TikTok's feed [64]. On the content side, the company's Symphony Creative Studio launched in 2024 with a text-to-video tool for advertisers, signaling that AI-generated clips can be injected directly into the same ranking pipeline [65]. YouTube's Dream Screen feature, announced in 2024, uses DeepMind's Veo model to create six-second video backgrounds from a simple text prompt, making generative video available inside the Shorts workflow [66]. In a recent patent, Meta provides the clearest integration with an ad system where a large-language model rewrites creative on the fly for each viewer, after which Meta's Horizon RL stack selects the best variant [67, 68]. These deployments illustrate a shift from selecting existing items to synthesizing fresh tailored content that is still optimized for the same engagement objective, tightening the feedback loop that drives habitual use.

By merging content generation with deep user personalization, digital platforms are poised to deliver an endless cycle of hyper-relevant stimuli. Longitudinal research has already shown a strong link between heavy personalized-feed exposure driven by screen-time maximizing algorithms and weakened impulse control as well elevated depression risk [11]. As such, these real-time generative systems coupled with objective-maximizing personalization are capable of satisfying core criteria for fostering addictive use: they deliver an immediate stream of novel content (rapid reward), tailor every snippet to the user's unique state (personalized salience), and may inject unpredictability in when or how the user is "rewarded" (variable reinforcement) [69]. Given the "always-learning" nature of these RL-driven personalization engines, each user interaction fuels further optimization, perpetually honing the model's ability to capture and hold attention. When aligned with platform objectives such as maximizing session length or ad revenue, this cycle can readily amplify compulsive engagement. By continuously probing a user's micro-signals (e.g., gaze tracking, pause durations, emotional cues) the model can recognize and exploit vulnerabilities, effectively "hacking" the user's reward pathways [16]. This mirrors computational accounts of drug addiction in which spiraling dopaminergic feedback progressively hands control from deliberative to habitual circuits [70].

In summary, personalization is no longer confined to selecting which posts a user sees next from a library of content. Instead, generative algorithms are beginning to synthesize custom-tailored content at low latency. While this progression undoubtedly enriches user experiences in many respects, it also sets the stage for an unprecedented level of behavioral targeting. If left unchecked, real-time hyper-personalization with misaligned objectives (e.g., maximizing watch time) could become the linchpin of "digital heroin," a potent engine for capturing and holding human attention with potentially serious consequences which we turn to in the next section.

3 Societal and Ethical Implications

By engaging users' reward circuits more effectively than previous "infinite-scroll" platforms, hyper-personalized generative content could transform casual interactions into compulsive usage, maximizing time-on-platform at the expense of user well-being [7, 71]. In this environment, digital engagement risks becoming a form of "digital heroin", reshaping attention spans, eroding mental resilience, and disrupting real-world social connections [8, 72]. Unlike physical substances, these content streams would operate continuously without any natural brakes, raising the specter of a global public health crisis. Left unregulated, such systems could stunt cognitive development (especially among minors), exacerbate existing social inequities, undercut workplace productivity, and ultimately impose severe societal costs [73, 74, 75, 7]. In this section, we explore these risks in more detail.

Public Health Fallout. Addictive AI-generated content would harm public health on three fronts: deteriorating mental health, declining physical health, and eroding social well-being. Clinical studies already link excessive social media use to attentional fragmentation [8, 76], increased anxiety, depression, and ADHD-like symptoms [77, 78, 79], chronic sleep disruption due to blue-light exposure [80], and heightened rates of loneliness and social isolation [81, 72]. With real-time generative AI, as these platforms further shorten feedback loops toward instant gratification, mental health burdens may surge dramatically, manifesting as longer sessions (tolerance), irritability and anxiety when offline (withdrawal), and loss of control, directly paralleling diagnostic behavioral addiction criteria [82]. Beyond psychiatric impacts, this addiction could significantly amplify physical health challenges, with prolonged sedentary screen-time behaviors precipitating obesity [83, 84], cardiovascular strain [85], and musculoskeletal disorders [86]. Social health may likewise deteriorate:

addiction to pervasive AI-curated content threatens to replace meaningful face-to-face relationships with isolated digital interactions [72], eroding communal bonds and dulling collective empathy through relentless exposure to sensationalized or traumatic narratives [87]. Left unchecked, the rapid proliferation of this addiction could severely escalate demands on the already strained public health systems, which remain under-resourced and unequipped for this looming digital-health epidemic [88].

Brain Development and Educational Disruption. Adolescents face elevated risks from this addiction: their developing prefrontal cortices (critical for impulse control and empathy) would be hijacked by hyper-personalized content's dopamine-driven feedback loops. Evidence suggests adolescent brains release more dopamine than adults to novel stimuli [73]. This plasticity, meant to foster learning, instead would become a liability, making them prone to compulsive engagement that mimics ADHD neural patterns (impaired inhibition, distractibility [89]) and disrupts sleep-dependent memory consolidation [90]. We highlight that OECD PISA data has already linked rising screen time to lower academic performance [91].

Economic Productivity and Labor-Market Effects. The workplace and economic ramifications of hyper-personalized AI addiction could mirror the productivity drag of substance dependencies: compulsive engagement would fracture workplace focus, costing firms billions through presenteeism and task-switching. Studies already estimate that digital interruptions reduce worker productivity by nearly 28% [74], a figure very likely to escalate with the rise of AI-generated media addiction. This may mirror Yemen's khat epidemic, where addiction consumed "one-quarter of usable work hours" [92], and the U.S. opioid crisis, which reduced workforce participation by nearly 2 million workers between 1999–2015, costing the U.S. economy nearly $1.6 trillion [93]. Addictive AI-generated content could globalize these harms without the physical constraints of traditional addictions, potentially causing widespread productivity losses across the workforce.

Differential Effects, Widening Gap. Hyper-personalized content addiction may increase inequality by disproportionately targeting vulnerable populations such as low-income households and those with pre-existing mental health conditions. Free, ad-supported tiers would amplify engagement at all costs, while premium "digital-wellness" subscriptions and AI usage dashboards may remain paywalled, mirroring Big Tobacco's predatory marketing in U.S.'s Black neighbourhoods during the 1950s-1970s [94, 75]. Neuro-divergent individuals, such as those with autism or ADHD, will remain uniquely susceptible to this addictive design: studies show they spend 26-50% more time on algorithmic platforms due to heightened sensory-seeking behaviors [95]. Globally, tech firms will be well-positioned to exploit weak data laws in the Global South to beta-test unregulated AI models – a "digital dumping" akin to 1990s pharmaceutical exploitation [96]. In emerging economies, where cheap smartphones outpace broadband access [97], addictive feeds may eclipse education, trapping a "scrolling class" in low-wage gig work. Lack of rehabilitation access would exacerbate inequality: high-income countries may offer detox clinics, while low-income regions face treatment deserts (currently <1 psychiatrist per 100k people vs. 10+ in wealthy nations [98]). The result would be an entrenched inequality in which those harmed by AI-driven addiction are largely low-income and marginalized groups, while corporations reap the profits.

Environmental and Geopolitical Externalities. The increased consumption of real-time personalized generative AI content could further strain the already significant energy demands from generative AI [99, 100, 101]. On a geopolitical front, analysts already map an emerging "AI bloc" of competition [102] and we may witness further disparity across countries as a result of the differences in regulation [96, 103]. Furthermore, algorithmic content is already being used as a tool to pacify dissent and shape opinion in authoritarian spheres [104] which may be increased further with real-time generative AI content generation.

In sum, the societal costs of unregulated real-time hyper-personalized AI may be profound. The potential consequences may further parallel past public health crises (like tobacco or opioids) where harmful effects were initially ignored, but here the potential scale and 24/7 nature of the harm could make it even more pervasive.

4 Recommendations and Proposed Guidelines

Left unchecked, a commercial race to optimize for user retention and engagement may lead to hyper-addictive real-time generative AI algorithms that may systematically hijack human reward circuitry [11, 105, 106]. We argue that these real-time, engagement-maximizing generative-AI systems warrant the same public-health safeguards applied to addictive substances.

Historically, products exploiting reward loops (e.g., alcohol, tobacco, gambling) have been regulated via: (i) age restrictions, (ii) warning labels, (iii) taxes or licenses, and (iv) liability for deceptive marketing [107, 108, 109]. Such measures have proven effective. For instance, U.S. teen smoking fell from 36% to 6% between 1997 and 2019 following the 1998 Tobacco Master Settlement Agreement [110]. Similarly, after the implementation of self-exclusion programs for problem gambling in the U.K., 83% of registrants reported that the program had helped them reduce or stop gambling [111].

Regulatory attention to addictive design is growing, however, efforts remain fragmented. The EU Digital Services Act [112] and AI Act [113] mandate risk audits for high-impact platforms, while the U.K.'s Online Safety Act 2023 [114] imposes a statutory duty of care. In the U.S., various states and the proposed Kids Online Safety Act 2025 [115, 34, 116] address "addictive feeds" with a focus on minors. Outside Western contexts, China enforces strict gaming curfews for minors [117], and Brazil proposes algorithmic-risk audits [118]. Despite existing precedents, regulations remain limited, targeting specific features (e.g., infinite scroll) or groups (e.g., children) but not real-time generative systems that personalize content [119]. Lacking mechanisms for rapid new AI deployments, they cannot anticipate novel addictive pathways [120]. Audit and reporting rules varied across jurisdictions further weaken enforcement [121]. Consequently, these measures fall short of safeguarding against next-generation, hyper-personalized "digital heroin." Below, we recommend targeted guidelines specific to real-time generative AI.

4.1 Policy and Regulatory Recommendations

Building on the EU AI Act's tiered risk model [113], we recommend adding an overlay label called the *Designated Addictive System (DAS)*. A generative AI service (whether a model or platform) would receive this label if its primary function is to maximize continuous user engagement (e.g., through infinite-scroll short-video feeds or a stream of real-time personalized content). Under this DAS classification, we propose operators be required to:

1. Submit a pre-market safety case showing built-in addiction mitigations (session caps, break reminders).
2. Undergo annual third-party audits measuring compulsive-use patterns and validating countermeasures.
3. Publish transparency reports on engagement percentiles and the effectiveness of their interventions.

These requirements function similarly to the Digital Services Act designation of "Very Large Online Platforms" and the AI Act's obligations for foundation model providers. By formalizing a DAS category, regulators can impose stricter obligations on services with the highest risk of abuse. We now outline specific proposals aimed at enforcement and practical safeguards.

Liability, Oversight, and Enforcement. Responsibility should rest with platform operators, not users. Building on the U.K.'s Online Safety Act [114] and the EU Digital Services Act [112], we propose: (i) strict or negligence-based liability for clinically verified addiction in minors; (ii) algorithmic-transparency subpoenas to expose internal engagement data; and (iii) designation of dark patterns as "unfair practices" by regulators (e.g., FTC). Large monetary penalties analogous to the tobacco settlements could align incentives toward safer design [122].

Design-Friction Mandates. Interface friction demonstrably curbs compulsive use [123, 124, 81]. Mandatory brake pedals for all DAS should therefore: (i) prohibit or heavily gate continuous engagement features (e.g., infinite scroll and autoplay); (ii) impose default session caps and high-salience break reminders, preventing binge usage; (iii) require a one-click option to slow or randomize AI-driven recommendations, reducing the potency of hyper-tailored stimuli; and (iv) require explicit, renewable consent to disable friction measures.

Age-Based Protections. Adolescents are at heightened risk of digital addiction. For example, the U.S. Surgeon General recently cautioned that social-media use may be unsafe for minors, mirroring WHO reports of problematic social-media behavior in 11% of adolescents [9, 125]. We propose: (1) privacy-preserving age-assurance so platforms can default minors into "low-engagement mode" (no late-night notifications or targeted ads); (2) parental controls including guardian dashboards with real-time usage and remote lockout; (3) extension of children's TV ad rules to AI-generated content; and (4) education programs that raise awareness of digital addiction risks.

Major platforms already offer mitigations such as screen-time reminders, daily limits, teen modes, family/parental controls, and periodic "take a break" nudges [126]. We view these as necessary but not sufficient. They are typically voluntary, inconsistently implemented, easy to bypass, and rarely audited for outcomes [127, 128]. Under the proposed DAS regime, these measures complement (but do not substitute for) enforceable obligations: on-by-default protections for minors, standardized configurations, independent outcome audits, and penalties for regressions.

Screen-Time Excise ("Attention Tax"). A more severe option to mitigate attention-related harms may be a Pigouvian levy on excessive screen-time. Building on the U.K. Soft Drinks Industry Levy [129] and Minnesota's draft Social-Media Excise Tax Bill [130], An example tax may include a per-minute tax imposed on platform operators for each user session exceeding 60 minutes within a 24-hour period. The levy internalizes attention-related externalities and aligns platform incentives with healthier engagement patterns which we believe warrants further exploration outside the scope of this paper.

We note that clinical interventions and public education are also vital. A levy on engagement-driven ad revenue could fund peer-support hotlines and research, much like the U.K. Gambling Commission's requirement to finance GambleAware [109]. Large-scale longitudinal studies, modeled on the NIH's Adolescent Brain Cognitive Development project [131], should track long-term neuropsychological outcomes. Moreover, an OECD–WHO consortium could coordinate data-sharing protocols, risk audits, and compliance mechanisms, mirroring the global tobacco-control framework [132]. Such cooperation would help ensure consistent standards, strengthen consumer protections, and preserve room for innovation.

To enable near-term adoption, we highlight three practical steps for DAS operators: (i) ship an on-by-default *slow mode* for minors (session caps, enforced breaks, and night-time curfews); (ii) expose a one-tap "de-intensify personalization" control directly in the feed and honor it across sessions; and (iii) publish a minimal monthly harm dashboard vetted by independent auditors. We recommend tracking a small standardized indicator set: 95^{th}/99^{th} percentile session length and weekly hours (age-stratified), hazard-of-stopping after each item, share of night-time minutes for minors, short-interval re-entry (e.g., within 5 minutes), and break-adherence rates.

4.2 Guidelines for Researchers and Practitioners

In addition to policy measures (Section 4.1), proactive actions by the AI community can help contain the addictive potential of real-time generative AI. Below, we outline practical guidelines for researchers, engineers, and product developers.

Design for Well-Being and User Agency Rather than focusing solely on engagement, embed user mental health metrics (e.g., stress, satisfaction) into optimization objectives [10, 7]. Design platforms and user interfaces to prevent compulsive behavior. For instance, introduce periodic break prompts, usage alerts, and mandatory pauses to disrupt continuous engagement loops [20, 133]. Provide an easily accessible "off switch" (on by default) that reduces personalization or slows content generation. Further, limit usage of "dark patterns" like infinite scroll or auto-play for generative content, especially for younger audiences [134, 135]. For vulnerable segments of the population, introduce design elements that reduce addictive triggers. For example, if a user's interaction history indicates signs of potential overuse, lower engagement-optimizing factors or provide slower response cadences.

Implement Robust Oversight and Testing Before deploying large-scale generative applications, form advisory boards or conduct pre-launch audits with mental health professionals, developmental

psychologists, and sociologists [9] to identify high-risk interface designs or personalization features. Conduct controlled experiments with limited user testing before releasing new generative models to the general public to identify and address any addictive or harmful effects, prior to wide-scale rollout [136, 137]. Run user studies that measure addictive potential, stress, and well-being outcomes over time, rather than focusing solely on short-term engagement metrics [10, 7], and make these findings public to inform broader community understanding and regulatory discussions. Extend Institutional Review Board (IRB) processes or equivalent ethics checks to evaluate addiction risks in user-facing AI research with projects conducting "addiction risk assessments" detailing mitigation plans [20, 138].

Foster Transparency and Accountability. Regularly release anonymized statistics on session duration, repeat visits, and peak engagement windows [19]. Such transparency allows external researchers to evaluate whether platforms nurture compulsive behavior. Provide concise "system cards" or technical briefs explaining how data are collected and how models tailor content in real time [139, 140] and maintain clear documentation to support audits of potentially addictive features [141, 142]. Openly report research sponsors and relevant corporate ties to clarify incentives driving system design [143, 144].

Build Interdisciplinary Collaboration and Education Expand collaborations to include public health, ethics, education, and medical experts [145] to help uncover subtler risks and ethical concerns that single-domain teams might overlook. Incorporate digital well-being topics into machine learning curriculum to cover how reward loops form, how to recognize addictive design, and what frameworks exist for safer user experiences [14, 142]

Research agenda for the NeurIPS community. We see specific roles for the ML research community beyond platform deployment. Research priorities include:

- *Addiction-risk benchmarks:* Create open, real-time personalization benchmarks that simulate users (state, micro-signals, vulnerability profiles) to evaluate behaviors offline.
- *Objective redesign:* Use multi-objective RL and constrained optimization to trade off engagement with well-being, exposure diversity, and long-term satisfaction, with provable guarantees or safety bounds.
- *Interpretability and auditing:* Develop methods to detect when policies exploit sensitive micro-signals (fatigue, hesitancy) or exhibit within-session escalation.
- *Red-teaming protocols:* Standardize addiction-risk red-teaming, paralleling safety evaluations, including teen-risk scenarios and neurodivergent user personas.

We encourage NeurIPS workshops, shared tasks, and datasets around these themes, and support the publication of null results or negative side-effects when optimizing beyond-engagement objectives.

5 Alternative Views

Technological Optimism. Some researchers highlight that real-time generative AI, rather than fueling addiction, can elevate human creativity, education, and well-being. Rapid, on-demand content generation may empower learners with personalized tutoring, assist artists in prototyping ideas, and provide accessible mental health chatbots in under-resourced regions [146, 14]. From this optimistic standpoint, stringent regulation risks stifling innovation and restricting socially beneficial applications. Yet, we argue that the very features enabling such benefits (e.g., continuous feedback, personalization, and emotional resonance) also amplify susceptibility to addictive behaviors, especially when commercial incentives prioritize screen time and engagement.

Libertarian or Anti-Regulatory Critiques. Another viewpoint contends that individuals have the right to consume content freely and assume corresponding risks, casting governmental oversight as paternalistic and potentially detrimental to innovation [141]. Proponents of this stance stress personal responsibility and highlight that excessive regulation could drive cutting-edge AI development overseas [147]. However, our position counters that real-time generative AI exploits power imbalances between profit-driven platforms and less-informed users, limiting genuine autonomy. Without at least

minimal guardrails, platforms can systematically leverage psychological vulnerabilities to maximize revenue, undermining consumer choice in practice.

Feasibility Challenges. Finally, critics note that regulating AI across borders and open-source communities is inherently difficult, risking uneven enforcement or policy evasion [148, 19]. Attempts to outlaw addictive design might be bypassed if smaller developers or overseas actors deploy comparable models without restrictions. In our view, these complications underscore the need for coordinated international standards, transparent auditing mechanisms, and multi-stakeholder coalitions. While perfect implementation is elusive, consistent guidelines can still pressure major market participants to adopt ethical practices and protect vulnerable users.

6 Conclusion

Real-time generative AI offers extraordinary benefits for creativity, problem-solving, and entertainment, yet it also poses a heightened risk of driving users toward addiction-like behavior [16, 17, 7], especially for vulnerable users such as adolescents [9, 73, 89] and individuals with existing mental health conditions [8, 35]. Coupled with misaligned incentives to maximize engagement, real-time generative platforms risk creating "digital heroin" on a scale beyond current social media. Existing research on digital addiction, including neuroimaging studies showing changes in reward circuitry [7, 13, 35], points to a need for focused policy intervention and ethical design practices to prevent prolonged, compulsive usage.

Moving forward, we urge the machine learning community, policymakers, and technology companies to collaborate on safeguards that mirror regulation in fields like gambling and substance control, while also leveraging our scientific understanding of how personalization and frequent feedback loops can harm mental well-being. Through transparent reporting, user-centric design, and age-appropriate restrictions, it is possible to harness the benefits of real-time generative AI without allowing it to devolve into an unchecked public health concern.

References

[1] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[3] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in Neural Information Processing Systems*, 35:8633–8646, 2022.

[4] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

[5] Xinyu Li, Ruiyang Zhou, Zachary C Lipton, and Liu Leqi. Personalized language modeling from personalized human feedback. *arXiv preprint arXiv:2402.05133*, 2024.

[6] OpenAI. Video generation models as world simulators. *OpenAI Technical Report*, Feb 2024.

[7] M León Méndez, I Padrón, A Fumero, and RJ Marrero. Effects of internet and smartphone addiction on cognitive control in adolescents and young adults: A systematic review of fmri studies. *Neuroscience & Biobehavioral Reviews*, 159:105572, 2024.

[8] David Greenfield and Shivan Bhavnani. Social media: Generative ai could harm mental health. *Nature*, 617(7962):676, 2023.

[9] Vivek H. Murthy. Social media and youth mental health: The u.s. surgeon general's advisory. Technical report, Office of the U.S. Surgeon General, 2023.

[10] Chokri Kooli, Youssef Kooli, and Eya Kooli. Generative artificial intelligence addiction syndrome: A new behavioral disorder? *Asian Journal of Psychiatry*, 107:104476, 2025.

[11] Debasmita De, Mazen El Jamal, Eda Aydemir, and Anika Khera. Social media algorithms and teen addiction: Neurophysiological impact and ethical considerations. *Cureus*, 17(1), 2025.

[12] Nancy Costello, Rebecca Sutton, Madeline Jones, Mackenzie Almassian, Amanda Raffoul, Oluwadunni Ojumu, Meg Salvia, Monique Santoso, Jill R Kavanaugh, and S Bryn Austin. Algorithms, addiction, and adolescent mental health: an interdisciplinary study to inform state-level policy action to protect youth from the dangers of social media. *American Journal of Law & Medicine*, 49(2-3):135–172, 2023.

[13] Qinghua He, Ofir Turel, and Antoine Bechara. Brain anatomy alterations associated with social networking site (sns) addiction. *Scientific reports*, 7(1):45064, 2017.

[14] Nizan Geslevich Packin and Karni Chagal-Feferkorn. This is not a game: The addictive allure of digital companions. *Seattle University Law Review*, 48(3):693, 2025.

[15] Miguel Barreda-Ángeles and Tilo Hartmann. Hooked on the metaverse? exploring the prevalence of addiction to virtual reality applications. *Frontiers in virtual reality*, 3:1031697, 2022.

[16] Kenji Kobayashi and Ming Hsu. Common neural code for reward and information value. *Proceedings of the National Academy of Sciences*, 116(26):13061–13066, 2019.

[17] Matthias J Koepp, Roger N Gunn, Andrew D Lawrence, Vincent J Cunningham, Alain Dagher, Tasmin Jones, David J Brooks, Christopher J Bench, and PM Grasby. Evidence for striatal dopamine release during a video game. *Nature*, 393(6682):266–268, 1998.

[18] Ido Hartogsohn and Amir Vudka. Technology and addiction: what drugs can teach us about digital media. *Transcultural Psychiatry*, 60(4):651–661, 2023.

[19] Michelle Nie. Algorithmic addiction by design: Big tech's leverage of dark patterns to maintain market dominance and its challenge for content moderation. *arXiv preprint arXiv:2505.00054*, 2025.

[20] Paula Helm and Tobias Matzner. Co-addictive human–machine configurations: Relating critical design and algorithm studies to medical-psychiatric research on "problematic internet use". *New Media & Society*, 26(12):7295–7313, 2024.

[21] É Duke, C Montag, and M Reuter. Internet addiction: Neuroscientific approaches and therapeutical implications including smartphone addiction. 2017.

[22] California State Legislature. Protecting our kids from social media addiction act. `https://legiscan.com/CA/text/SB976/id/3013535`, September 2024. California Senate Bill 976, Chapter 321, Statutes of 2024.

[23] Maleeha Masood, Shreya Kannan, Zikun Liu, Deepak Vasisht, and Indranil Gupta. Counting how the seconds count: Understanding algorithm-user interplay in tiktok via ml-driven analysis of video content. *arXiv preprint arXiv:2503.20030*, 2025.

[24] Backlinko. Tiktok statistics you need to know in 2025. `https://backlinko.com/tiktok-users`, 2025. Accessed: 2025-05-18.

[25] Xiaocong Chen, Lina Yao, Julian McAuley, Guanglin Zhou, and Xianzhi Wang. Deep reinforcement learning in recommender systems: A survey and new perspectives. *Knowledge-Based Systems*, 264:110335, 2023.

[26] Ruiyang Xu, Jalaj Bhandari, Dmytro Korenkevych, Fan Liu, Yuchen He, Alex Nikulkov, and Zheqing Zhu. Optimizing long-term value for auction-based recommender systems via on-policy reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.

[27] Cai Yang, Sepehr Mousavi, Abhisek Dash, Krishna P Gummadi, and Ingmar Weber. Studying behavioral addiction by combining surveys and digital traces: A case study of tiktok. *arXiv preprint arXiv:2501.15539*, 2025.

[28] Marius Manic. Short-form video content and consumer engagement in digital landscapes. *Bulletin of the Transilvania University of Brasov. Series V: Economic Sciences*, pages 45–52, 2024.

[29] Clara Virós-Martín, Mireia Montaña-Blasco, and Mònika Jiménez-Morales. Can't stop scrolling! adolescents' patterns of tiktok use and digital well-being self-perception. *Humanities and Social Sciences Communications*, 11(1):1–11, 2024.

[30] Universitat Oberta de Catalunya (UOC). 20% of young people spend too much time on tiktok. `https://www.uoc.edu/en/news/2024/adolescents-addiction-to-tiktok`, 2024. Accessed: 2025-05-18.

[31] Knight First Amendment Institute. Understanding social media recommendation algorithms. `https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms`, 2022. Accessed: 2025-05-18.

[32] Thomas Mildner and Gian-Luca Savino. Ethical user interfaces: Exploring the effects of dark patterns on facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[33] Vidico. 25+ instagram reels statistics, data, & trends. `https://vidico.com/news/instagram-reels-statistics/`, 2025. Accessed: 2025-05-18.

[34] New York State Assembly. Assembly bill a8148a—*Stop Addictive Feeds Exploitation (SAFE) for Kids Act*. `https://www.nysenate.gov/legislation/bills/2023/A8148`, 2024.

[35] Keya Ding, Yining Shen, Qianming Liu, and Hui Li. The effects of digital addiction on brain function and structure of children and adolescents: a scoping review. In *Healthcare*, volume 12, page 15. MDPI, 2023.

[36] Jeremy E Solly, Roxanne W Hook, Jon E Grant, Samuele Cortese, and Samuel R Chamberlain. Structural gray matter differences in problematic usage of the internet: a systematic review and meta-analysis. *Molecular psychiatry*, 27(2):1000–1009, 2022.

[37] Mingli Liu, Kimberly E Kamper-DeMarco, Jie Zhang, Jia Xiao, Daifeng Dong, and Peng Xue. Time spent on social media and risk of depression in adolescents: a dose–response meta-analysis. *International journal of environmental research and public health*, 19(9):5164, 2022.

[38] Mark D Griffiths and Daria Kuss. Adolescent social media addiction (revisited). *Education and Health*, 35(3):49–52, 2017.

[39] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*, October 2022. Version 1.

[40] Google DeepMind. Veo 2: Google deepmind's state-of-the-art video generation model. `https://deepmind.google/technologies/veo/veo-2/`, December 2024. Model overview page, accessed 18 May 2025.

[41] Runway Research. Introducing runway gen-4. `https://runwayml.com/research/introducing-runway-gen-4`, March 2025. Accessed 18 May 2025.

[42] Midjourney. Version 7. `https://docs.midjourney.com/hc/en-us/articles/32199405667853-Version`, April 2025. Accessed 19 May 2025.

[43] Google DeepMind. Veo 3: Video, meet audio. `https://blog.google/technology/ai/generative-media-models-io-2025/`, May 2025. Accessed 21 May 2025.

[44] Weijie Kong, Qi Tian, Zijian Zhang, Rox Min, Zuozhuo Dai, Jin Zhou, Jiangfeng Xiong, Xin Li, Bo Wu, Jianwei Zhang, et al. Hunyuanvideo: A systematic framework for large video generative models. *arXiv preprint arXiv:2412.03603*, 2024.

[45] Zhixing Zhang, Yanyu Li, Yushu Wu, Anil Kag, Ivan Skorokhodov, Willi Menapace, Aliaksandr Siarohin, Junli Cao, Dimitris Metaxas, Sergey Tulyakov, et al. Sf-v: Single forward video generation model. *Advances in Neural Information Processing Systems*, 37:103599–103618, 2024.

[46] Jiachen Li, Weixi Feng, Tsu-Jui Fu, Xinyi Wang, Sugato Basu, Wenhu Chen, and William Yang Wang. T2v-turbo: Breaking the quality bottleneck of video consistency model with mixed reward feedback. *arXiv preprint arXiv:2405.18750*, 2024.

[47] Ziqi Huang, Fan Zhang, Xiaojie Xu, Yinan He, Jiashuo Yu, Ziyue Dong, Qianli Ma, Nattapol Chanpaisit, Chenyang Si, Yuming Jiang, et al. Vbench++: Comprehensive and versatile benchmark suite for video generative models. *arXiv preprint arXiv:2411.13503*, 2024.

[48] CoreWeave Inc. Gpu instance pricing. `https://docs.coreweave.com/docs/pricing/pricing-instances`, 2025. Accessed: 2025-05-18.

[49] Meta Platforms, Inc. Labeling ai content. `https://transparency.meta.com/governance/tracking-impact/labeling-ai-content/`, February 2025.

[50] Mandy Dalugdug. Tiktok to start automatically labeling AI-generated content. `https://www.musicbusinessworldwide.com/tiktok-to-start-automatically-labeling-ai-generated-content/`, May 2024.

[51] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[52] Xiangyu Zhao, Long Xia, Liang Zhang, Zhuoye Ding, Dawei Yin, and Jiliang Tang. Deep reinforcement learning for page-wise recommendations. In *Proceedings of the 12th ACM conference on recommender systems*, pages 95–103, 2018.

[53] Renpu Liu, Peng Wang, Donghao Li, Cong Shen, and Jing Yang. A shared low-rank adaptation approach to personalized RLHF. *arXiv preprint arXiv:2503.19201*, 2025.

[54] Meihua Dang, Anikait Singh, Linqi Zhou, Stefano Ermon, and Jiaming Song. Personalized preference fine-tuning of diffusion models. *arXiv preprint arXiv:2501.06655*, 2025.

[55] Runtao Liu, Haoyu Wu, Ziqiang Zheng, Chen Wei, Yingqing He, Renjie Pi, and Qifeng Chen. Videodpo: Omni-preference alignment for video diffusion generation. *arXiv preprint arXiv:2412.14167*, 2024. Accepted at CVPR 2025.

[56] Jiazheng Xu, Yu Huang, Jiale Cheng, Yuanming Yang, Jiajun Xu, Yuan Wang, Wenbo Duan, Shen Yang, Qunlin Jin, Shurun Li, et al. Visionreward: Fine-grained multi-dimensional human preference learning for image and video generation. *arXiv preprint arXiv:2412.21059*, 2024.

[57] Alicia Heraz, Kiran Kumar Ashish Bhyravabhottla, and Nandith Sajith. Predicting user engagement levels through emotion-based gesture analysis of initial impressions. *Electronic Commerce Research*, pages 1–17, 2024.

[58] Seyed Mohsen Ebadi Jokandan, Peyman Bayat, and Mehdi Farrokhbakht Foumani. Targeted advertising in social media platforms using hybrid convolutional learning method besides efficient feature weights. *Journal of Electrical and Computer Engineering*, 2022(1):6159650, 2022.

[59] Tom B. Brown, Benjamin Mann, and Nick et al. Ryder. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[60] Yunhang Shen, Chaoyou Fu, Shaoqi Dong, Xiong Wang, Yi-Fan Zhang, Peixian Chen, Mengdan Zhang, Haoyu Cao, Ke Li, Xiawu Zheng, et al. Long-vita: Scaling large multi-modal models to 1 million tokens with leading short-context accuracy. *arXiv preprint arXiv:2502.05177*, 2025.

[61] Linda He, Jue Wang, Maurice Weber, Shang Zhu, Ben Athiwaratkun, and Ce Zhang. Scaling instruction-tuned llms to million-token contexts via hierarchical synthetic data generation. *arXiv preprint arXiv:2504.12637*, 2025.

[62] Yuwei Guo, Ceyuan Yang, Ziyan Yang, Zhibei Ma, Zhijie Lin, Zhenheng Yang, Dahua Lin, and Lu Jiang. Long context tuning for video generation. *arXiv preprint arXiv:2503.10589*, 2025.

[63] Yuchao Gu, Weijia Mao, and Mike Zheng Shou. Long-context autoregressive video modeling with next-frame prediction. *arXiv preprint arXiv:2503.19325*, 2025.

[64] Zhuoran Liu, Leqi Zou, Xuan Zou, Caihua Wang, Biao Zhang, Da Tang, Bolin Zhu, and Youlong Cheng. Monolith: Real-time recommendation system with collisionless embedding table. *arXiv preprint arXiv:2209.07663*, 2022.

[65] Tiktok launches ai-powered video platform to advertisers globally. Reuters Technology, Nov 2024. Accessed 18 May 2025.

[66] YouTube Official Blog. Made on youtube: Empowering anyone to create on youtube. blog.youtube, 2023. Accessed 18 May 2025.

[67] Meta Platforms Inc. Systems and methods for generation and delivery of enhanced content utilizing remote rendering and data streaming, May 2023.

[68] Alex Heath. Meta's ai now animates still images to create personalised video ads. The Verge, Oct 2024. Accessed 18 May 2025.

[69] Kent C. Berridge and Terry E. Robinson. Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, 74(11):1264–1276, 2019.

[70] Mehdi Keramati and Boris Gutkin. Imbalanced decision hierarchy in addicts emerging from drug-hijacked dopamine spiraling circuit. *PloS one*, 8(4):e61489, 2013.

[71] Ofir Turel, Qinghua He, Damien Brevers, and Antoine Bechara. Delay discounting mediates the association between posterior insular cortex volume and social media addiction symptoms. *Cognitive, Affective, & Behavioral Neuroscience*, 18(4):694–704, 2018.

[72] Rebecca Nowland, Elizabeth A Necka, and John T Cacioppo. Loneliness and social internet use: pathways to reconnection in a digital world? *Perspectives on psychological science*, 13(1):70–87, 2018.

[73] Linda P Spear. The adolescent brain and age-related behavioral manifestations. *Neuroscience & biobehavioral reviews*, 24(4):417–463, 2000.

[74] Gloria Mark. *Attention span: A groundbreaking way to restore balance, happiness and productivity*. Harlequin, 2023.

[75] Phillip S Gardiner. The african americanization of menthol cigarette use in the united states. *Nicotine & Tobacco Research*, 6(Suppl_1):S55–S65, 2004.

[76] Susanna Paasonen. Fickle focus: Distraction, affect and the production of value in social media. *First Monday*, 2016.

[77] Jean M Twenge and W Keith Campbell. Media use is linked to lower psychological well-being: Evidence from three datasets. *Psychiatric Quarterly*, 90:311–331, 2019.

[78] Jean M Twenge, Thomas E Joiner, Megan L Rogers, and Gabrielle N Martin. Increases in depressive symptoms, suicide-related outcomes, and suicide rates among us adolescents after 2010 and links to increased new media screen time. *Clinical psychological science*, 6(1):3–17, 2018.

[79] Chaelin K Ra, Junhan Cho, Matthew D Stone, Julianne De La Cerda, Nicholas I Goldenson, Elizabeth Moroney, Irene Tung, Steve S Lee, and Adam M Leventhal. Association of digital media use with subsequent symptoms of attention-deficit/hyperactivity disorder among adolescents. *Jama*, 320(3):255–263, 2018.

[80] Anne-Marie Chang, Daniel Aeschbach, Jeanne F Duffy, and Charles A Czeisler. Evening use of light-emitting ereaders negatively affects sleep, circadian timing, and next-morning alertness. *Proceedings of the National Academy of Sciences*, 112(4):1232–1237, 2015.

[81] Melissa G Hunt, Rachel Marx, Courtney Lipson, and Jordyn Young. No more fomo: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, 37(10):751–768, 2018.

[82] Michael B First, Lamyaa H Yousif, Diana E Clarke, Philip S Wang, Nitin Gogtay, and Paul S Appelbaum. Dsm-5-tr: Overview of what's new and what's changed. *World Psychiatry*, 21(2):218, 2022.

[83] World Health Organization. International Classification of Diseases, 11th Revision (ICD-11): Code 6C51 *Gaming disorder*. `https://www.who.int/standards/classifications/frequently-asked-questions/gaming-disorder`, 2019.

[84] Mark S Tremblay and J Douglas Willms. Is the canadian childhood obesity epidemic related to physical inactivity? *International journal of obesity*, 27(9):1100–1105, 2003.

[85] Vaishnavi S Nakshine, Preeti Thute, Mahalaqua Nazli Khatib, Bratati Sarkar, et al. Increased screen time as a cause of declining physical, psychological health, and sleep patterns: a literary review. *Cureus*, 14(10), 2022.

[86] Alexandra-Regina Tsantili, Dimosthenis Chrysikos, and Theodore Troupis. Text neck syndrome: disentangling a new epidemic. *Acta medica academica*, 51(2):123, 2022.

[87] Katherine Ormerod. *Why social media is ruining your life*. Hachette UK, 2018.

[88] Shekhar Saxena, Graham Thornicroft, Martin Knapp, and Harvey Whiteford. Resources for mental health: scarcity, inequity, and inefficiency. *The lancet*, 370(9590):878–889, 2007.

[89] Betul Keles, Niall McCrae, and Annmarie Grealish. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International journal of adolescence and youth*, 25(1):79–93, 2020.

[90] Susanne Diekelmann and Jan Born. The memory function of sleep. *Nature reviews neuroscience*, 11(2):114–126, 2010.

[91] OECD. Students, digital devices and success. *OECD Education Policy Perspectives*, (102), 2024.

[92] World Bank. Yemen-towards qat demand reduction. Technical report, World Bank, 2007.

[93] Ben Gitis. The workforce and economic implications of the opioid crisis. *Statement before the US House of Representatives Committee on Small Business, Washington, DC, September*, 13, 2018.

[94] Mark Wolfson. *The fight against big tobacco: the movement, the state and the public's health*. Routledge, 2017.

[95] Aviva Must, Misha Eliasziw, Heidi Stanish, Carol Curtin, Linda G Bandini, and April Bowling. Passive and social screen time in children with autism and in association with obesity. *Frontiers in Pediatrics*, 11:1198033, 2023.

[96] Nick Couldry and Ulises A Mejias. The costs of connection: How data is colonizing human life and appropriating it for capitalism. In *The costs of connection*. Stanford University Press, 2019.

[97] Jacob Poushter et al. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew research center*, 22(1):1–44, 2016.

[98] World Health Organization. Mental health atlas 2021. `https://www.who.int/publications/i/item/9789240036703`, 2021.

[99] Martin Oteng-Ababio and Richard Grant. E-waste recycling slum in the heart of accra, ghana: the dirty secrets. In *Handbook of Electronic Waste Management*, pages 355–376. Elsevier, 2020.

[100] David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David R So, Maud Texier, and Jeff Dean. The carbon footprint of machine learning training will plateau, then shrink. *Computer*, 55(7):18–28, 2022.

[101] Pengfei Li, Jianyi Yang, Mohammad A Islam, and Shaolei Ren. Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models. *arXiv preprint arXiv:2304.03271*, 2023.

[102] Michael Raska and Richard A Bitzinger. *The AI wave in defence innovation: Assessing military artificial intelligence strategies, capabilities, and trajectories*. Taylor & Francis, 2023.

[103] Reuters. France bans tiktok, other 'recreational' apps on government phones. *Reuters*, March 2023.

[104] Alina Polyakova and Chris Meserole. Exporting digital authoritarianism: The russian and chinese models. *Policy brief, democracy and disorder series*, (August 2019):1–22, 2019.

[105] Hunt Allcott, Matthew Gentzkow, and Lena Song. Digital addiction. *American Economic Review*, 112(7):2424–2463, 2022.

[106] Ala Yankouskaya, Magnus Liebherr, and Raian Ali. Can chatgpt be addictive? a call to examine the shift from support to dependence in ai conversational large language models. *Human-Centric Intelligent Systems*, pages 1–13, 2025.

[107] Sally Casswell and Thaksaphon Thamarangsi. Reducing harm from alcohol: call to action. *The Lancet*, 373(9682):2247–2257, 2009.

[108] Robert B Wallace, Kathleen Stratton, and Richard J Bonnie. *Ending the tobacco problem: a blueprint for the nation*. National Academies Press, 2007.

[109] Heather Wardle, Gerda Reith, Erika Langham, and Robert D Rogers. Gambling and public health: we need policy action to prevent harm. *Bmj*, 365, 2019.

[110] National Association of Attorneys General. The master settlement agreement. `https://www.naag.org/our-work/naag-center-for-tobacco-and-public-health/the-master-settlement-agreement/`. Accessed: 2025-05-20.

[111] Chrysalis Research. Evaluation of the multi-operator self-exclusion scheme (moses). Technical report, GambleAware, London, 2017. Commissioned under the Gambling Act 2005.

[112] Regulation (EU) 2022/2065 on a single market for digital services (digital services act). Official Journal of the European Union, 2022.

[113] Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (artificial intelligence act). Official Journal of the European Union, 2024.

[114] Online safety act 2023, c. 50. UK Public General Acts, 2023.

[115] California Legislature. Senate bill 976—*Protecting Our Kids from Social Media Addiction Act*, 2024. Chapter 321.

[116] U.S. Congress. Kids Online Safety Act, s.1748, 119th congress (2025–2026). `https://www.congress.gov/bill/119th-congress/senate-bill/1748`, 2025.

[117] State Administration of Press and Publication, PRC. Notice on Further Strict Management and Effective Prevention of Minors' Addiction to Online Games (guo xin chu fa [2021] 14). `https://www.gov.cn/zhengce/zhengceku/2021-09/01/content_5634661.htm`, August 2021.

[118] Câmara dos Deputados. Substitutive Report on Draft Bill No. 2630/2020 – *Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet*. https://static.poder360.com.br/2023/04/pl-fake-news-camara.pdf, 2024.

[119] Ruth Elisabeth Appel. Generative ai regulation can learn from social media regulation. *arXiv preprint arXiv:2412.11335*, 2024.

[120] Noam Kolt, Michal Shur-Ofry, and Reuven Cohen. Lessons from complexity theory for ai governance. *arXiv preprint arXiv:2502.00012*, 2025.

[121] David Krause. Addressing the challenges of auditing and testing for ai bias: A comparative analysis of regulatory frameworks. *Available at SSRN*, 2024.

[122] California Department of Justice. Master Settlement Agreement Fact Sheet. https://oag.ca.gov/tobacco/msa, 1998.

[123] Information Commissioner's Office. Age Appropriate Design Code: A code of practice for online services. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/, August 2020.

[124] Alex Hern. Social media giants increase global child safety after uk regulations introduced. *The Guardian*.

[125] Meyran Boniel-Nissim, Claudia Marino, Tommaso Galeotti, Lukas Blinka, Kristīne Ozoliņa, Wendy Craig, Henri Lahti, Suzy L Wong, Judith Brown, Mary Wilson, et al. A focus on adolescent social media use and gaming in europe, central asia and canada: Health behaviour in school-aged children international report from the 2021/2022 survey. volume 6. 2024.

[126] Jan Keller, Tobias Herrmann-Schwarz, Christina Roitzheim, Lea Mertens, Lina Christin Brockmeier, and Aditya Kumar Purohit. A digital nudge-based intervention to interrupt instagram usage: Randomized controlled pilot study. *European Journal of Health Psychology*, 2024.

[127] Mariya Stoilova, Monica Bulger, and Sonia Livingstone. Do parental control tools fulfil family expectations for child protection? a rapid evidence review of the contexts and outcomes of use. *Journal of Children and Media*, 18(1):29–49, 2024.

[128] Alberto Monge Roffarello and Luigi De Russis. Achieving digital wellbeing through digital self-control tools: A systematic review and meta-analysis. *ACM Transactions on Computer-Human Interaction*, 30(4):1–66, 2023.

[129] The soft drinks industry levy regulations 2018. https://www.legislation.gov.uk/uksi/2018/41/contents/made, 2018. U.K. Statutory Instrument 2018 No. 41.

[130] Hf 3117 / sf 3197 — social media platform excise tax bill. https://www.revenue.state.mn.us/sites/default/files/2025-04/sf3197hf3117-social-media-gross-receipts-tax-2.pdf, 2025.

[131] Betty Jo Casey, Tariq Cannonier, May I Conley, Alexandra O Cohen, Deanna M Barch, Mary M Heitzeg, Mary E Soules, Theresa Teslovich, Danielle V Dellarco, Hugh Garavan, et al. The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience*, 32:43–54, 2018.

[132] Convention Secretariat, WHO Framework Convention on Tobacco Control. 2023 Global Progress Report on Implementation of the WHO Framework Convention on Tobacco Control. Technical report, World Health Organization, Geneva, 2023.

[133] Luke Haliburton, David J. Grüning, Frank Riedel, Albrecht Schmidt, and Nina Terzimehić. A longitudinal in-the-wild investigation of design frictions to prevent smartphone overuse. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, 2024.

[134] Fabrizio Esposito and Thaís Maciel Cathoud Ferreira. Addictive design as an unfair commercial practice: The case of hyper-engaging dark patterns. *European Journal of Risk Regulation*, 15(4):999–1016, 2024.

[135] European Parliament. European Parliament resolution of 12 December 2023 on addictive design of online services and consumer protection in the EU single market. P9_TA(2023)0459, December 2023. Procedure 2023/2043(INI). Official Journal C 4164, 2 Aug 2024.

[136] Irene Dankwa-Mullan. Health equity and ethical considerations in using artificial intelligence in public health and medicine. *Preventing chronic disease*, 21:E64, 2024.

[137] Sonya Falahati, Morteza Alizadeh, Zhino Safahi, Navid Khaledian, Mohsen Alambardar Meybodi, and Mohammad R Salmanpour. An ai-powered public health automated kiosk system for personalized care: An experimental pilot study. *arXiv preprint arXiv:2504.13880*, 2025.

[138] Jonas Schuett, Ann-Katrin Reuel, and Alexis Carlier. How to design an ai ethics board. *AI and Ethics*, pages 1–19, 2024.

[139] OpenAI. Creating video from text (sora system card). `https://openai.com/product/sora`, Feb 2024.

[140] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.

[141] MB Lawrence. Public health law's digital frontier: Addictive design, section 230, and the freedom of speech, 2023.

[142] Christian Montag and Jon D Elhai. On social media design,(online-) time well-spent and addictive behaviors in the age of surveillance capitalism. *Current Addiction Reports*, 10(3):610–616, 2023.

[143] Maarten Bosten and Bennett Kleinberg. Conflicts of interest in published nlp research 2000-2024. *arXiv preprint arXiv:2502.16218*, 2025.

[144] Thilo Hagendorff and Kristof Meding. Ethical considerations and statistical analysis of industry involvement in machine learning research. *Ai & Society*, pages 1–11, 2023.

[145] Polat Goktas and Andrzej Grzybowski. Shaping the future of healthcare: Ethical clinical challenges and pathways to trustworthy ai. *Journal of Clinical Medicine*, 14(5):1605, 2025.

[146] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

[147] Filippo Lancieri, Laura Edelson, and Stefan Bechtold. Ai regulation: Competition, arbitrage & regulatory capture. *Georgetown University Law Center Research Paper No. 2025/05*, 2025. SSRN working paper, posted 31 January 2025.

[148] Tucker Craven. Kids, no phones at the dinner table: Analyzing the people's republic of china's proposed" minor mode" regulation and an international right to the internet. *Chi. J. Int'l L.*, 25:219, 2024.