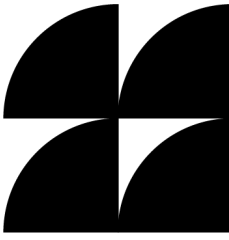
Knight Georgetown Institute

PANOPTYKON FOUNDATION

7 APRIL 2025

# European Board for Digital Services and European Commission Report on Systemic Risks and Mitigations under the Digital Services Act

## Comments of the Knight-Georgetown Institute (KGI) and Panoptykon Foundation



# About the Knight-Georgetown Institute

The Knight-Georgetown Institute (KGI) is dedicated to connecting independent research with technology policy and design. KGI serves as a central hub for the growing network of scholarship that seeks to shape how technology is used to produce, disseminate, and access information. KGI is designed to provide practical resources that policymakers, journalists, and private and public sector leaders can use to tackle information and technology issues in real time. Georgetown University and the Knight Foundation came together to launch the institute in 2024. Learn more about KGI at https://kgi.georgetown.edu.

# About the Panoptykon Foundation

Panoptykon Foundation is a Polish watchdog NGO, founded in 2009, with a mission to protect human rights and agency in the context of growing surveillance and intrusive information technologies. Panoptykon works at the national (Polish) and European level. Through their research, litigation, and strategic communication, Panoptykon exposes risks related to commercial and public surveillance. They keep track of new legislation that regulates technology, develop human-centric solutions, advocate for them, and test them in practice. Panoptykon is an active member of the European Digital Rights, the DSA Working Group chaired by the Center for Democracy and Technology, and the People VS Big Tech movement. Learn more at https://en.panoptykon.org/.

# Introduction

The Knight-Georgetown Institute (KGI)[1] and the Panoptykon Foundation[2] welcome the opportunity to provide joint input to the report on systemic risks being developed by the European Board for Digital Services, in cooperation with the European Commission. Our comments respond to the four sets of questions provided by the Commission.

Systemic risk assessment and mitigation requirements are core components of the DSA framework and can be more fully integrated into the day-to-day operations of platforms. Our responses and recommendations below seek to help evolve disparate DSA implementation approaches into a more cohesive whole that links upstream design considerations with user agency and robust transparency.

# Question 1: Prominent or recurrent systemic risks

Product design may impact the full range of systemic risks, including threats to fundamental rights, the spread of illegal content, and problematic overuse or other harms to minors. Numerous comprehensive studies and policy frameworks draw a link between design and risk, including the Australia eSafety Commissioner's Safety by Design framework,[3] the Dutch Code for Children's Rights,[4] OECD's Digital Safety by Design for Children,[5] the UK's Age Appropriate Design Code,[6] the US National Academies of Sciences, Engineering, and Medicine's Consensus Report on Social Media and Adolescent Health,[7] and many others.

Through a focus on platform design, researchers and scholars across a range of disciplines are working to build frameworks for effective, ongoing, and comparable measurement of social media risks and mitigation strategies. A new Integrated Harm Framework, developed by the Center for Digital Health (CDH) and Social Media Lab (SML) at Stanford University, documents youth-specific social media harms identified in empirical studies as well as the strength of available evidence for specific mitigation strategies.[8] Integration of these measurement tools into the systemic risk assessment

[1] Based at Georgetown University, the Knight-Georgetown Institute is dedicated to connecting independent research with technology policy and design, with a focus on the online information environment. See Knight-Georgetown Institute, "About."
[2] Panoptykon Foundation is a Polish watchdog NGO, founded in 2009, with a mission to protect human rights and agency in the context of growing surveillance and intrusive information technologies. See Panoptykon Foundation. "What is Panoptykon."
[3] eSafety Commissioner, "Safety by Design Overview."
[4] University of Leiden and the Waag organisation, "Code for Children's Rights."
[5] Organisation for Economic Co-operation and Development, "Towards Digital Safety by Design for Children."
[6] Information Commissioner's Office, "Age appropriate design."
[7] National Academies of Sciences, Engineering, and Medicine, *Social Media and Adolescent Health*.
[8] Center for Digital Health, "Stanford Youth Safety and Digital Wellbeing Report, 2025."

process will strengthen assessment by allowing for objective and measurable assessment of platform risks and mitigation strategies.

Working with the USC Neely Center and the Tech Justice Law Project, KGI recently developed a taxonomy to map risks to specific platform designs, discussed below.[9] Alongside these risks associated with design, our comments also considers polarization and trust, as well as risks to minors.

Our comments highlight empirical work across risk areas relevant to the Commission's understanding of prominent and recurrent systemic risks, as well as effective mitigations. Selected documentation, research, and resources are presented in each of the following sections.

## A. Problematic and harmful use

Independent research links excessive or problematic social media use with physical and mental well-being impacts referenced in Article 34(1)(d).

Problematic social media use can disrupt everyday activities, such as school attendance[10] or sleep,[11] and can be associated with negative mental health outcomes, including depression.[12] Existing research shows that social media use contributes to a drop in time associated with sleep and exercise.[13] Problematic social media use is closely tied to the design choices platforms make. Multiple platform design features may extend time spent on social media, including infinite scroll,[14] autoplay,[15] gamification of engagement,[16] ephemeral content,[17] and the timing and clustering of notifications. Minors may be particularly vulnerable to platform design features intended to prolong use.[18]

To deliver hyper-personalized experiences, social media recommender systems may rely on behavioural patterns. These patterns may reveal individual vulnerabilities such as addictions, eating disorders, body complexes, anxiety, or depressive disorders. Recommender systems designed to maximize user engagement may purposefully or inadvertently exploit or exacerbate these individual vulnerabilities. Depending on the individual, these systems can also create feedback loops that drive users into narrower selections of content, corresponding to their vulnerabilities.[19] Such content may not be dangerous per se, and may be entirely acceptable when considered in isolation, but becomes harmful if consumed consistently over time by vulnerable individuals.

[9] Knight-Georgetown Institute et al., "Taxonomy - Mapping Consumer Harm to Specific Social Media Design Elements."
[10] Caplan, "Theory and Measurement of Generalized Problematic Internet Use."
[11] Casale and Banchi, "Narcissism and Problematic Social Media Use."
[12] Bányai et al., "Problematic Social Media Use."
[13] Brautsch et al., "Digital Media Use and Sleep in Late Adolescence and Young Adulthood."
[14] Langvardt, "Regulating Habit-Forming Technology."
[15] Lukoff et al., "How the Design of YouTube Influences User Sense of Agency."
[16] Bernstein, *Unwired*.
[17] Zhang et al., "Ephemerality in Social Media."
[18] Paakkari et al., "Problematic Social Media Use and Health among Adolescents."
[19] Costello et al., "Algorithms, Addiction, and Adolescent Mental Health"; Griffiths et al., "Does TikTok Contribute to Eating Disorders?".

Falling into a "doomscrolling" trap (e.g., excessive exposure to self-harm, diet-related content, or idealised body images) can trigger unhealthy engagement, which negatively impacts users' wellbeing and may exacerbate their pre-existing mental health issues.[20] For certain vulnerable individuals, platforms may not introduce sufficient friction that would prevent them from getting stuck in a pattern of scrolling for negative information.

## B. Unwanted and harmful contact

Platform design elements that make users' accounts broadly visible or discoverable by default and recommend connection with strangers may contribute to a range of systemic risks including the dissemination of illegal content, impact on fundamental rights, and negative effects on physical and mental well-being.

Platform design plays an important role in creating risks of unwanted and harmful contact. Some platforms, for example, enable user visibility by default and recommend user accounts to others outside their network, as well as accounts outside a user's network to them. These designs pose particular risks to minors by enabling bad actors to target and/or mass contact minors' accounts. Research has found that expansive default account visibility and account recommendations are crucial design vulnerabilities for sextortion targeting minors.[21]

Litigation in the United States contends that platform design can encourage unwanted contact. Internal data and research from Snap, for example, shows that these risks can be systemic. In the US, Snap's internal data is alleged to show that "one-third of teen girls and 30% of teen boys were exposed to unwanted contact on its platform."[22] An internal Instagram user survey found that over a seven-day period, nearly twenty percent of 13-15-year-olds reported receiving unwanted sexually explicit content, and 13% experienced an unwanted sexual advance.[23]

## C. Unwanted and harmful content

Platform design can incentivize and enable the development and distribution of unwanted and harmful content. Such designs contribute to a range of systemic risks, including dissemination of illegal content, impact on fundamental rights, and negative effects on physical and mental well-being.

Recommender systems may play a critical role, as described in our response to Questions 2 and 3 below. Optimizing recommender systems to maximize predicted short-term engagement can lead to the spread of sensational or borderline content that may exacerbate systemic risks. There is empirical research documenting that optimizing for engagement contributes to increased encounters with

[20] Ibid.
[21] FBI San Francisco, "FBI San Francisco Warns of Increase in Sextortion Schemes Targeting Young Boys"; Raffile et al., "A Digital Pandemic"; Raffile et al., "A Digital Pandemic."
[22] New Mexico Department of Justice, "Attorney General Raúl Torrez Files Unredacted Complaint Against Snapchat, Exposing Internal Messages That Snap Knowingly Contributed to Harm Amongst Children."
[23] Béjar, "Bad Experiences and Encounters Framework (BEEF) Survey."

borderline content (such as insults and targeted cursing) and low-quality information about news events.[24]

Researchers have found that engagement is negatively related to quality.[25] Content with the highest predicted engagement may score low in terms of quality and trustworthiness (attributes of quality used by platforms include sensationalised language, engagement bait, among others).[26] A 2021 study by Mozilla Foundation found that 71% of videos that were reported as harmful (referred to as 'Regret reports' on YouTube) were recommended to viewers on the platform. Recommended videos were 40% more likely to be reported as harmful than videos users found via specific searches. These videos often contained harmful content such as violence, misinformation, hate speech, and scams.[27] The Integrity Institute has similarly shown that engagement-based ranking can amplify low quality content through recommendations from non-followed accounts and frictionless resharing mechanisms.[28] Internal Facebook documents leaked to *The Wall Street Journal* have shown that Facebook's recommender systems proactively connected users with extremists. Internal research cited by the *Journal* reportedly states that "64% of all extremist group joins are due to [Facebook's] recommendation tools."[29]

Aside from recommender systems, platforms may deploy other design patterns that incentivize users to develop and disseminate harmful content. For example, some platforms enable lenses and filters that let users alter images and their appearance. Filters may include features like touching up an image of a sunset or adding animal features to a human face. But lenses may also apply so-called "beautification" to a user's photos.[30] For the last several years, research has explored associations between social media, body image, and distorted eating.[31] Recent work has shown filters may negatively impact users' body image and promote unrealistic social comparison.[32] Much of this research focuses on specific risks to women and girls.[33]

## D. Polarization and trust

Research explores how engagement-based feeds may impact polarization and trust.

Though research attempting to link social media with political polarization in the aggregate has had mixed effects,[34] experimental studies have found that, compared with alternative designs,

[24] Allen et al., "Better Feeds."
[25] Cunningham et al., "What We Know About Using Non-Engagement Signals in Content Ranking."
[26] Meta, "About Quality Ranking."
[27] McCrosky and Geurkink, "YouTube Regrets."
[28] Allen, "Misinformation Amplification Analysis and Tracking Dashboard."
[29] Horowitz et al., "Facebook Executives Shut Down Efforts to Make the Site Less Divisive."
[30] Ryan-Mosley, "Beauty Filters Are Changing the Way Young Girls See Themselves."
[31] Holland and Tiggemann, "A Systematic Review of the Impact of the Use of Social Networking Sites on Body Image and Disordered Eating Outcomes."
[32] Dijkslag et al., "To Beautify or Uglify! The Effects of Augmented Reality Face Filters on Body Satisfaction Moderated by Self-Esteem and Self-Identification"; Schroeder and Behm-Morawitz, "Digitally Curated Beauty."
[33] Pescott, "'I Wish I Was Wearing a Filter Right Now.'"
[34] Kubin and von Sikorski, "The Role of (Social) Media in Political Polarization."

engagement-based ranking elevates negative emotions (including anger and sadness) and hostility toward outgroups among users, as well as the share of items expressing this negativity and hostility.[35]

Optimizing for engagement may also shape the kinds of items users are exposed to in detrimental ways. Empirical research has documented how engagement contributes to increased encounters with borderline abuse (such as insults and targeted cursing)[36] and low-quality information about news events.[37] A 2021 EU-representative survey of 2,000 people between the ages of 18 and 80, for example, found that significant rates of internet users, especially women, experience hate and violence online and thus change their behaviour or withdraw from social media.[38]

Research also shows that certain groups may benefit from algorithmic amplification while others face exclusion.[39] A 2023 study focused on Google, Google News, Facebook, YouTube, and Twitter, for example, found that recommender systems delivered homogenized results.[40] By dialing recommender system weights up and down, platform design may give visibility boosts to certain media outlets or voices leaning towards certain categories of people, political perspectives, or sides of a dispute.[41] These mechanisms may impact the pluralism of public debate and trust.

## E. Unique risks to minors

Research into the cognitive and social-emotional development of adolescents indicates they may be more vulnerable to social media risks than adults.[42] Several traits unique to adolescent development affect their social media use,[43] including sensitivity to social acceptance from peers.[44] Adolescents' opinions and decisions may be more influenced by peers than by adults.[45] In addition, during adolescence, regions of the brain associated with emotional processing develop faster than those involved with reasoning and impulse control,[46] which may shape how adolescents use and respond to social media content.[47]

[35] Milli et al., "Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media"; Piccardi et al., "Social Media Algorithms Can Shape Affective Polarization via Exposure to Antidemocratic Attitudes and Partisan Animosity."
[36] Bandy and Lazovich, "Exposure to Marginally Abusive Content on Twitter."
[37] Moehring, "Personalization, Engagement, and Content Quality on Social Media."
[38] HateAid, "Boundless Hate on the Internet – Dramatic Situation across Europe."
[39] Nobel, "Algorithms of Oppression."
[40] Nechushtai et al., "More of the Same?"
[41] Moehring, "Personalization, Engagement, and Content Quality on Social Media."
[42] See, e.g., Office of the Surgeon General, "Social Media and Youth Mental Health."
[43] Crone and Konijn, "Media Use and Brain Development during Adolescence."
[44] Somerville, "The Teenage Brain."
[45] Ibid.
[46] Casey et al., "The Adolescent Brain."
[47] 5Rights Foundation, "Pathways"; Costello et al., "Algorithms, Addiction, and Adolescent Mental Health"; Chen et al., "The Engagement-Prolonging Designs Teens Encounter on Very Large Online Platforms"; Pizzo Frey et al., "Recommendation Systems in Social Media."

Given the reality of cognitive development in adolescents, numerous scholars have conducted research on the experiences of adolescents in digital environments. While research shows that social media use can deliver benefits to young people,[48] current science also suggests that certain features of social media can be harmful to the health of some adolescents.[49] Adolescents use social media as a key site of discussion for a wide range of topics, such as issues of body image and weight amongst teenage girls.[50] Researchers have documented important patterns in the effects of social media on adolescents which contribute to systemic risks related to illegal content, fundamental rights, gender-based violence, public health, and serious negative consequences to the person's physical and mental well-being. Relevant research exploring risks to minors is briefly summarized below.[51]

| Risk to minors | Evidence of risk |
|---|---|
| **Social comparison** | Social comparison is a normal feature of cognitive development for adolescents. However, comparison on social media may be particularly associated with negative emotional responses for adolescents.[52] Social comparison online may be related to depression, particularly among girls,[53] and certain subgroups of adolescents may be more prone to harmful effects than others.[54] Research has particularly found that "upward" social comparison, which involves viewing others such as celebrities and models as better off, is often harmful.[55] |

[48] See for example: Hadjipanayis et al., "Social Media and Children"; National Academies of Sciences, Engineering, and Medicine, "The Relation between Social Media and Health"; Reid Chassiakos et al., "Children and Adolescents and Digital Media"; Yau et al., "Coping With Stress Through Texting."
[49] See, e.g., American Psychological Association, "Potential Risks of Content, Features, and Functions."
[50] Crone and Konijn, "Media Use and Brain Development during Adolescence."
[51] This table was submitted in KGI's comments to the European Commission on its guidelines to enforce the protection of minors online in October 2024. See Chapman and Arnao, "KGI Comments to the European Commission on Guidelines to Enforce the Protection of Minors Online."
[52] Nesi and Prinstein, "Using Social Media for Social Comparison and Feedback-Seeking"; Vandenbosch et al., "Social Media and Body Image."
[53] Cataldo et al., "Social Media Usage and Development of Psychiatric Disorders in Childhood and Adolescence."
[54] Cingel et al., "U.S. Adolescents' Attitudes toward School, Social Connection, Media Use, and Mental Health during the COVID-19 Pandemic"; Weinstein, "Adolescents' Differential Responses to Social Media Browsing."
[55] de Vries et al., "Social Comparison as the Thief of Joy"; McComb and and Tobin, "A Meta-Analysis of the Effects of Social Media Exposure to Upward Comparison Targets on Self-Evaluations and Emotions."

| | |
|---|---|
| **Body image, dissatisfaction, and disordered eating** | Social media use may contribute to disordered eating and associated mental health issues in adolescents.[56] Disentangling the direction of causality is a subject of ongoing study,[57] yet experimental research has observed more image-based comparison and lower body satisfaction among adolescents who often use social media,[58] particularly those using platforms that emphasize sharing of images.[59] Longitudinal studies in Europe and the US have further confirmed the association between extended social media use by adolescents and risks of developing eating disorders.[60] Identifying the specific mechanisms mediating the connection between social media and negative mental health outcomes is an area of increasing study that will be relevant to the Commission. |
| **Displacement of healthy behaviors** | Research has shown that among adolescents, social media use straightforwardly contributes to a drop in time associated with healthier activities such as sleep and exercise.[61] Social media can disrupt sleep in three main ways: delaying and worsening the quality of sleep time, distorting circadian rhythms via light emissions from screens, and increasing psychological stimulation before bedtime.[62] Such disruption is concerning to researchers as insufficient sleep contributes to various other health issues.[63] |
| **Feelings of sadness, anxiety, depression, and stress** | The connection between mental health and social media use is the most studied risk in the literature; however, research does not contain consistent findings.[64] Many reviews and meta-analyses report a weak or even absent association between poor mental health and social media use.[65] The studies reviewed provide limited insight into the causal connection between social media and mental health due to variability in use across minors and platforms and the |

[56] Fitzsimmons-Craft et al., "Adolescents and Young Adults Engaged with Pro-Eating Disorder Social Media."
[57] Barakat et al., "Risk Factors for Eating Disorders"; Yurtdaş-Depboylu et al., "The Association between Social Media Addiction and Orthorexia Nervosa, Eating Attitudes, and Body Image among Adolescents."
[58] Ryding and Kuss, "The Use of Social Networking Sites, Body Image Dissatisfaction, and Body Dysmorphic Disorder."
[59] Costello et al., "Algorithms, Addiction, and Adolescent Mental Health."
[60] See, e.g., Coyne et al., "Suicide Risk in Emerging Adulthood"; Jarman et al., "Direct and Indirect Relationships between Social Media Use and Body Satisfaction."
[61] Brautsch et al., "Digital Media Use and Sleep in Late Adolescence and Young Adulthood"; Carter et al., "Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes."
[62] LeBourgeois et al., "Digital Media and Sleep in Childhood and Adolescence."
[63] Paruthi et al., "Consensus Statement of the American Academy of Sleep Medicine on the Recommended Amount of Sleep for Healthy Children."
[64] National Academies of Sciences, Engineering, and Medicine, "The Relation between Social Media and Health."
[65] See, e.g., Kreski et al., "Social Media Use and Depressive Symptoms Among United States Adolescents."

| | |
|---|---|
| | measures employed by studies.[66] One consistent finding is that minors who use social media at high levels tend to have more depressive symptoms.[67] |
| **Problematic use** | Problematic social media use occurs when it causes dysfunctions in everyday life activities, such as attending school and sleep,[68] and is associated with negative mental health outcomes, including depression.[69] While problematic social media use is not yet a formally recognized disorder by the International Classification of Diseases,[70] it may resemble a formal diagnosis, gaming disorder, both of which may in fact be subsets of a common disordered use of technology.[71] |

[66] National Academies of Sciences, Engineering, and Medicine, "The Relation between Social Media and Health."
[67] Shannon et al., "Problematic Social Media Use in Adolescents and Young Adults."
[68] Caplan, "Theory and Measurement of Generalized Problematic Internet Use"; Casale and Banchi, "Narcissism and Problematic Social Media Use."
[69] Bányai et al., "Problematic Social Media Use"; Paakkari et al., "Problematic Social Media Use and Health among Adolescents."
[70] Paschke et al., "ICD-11-Based Assessment of Social Media Use Disorder in Adolescents."
[71] Moreno et al., "Measuring Problematic Internet Use, Internet Gaming Disorder, and Social Media Addiction in Young Adults."

# Question 2: Best practices for mitigation measures

The design choices that platforms make contribute to systemic risk, as well as how risks are mitigated.[72] Design is increasingly viewed as a central place to ensure risks are anticipated and mitigated during product development.[73] The DSA Civil Society Coordination Group comments in response to the first round of systemic risk assessment highlights the importance of design-based mitigations in responding to systemic risk.[74] The Integrity Institute similarly focuses on the role of platform design in mitigation.[75] While a range of design choices may impact systemic risks, this section focuses on two specific strategies for effective mitigation of prominent and recurrent systemic risks: (A) the design of algorithmic recommender systems, and (B) ongoing experiential and observational monitoring of design risk.

## A. Recommender System Design

Recommender system design is a key site of concern for systemic risks. As discussed in our answer to Questions 1 and 3, research shows that recommender system design may contribute to a wide range of risks including threats to fundamental rights, the spread of illegal content, and problematic overuse or other harms to minors.

Risks may be experienced by both the general population as well as minors or other specific groups. This section identifies effective measures for mitigating systemic risks associated with recommender system design. These recommendations are described in detail in KGI's recent report, *Better Feeds: Algorithms That Put People First*,[76] KGI's *Better Feeds* EU Policy Brief,[77] and Panoptykon and People vs Big Tech's briefing *Safe by Default*.[78]

Recommender systems aligned with *long-term user value* help to manage systemic risks. Long-term user value aligns outcomes with users' deliberative, forward-looking aspirations or preferences. Finding out what users value over the long term and optimizing for it requires platforms to rely on data other than engagement. Platforms that are optimized to support long-term value to users may:

- ask users directly to state their explicit preferences;
- rely on surveys, quality indicators selected by the user, or predictions of each;
- rely on user engagement or behavioral signals that are deliberative, clearly express the user's desires, or require significant effort from the user; or

[72] Chapman, "Advancing Platform Accountability."
[73] Iyer, "Introducing the Neely Center Design Code for Social Media."
[74] Center for Democracy and Technology, "Civil Society Responds to DSA Risk Assessment Reports."
[75] Allen et al., "Risk Dimensions and Mitigation Effectiveness."
[76] Allen et al., "Better Feeds."
[77] Knight-Georgetown Institute, "Better Feeds EU Policy Brief."
[78] Szymielewicz, "Safe by Default."

- combine aspects of these or other approaches.

As described in the preceding section, optimizing recommender systems to maximize predicted short-term engagement can contribute to excessive or problematic overuse, and lead to the spread of sensational or borderline content that may exacerbate systemic risks. Focusing on long-term user value aligns the system design to users' aspirations and preferences, thereby better promoting fundamental rights and physical and mental health among users.

When recommender systems are designed for maximum engagement they typically do not promote long-term user value. The mere fact that recommended items succeed in engaging users is not sufficient to establish that recommendations align with long-term user value and effectively mitigate risks. Rather, predictions of long-term value must be supported by evidence of explicit, expressed desires held by individual users or representative subsets of users, not ambiguous behaviors that may correlate with dubious inferred "preferences."

The *Better Feeds* guidelines describe three key strategies for mitigating risks associated with recommender systems: design and public content transparency, better user choices and defaults, and assessments of long-term impact. The following guidelines can be seen as best practices for platform compliance with DSA requirements related to recommender systems.

### 1. Design and Public Content Transparency

Detailed disclosures about the design of recommender systems would allow the Commission, outside experts, and the European public to understand the tradeoffs being made in the design of these systems. Disclosures could motivate platforms to show how their system designs mitigate risks. The *Better Feeds* guidelines propose that platforms publicly disclose:

- information about the specific input data and weights used in their recommender systems;
- the metrics they use to measure long-term user value;
- the metrics they use to evaluate product teams responsible for recommender system design;
- a sample of the public content that is most highly disseminated on the platform and a sample of the public content that receives the highest engagement; and
- a representative sample of public content consumed during a typical user session.

These mitigation recommendations can inform implementation and operationalization of several obligations under the DSA. Article 27 requires platforms to disclose the main parameters of recommender systems in plain and intelligible language. This includes the "most significant" criteria for recommending information to users, as well as the "reasons for the relative importance of those parameters." These specifications can be interpreted in multiple ways, some of which yield qualitative descriptions of recommender system parameters that do not allow independent experts to examine or compare how these systems are optimized. The first round of recommender system audits reveal significant inconsistencies in how platforms interpret these definitions.[79] Independent analysis would

[79] Chapman, "Advancing Platform Accountability."

be more feasible if the Commission and platforms adopted the *Better Feeds* design transparency proposals as the baseline interpretation of "main parameters."

In order to promote transparency, Article 40 of the DSA requires platforms to enable the sharing of real-time publicly accessible content data with researchers working to identify and understand systemic risks. Analysis of publicly available data is particularly important for understanding recommender system design. The *Better Feeds* guidelines recommend that platforms proactively share:

- public content that is most highly disseminated on the platform;
- the provision of a sample of public content that receives the highest engagement; and
- a representative sample of public content consumed during a typical user session.

KGI is working with leading experts to develop a framework for what kind of platform data should be made publicly available, under what circumstances, and in what format.[80]

### 2. User Choices and Defaults

To help proactively mitigate systemic risks, all users should be defaulted into designs optimized for long-term value – even if it means sacrificing short-term engagement. Additional, enforceable user choices would allow individuals finer-grained control to tailor their platform experiences. The *Better Feeds* guidelines propose that:

- By default, platforms must optimize users' recommender systems to support long-term user value. If platforms have insufficient information about long-term value to minors specifically, they must default minors to non-personalized recommender systems.
- Platforms must provide easily accessible ways for users to set their preferences about the types of items to be recommended and blocked and honor those preferences.

Panoptykon's *Safe by default* guidelines align with these recommendations and suggest that:

- By default, platforms' recommender systems should *not* use behavioural profiling i.e. observing and collecting passive data about how users behave and interact on the platform in order to infer their interests. Instead, the default feed should only use as input signals and data actively provided by the user for this very purpose (e.g. interests declared by the user when building their profile), as well as explicit user feedback on specific content (e.g. "show me more/show me less" signal sent by clicking a relevant button).
- By default, platforms should introduce positive friction to give users a chance to think before posting, commenting, or sharing content. This includes "think before you share" messages and limits on resharing as well as a series of practical recommendations aimed at countering platform stickiness so that users are nudged towards disconnecting from social media rather than compulsively engaging, as well as being provoked to be more intentional about what they want to get out of a given social media session.
- Platforms should stop using addictive and deceptive design features such as: extensive notifications turned on by default, infinite scroll, and video autoplay.

[80] Knight-Georgetown Institute, "The Gold Standard for Publicly Available Platform Data."

- Platforms should create features that facilitate conscious, authentic personalisation of the feed by their users and protect their wellbeing. This includes a range of measures such as sliders to set different optimization goals for recommendations (e.g. more long-form vs short-form content, local vs global relevance, etc.), a 'hard stop' button to remove unwanted classifications of content from appearing altogether, a button to 'reset' an individual's feed, prompts to share declared interests, and settings to allow users to explore how their feed changes based on their choices and interactions.[81]
- The Commission should consider obliging platforms to open up their infrastructure to allow independent, third-party content curation and moderation services (sometimes called "algorithmic pluralism") as a mitigation measure.[82]

These proposals could be incorporated into platform implementation of DSA expectations. Indeed, the DSA requires platforms to modify their interfaces and recommender systems in order to mitigate systemic risks (Article 35(d)), which include, among others, risks related to civic discourse, media freedom and pluralism, and mental well-being (i.e., users' addiction). Algorithmic pluralism is a promising strategy for mitigating these systemic risks. Providing consumers with a marketplace of options for content curation tools would mean that large platforms are no longer the only arbiters of quality and credibility in ranking algorithms.

Article 38 requires large platforms to offer at least one recommender system option that is not based on user profiling. The *Better Feeds* guidelines propose that the non-profiling option be optimized for long-term user value, rather than non-personalized or chronological feeds. Panoptykon's *Safe by default* guidelines similarly recommend the use of input signals and data actively provided by the user, as well as explicit user feedback on specific content. This could be seen as best practice.

Finally, Article 28 requires platforms to take proportionate measures to ensure a high level of privacy, safety, and security for minors. Given the unique risks related to cognitive and social-emotional development of minors discussed above, evidence from the *Better Feeds* report suggest that optimizing minors' recommender systems for long-term value can be an effective way to mitigate some of the unique risks that minors face.

### 3. Assessments of Long-Term Impact

Platforms can only deliver long-term value to users if they continuously test the impact of algorithmic changes over time.

Online platforms may run thousands or tens of thousands of experiments each year to test out different design aspects of their systems, including changes to recommender system design.[83] Many of these experiments might last for days or weeks, after which time product teams evaluate their effectiveness against company-selected metrics and decide whether to maintain the changes, revert

[81] For examples see: McCrosky et al., "Prototyping User Empowerment."
[82] See Elsayed-Ali and Berjon, "Algorithmic Pluralism."
[83] See, e.g., Donovan, "The Role of Experimentation at Booking.com."

back, or continue experimenting. Over the course of a year, platforms might make dozens or hundreds of changes to the algorithms that power the recommendations each user sees.

Given the frequency of experimentation, many platforms also maintain a holdout group – a group of users that are exempt from having design changes applied to their accounts, and who function as a control group for comparison with the rest of the user base.[84] The holdout group size varies greatly from platform to platform. Most users never become aware that they are in a holdout group even though their user experience can vary significantly from all other users.

On some platforms, holdout experiments can be long-running, with users staying in the holdout group for years at a time. While most holdout experiments go undisclosed, in the past companies have voluntarily published the results of long-term holdouts demonstrating the effects of specific design choices on metrics of user welfare. For example, Meta and Twitter (now X) have previously shared the results of long-term holdouts that withheld users from receiving advertisements and personalized feeds, respectively.[85] These holdouts were run continuously for years, underpinning their usefulness for understanding the design features they examined. Running holdouts of this length should be considered a best practice for platform design and the ongoing mitigation of systemic risk.

If long-term holdout experiments were to become more institutionalized and more available for scrutiny by the Commission and the European public, they could become powerful tools to shift platforms' incentives towards designs that optimize for long term user retention, value, and satisfaction. As explained in the introduction to this section, users often make choices to engage with content in the moment that are not indicative of what the users aspire to or prefer in the long run, or what makes them happy about their overall experience with a platform later on. When platforms optimize for short-term engagement, they exacerbate this dynamic. If platforms were required to demonstrate long-term user value, they would design their recommender systems differently. Requiring long-term holdout experiments and the public disclosure of their results will incentivize platforms to give more priority to long-term user retention, value, and satisfaction.

The *Better Feeds* guidelines propose that:

- Platforms must run long-term (12-month or longer) holdout experiments on a continuous basis.
- Platforms must report the aggregate, anonymized results of the holdout experiments publicly.
- Holdout experiments must be subject to an audit by an independent third party.
- Platforms must measure the aggregate harms to at-risk populations that result from recommender systems and publicly disclose the results of those measurements.

[84] See, e.g., Pinterest Engineering, "How Holdout Groups Drive Sustainable Growth."
[85] Brynjolfsson et al., "The Consumer Welfare Effects of Online Ads"; Huszár et al., "Algorithmic Amplification of Politics on Twitter"; Mercer and Meakin, "The Power of Asymmetric Experiments @ Meta."

The results of long-term holdout experiments could serve as an integral step in how platforms assess the effects of their product designs on systemic risks under the DSA. Articles 34 and 35 establish specific requirements around systemic risk assessments and the implementation of proportionate and effective mitigation measures related to (among other design elements) their recommender systems. Article 37 mandates independent audits to assess compliance with these, and other, due diligence expectations. Results from long-term holdout experiments should serve as key data in these risk assessments and evidence of their mitigation strategies' overall effectiveness.

A key aspect of the measurements proposed in the *Better Feeds* guidelines is that they are designed to evaluate effects on populations and not on individuals – reflective of the types of harms that system architecture is capable of causing.[86] Because these types of assessments and metrics operate at the population level, they can generally be revealed without implicating the privacy of individual users.

When considering systemic risk and harms to at-risk populations, the Commission can communicate specific measurements requirements. While these may be tailored to specific platforms, examples include:

- *Unwanted or harmful contact and content*. These can be measured through surveys about negative experiences,[87] by tracking user behaviors that indicate negative experiences (e.g., hiding content, blocking users, reporting content), or by measuring engagement with content that violates platform policies or that is predicted to violate platform policies.

- *Unwanted or harmful usage*. To track harmful usage or sleep effects, platforms might track the percentage of users who use the platform for an excessive number of hours per day, during school and nighttime hours, or very frequently. Platforms can combine those metrics with survey data about unwanted usage, sleep, regret, and activity displacement to understand when users themselves feel that their usage is problematic.

- *Systemic harms*. These vary widely. For example, measurements of conflict or polarization on a platform might combine user surveys with engagement metrics that measure engagement with content identified as toxic towards outgroups.[88]

[86] Lubin et al., "Social Media Harm Abatement"; Lubin and Gilbert, "Accountability Infrastructure."
[87] For example, Meta has conducted internal research using its "Bad Experiences and Encounters Framework (BEEF)" survey to ask minors about various online harms, including hate speech and unwanted sexual advances. The findings of this research were not made public until they were disclosed as part of ongoing litigation with state Attorneys General. In the absence of transparency about internal survey results like this one, outside groups have launched projects such as the Neely Social Media Index to ask users about their negative experiences online. See Fast et al., "Unveiling the Neely Ethics & Technology Indices"; Horwitz, "His Job Was to Make Instagram Safe for Teens."
[88] Stray, "Dependent Variables."

## B. Experimental and Observational Monitoring

This section describes how measurement and transparency tools can help advance mitigation strategies throughout the product design and experimentation lifecycle. Platforms measure multiple variables when assessing design choices. Integrating risk assessment and mitigation measurement into the product development lifecycle can serve as a valuable complement to regular enterprise risk management, like the systemic risk assessment processes.[89]

A forthcoming article, *Social Media Harm Abatement: Mechanisms for Transparent Public Health Assessment*, offers recommendations to proactively mitigate platform design risk through the ongoing measurement of specific harms. The paper highlights the importance of integrating risk measurement into ongoing product assessments conducted by teams within companies as well as the ongoing monitoring of broader user experiences through longitudinal user studies, cross platform comparisons, among others.[90] *Abatement* makes several recommendations that are directly relevant for how platforms build internal mechanisms to respond to DSA risk assessment and mitigation requirements, including:

- *Measuring interface design and mitigation through universal holdouts*: *Abatement* recommends the use of holdout groups in relation to design. While the *Better Feeds* guidelines focused on the study of recommender systems, *Abatement* argues that holdout groups can improve understanding of broader design risks and mitigations. The results of holdout group experiences related to the full range of platform design decisions should be reported publicly through systemic risk assessments and audits, acting as a control group for interface and product design changes. This would enable the platform, the Commission, and external EU stakeholders to understand whether the totality of product changes are effectively mitigating risk.[91]

- *Improving operational and product experiment transparency*: *Abatement* recommends that platforms should publicly communicate high level goals of product teams and aggregated product design experimentation results. Sharing high level goals and metrics of platform design teams would allow stakeholders to understand how platform growth goals are aligned with risk prevention and mitigation goals.[92] Platform disclosure of an inventory of product design experimentation and results would enable the Commission and external EU stakeholders to understand the degree to which platforms are considering risks and externalities as they iterate design across the platform.[93] Platform compliance with systemic risk assessment, mitigation, and audit requirements could integrate greater transparency across product design goals and

[89] Lubin and Iyer, "How Tech Regulation Can Leverage Product Experimentation Results."
[90] See Lubin et al., "Social Media Harm Abatement." While this paper is focused on the United States litigation context, many of the recommendations related to the "internal mechanism" are directly relevant for proactive and ongoing management of prominent and recurrent systemic risks in Europe.
[91] Lubin et al., "Social Media Harm Abatement.
[92] Ibid.
[93] Ibid.

metrics. Companies deploy A/B tests to understand the potential impacts of a specific design change. These product-focused tests do not present a picture of how design changes across the platform aggregate to influence systemic risk.[94]

- *Tracking population-level effects*: Lastly, the *Abatement* paper calls for the integration of specific metrics to track risks on a longer-term basis across measurement strategies. While *Abatement* is focused on harms alleged in the US litigation context,[95] the recommended approach could inform how platforms assess systemic risk in Europe. At present, the annual systemic risk assessment and audits under Article 34 and 35 are the primary site through which risk is assessed. Platforms could significantly bolster these annual assessments by integrating a set of intermediate indicators to track longer-term outcomes associated with social media use. Specific indicators of risks and harm could be integrated into existing platform measurement infrastructure and reported out to the Commission through systemic risk assessment. Questions around harm and risk could be further integrated into existing platform surveys about user experiences,[96] by tracking user behaviors that indicate negative experiences, or by measuring engagement with content that violates platform policies or that is predicted to violate platform policies.

[94] Bojinov and Gupta, "Online Experimentation."
[95] Lubin et al., "Social Media Harm Abatement.
[96] Béjar, "Bad Experiences and Encounters Framework (BEEF) Survey."

# Question 3: Factors Influencing Systemic Risk

Platform design choices influence the full range of systemic risks under the DSA.[97] As articulated in KGI's previous submission to the Commission's consultation on Guidelines to Enforce the Protection of Minors Online,[98] design risks can include profile design features, usage-extending design features,[99] and recommender system design.[100] A 2023 Briefing Note from the Recommender Systems Task Force, led by Panoptykon, specifically analyzes how recommender systems may contribute to systemic risks.[101]

The Recommender Systems Task Force Briefing Note and *Better Feeds* both highlight how maximizing predicted engagement, in particular, may influence systemic risk.[102] Optimizing a recommender system for predicted engagement can lead to negative experiences including dissatisfaction with the overall amount of time spent on the platform, experience using the product, or both.[103]

Academic research has identified connections between various harms and recommender systems designed to maximize predicted engagement. However, this research faces important challenges, including lack of access to necessary data, inability to experiment with alternative designs in realistic settings, and legal risks associated with studying online platforms. These dynamics are explored further in our response to Question 4. Researchers have attempted to surmount these difficulties through various creative empirical methodologies, and though imperfect, findings have identified important factors which connect recommender system design to systemic risks enumerated in the DSA.

At a high level, these risks include increased polarization and reductions in trust, direct incitement to violence, harms to mental and physical health, and harms stemming from privacy invasion, commercial surveillance, and user profiling, among others.[104]

---

[97] Knight-Georgetown Institute, et al., "Taxonomy - Mapping Consumer Harm to Specific Social Media Design Elements."
[98] Chapman and Arnao, "KGI Comments to the European Commission on Guidelines to Enforce the Protection of Minors Online."
[99] Monge Roffarello, Lukoff, and De Russis, "Defining and Identifying Attention Capture Deceptive Designs in Digital Interfaces." Note that some research uses the term 'dark' patterns, but we use deceptive patterns.
[100] Allen et al., "Better Feeds."
[101] Szymielewicz and Głowacka, "Fixing Recommender Systems."
[102] Allen et al., "Better Feeds."
[103] See, e.g., Allcott et al., "Digital Addiction"; Cho et al., "Reflect, Not Regret."
[104] Bavel et al., "How Social Media Shapes Polarization"; Brailovskaia et al., "Experimental Longitudinal Evidence for Causal Role of Social Media Use and Physical Activity in COVID-19 Burden and Mental Health"; Park et al., "Global Mistrust in News"; United Nations Independent Investigative Mechanism for Myanmar, "Anti-Rohingya Hate Speech On Facebook"; Pasquale, *The Black Box Society*; Turow, *The Daily You*.

In particular, research points to a central role of recommender systems in extending social media use, which can negatively affect physical and mental health. Empirical literature has documented that among adolescents, extended use of social media (spurred by engagement-based designs) straightforwardly contributes to a decrease in time associated with healthier activities such as sleep.[105] When this happens, sleep may be disrupted through various mechanisms, including delayed and worsened quality, increased psychological stimulation before bedtime, and distorted circadian rhythms from light emissions.[106] Indeed, research has found that adolescents often report using social media late at night and losing track of time when doing so.[107] This occurrence is concerning because insufficient sleep can affect various other health issues, such as the likelihood of learning problems, depression, and suicidal ideation.[108] Engagement-based feeds may plausibly contribute to these outcomes in adolescents, although more research is needed to examine this connection.

[105] Alonzo et al., "Interplay between Social Media Use, Sleep Quality, and Mental Health in Youth"; Brautsch et al., "Digital Media Use and Sleep in Late Adolescence and Young Adulthood"; Carter et al., "Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes."
[106] LeBourgeois et al., "Digital Media and Sleep in Childhood and Adolescence."
[107] Common Sense Media, "Constant Companion."
[108] Paruthi et al., "Consensus Statement of the American Academy of Sleep Medicine on the Recommended Amount of Sleep for Healthy Children."

# Question 4: The importance of transparency and data access

The DSA's transparency and data access regime is essential for assessing systemic risks and the effectiveness of platform mitigations. The DSA includes specific requirements related to risk assessment in Article 34, audits in Article 37, and requirements to enable independent research with publicly accessible data under Article 40(12) and non-public data through vetted researchers under Article 40(4).[109]

Despite these important requirements, independent research focused on digital platforms is severely constrained. The first round of systemic risk assessments and audits have not revealed fundamentally new information on platform risks or the effectiveness mitigations measures.[110] In recent years, platforms have curtailed access to data that was previously accessible by independent researchers, including academics, civil society, and journalists. For example, Meta has shutdown access to CrowdTangle,[111] X took steps to restrict researcher API access,[112] Reddit updated access to its Data API with new rate limits,[113] and TikTok established researcher access through an API that researchers describe as "strict."[114] Civil society analysis of emerging platforms data access programs under the DSA has found significant gaps.[115] Indeed, the Commission has opened proceedings against multiple platforms for alleged shortcomings in enabling researcher access to data including investigations focused on X,[116] TikTok,[117] AliExpress,[118] and Facebook/Instagram.[119]

These research tools have been essential for improving understanding of our online information ecosystem. Meaningful independent analysis of platform risks and mitigations is vital for understanding prominent and recurring risks and mitigation measures. The Commission and European Board for Digital Services should work with key academic, civil society, and journalism stakeholders to guarantee meaningful researcher and public access to platform data.

---

[109] European Union, *Digital Services Act*.
[110] Center for Democracy and Technology, "Civil Society Responds to DSA Risk Assessment Reports."
[111] Center for American Progress et al., "CrowdTangle Letter"; Mozilla Foundation et al., "Open Letter To Meta."
[112] Coalition for Independent Technology Research et al., "Letter: Imposing Fees to Access the Twitter API Threatens Public-Interest Research."
[113] Reddit, "Creating a Healthy Ecosystem for Reddit Data and Reddit Data API Access."
[114] Counts, "TikTok's Rules Deter Researchers from Crunching Data on Users, Misinformation."
[115] Hickey et al., "Public Data Access Programs"; Gotfredsen and Dowling, "Meta Is Getting Rid of CrowdTangle — and Its Replacement Isn't As Transparent or Accessible."
[116] European Commission, "Commission Sends Preliminary Findings to X for Breach of DSA."
[117] European Commission, "Commission Opens Formal Proceedings against TikTok under the Digital Services Act."
[118] European Commission, "Commission Opens Formal Proceedings against AliExpress under the Digital Services Act."
[119] European Commission, "Commission Opens Formal Proceedings against Facebook and Instagram under the Digital Services Act."

# Contact

Please contact us should you have any questions.

**KGI:**

Alissa Cooper
Executive Director

Peter Chapman
Associate Director

knightgeorgetown@georgetown.edu

**Panoptykon Foundation:**

Katarzyna Szymielewicz
Advocacy & Strategy Director, President

Dorota Głowacka
Advocacy & Litigation Expert

fundacja@panoptykon.org

# Bibliography

5Rights Foundation. “Pathways: How Digital Design Puts Children at Risk.” 5Rights Foundation, September 17, 2021. https://5rightsfoundation.com/resource/pathways-how-digital-design-puts-children-at-risk/

Allcott, Hunt, Matthew Gentzkow, and Lena Song. “Digital Addiction.” *American Economic Review* 112, no. 7 (July 2022): 2424–63. https://doi.org/10.1257/aer.20210867.

Allen, Jeff. “Misinformation Amplification Analysis and Tracking Dashboard.” Integrity Institute, October 2022. https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard.

Allen, Jeff, Zander Arnao, Alissa Cooper, et al. “Better Feeds: Algorithms That Put People First.” Knight-Georgetown Institute, March 4, 2025. https://kgi.georgetown.edu/research-and-commentary/better-feeds/.

Allen, Jeff, Spencer Gurley, Sofia Bonilla, and Nick Shen. “Risk Dimensions and Mitigation Effectiveness.” Integrity Institute, February 2025. https://drive.google.com/file/d/1V42wIc-2NsP89AxLUqqY5OGQrVrzUY2y/view?usp=sharing.

Alonzo, Rea, Junayd Hussain, Saverio Stranges, and Kelly K. Anderson. “Interplay between Social Media Use, Sleep Quality, and Mental Health in Youth: A Systematic Review.” *Sleep Medicine Reviews* 56 (April 1, 2021). https://doi.org/10.1016/j.smrv.2020.101414.

American Psychological Association. “Potential Risks of Content, Features, and Functions: The Science of How Social Media Affects Youth.” American Psychological Association, April 2024. https://www.apa.org/topics/social-media-internet/youth-social-media-2024.

Bandy, Jack, and Tomo Lazovich. “Exposure to Marginally Abusive Content on Twitter.” *Proceedings of the International AAAI Conference on Web and Social Media* 17, no. 1 (June 2, 2023): 24–33. https://doi.org/10.1609/icwsm.v17i1.22123.

Bányai, Fanni, Ágnes Zsila, Orsolya Király, et al. “Problematic Social Media Use: Results from a Large-Scale Nationally Representative Adolescent Sample.” *PLOS ONE* 12, no. 1 (January 9, 2017). https://doi.org/10.1371/journal.pone.0169839.

Barakat, Sarah, Siân A. McLean, Emma Bryant, et al. “Risk Factors for Eating Disorders: Findings from a Rapid Review.” *Journal of Eating Disorders* 11, no. 1 (January 17, 2023). https://doi.org/10.1186/s40337-022-00717-4.

Bavel, Jay J. Van, Steve Rathje, Elizabeth Harris, Claire Robertson, and Anni Sternisko. “How Social Media Shapes Polarization.” *Trends in Cognitive Sciences* 25, no. 11 (November 1, 2021): 913–16. https://doi.org/10.1016/j.tics.2021.07.013.

Béjar, Arturo. "Bad Experiences and Encounters Framework (BEEF) Survey," 2021. https://storage.courtlistener.com/recap/gov.uscourts.nmd.496039/gov.uscourts.nmd.496039.37.2_1.pdf.

Bernstein, Gaia. *Unwired: Gaining Control over Addictive Technologies*. Cambridge University Press, 2023.

Biddle, Sam, Paulo Victor Ribeiro, and Tatiana Dias. "TikTok Told Moderators: Suppress Posts by the 'Ugly' and the Poor to Attract New Users." *The Intercept*, March 16, 2020. https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/.

Bojinov, Iavor, and Somit Gupta. "Online Experimentation: Benefits, Operational and Methodological Challenges, and Scaling Guide." *Harvard Data Science Review*, no. 4.3 (July 2022). https://hdsr.mitpress.mit.edu/pub/aj31wj81/release/1.

Brailovskaia, Julia, Verena J. Swarlik, Georg A. Grethe, Holger Schillack, and Jürgen Margraf. "Experimental Longitudinal Evidence for Causal Role of Social Media Use and Physical Activity in COVID-19 Burden and Mental Health." *Journal of Public Health* (September 2, 2022): 1–14. https://doi.org/10.1007/s10389-022-01751-x.

Brautsch, Louise AS., Lisbeth Lund, Martin M. Andersen, Poul J. Jennum, Anna P. Folker, and Susan Andersen. "Digital Media Use and Sleep in Late Adolescence and Young Adulthood: A Systematic Review." *Sleep Medicine Reviews* 68 (April 1, 2023). https://doi.org/10.1016/j.smrv.2022.101742.

Brynjolfsson, Erik, Avinash Collis, Asad Liaqat, et al. "The Consumer Welfare Effects of Online Ads: Evidence from a 9-Year Experiment." National Bureau of Economic Research, August 2024. https://doi.org/10.3386/w32846.

Caplan, Scott E. "Theory and Measurement of Generalized Problematic Internet Use: A Two-Step Approach." *Computers in Human Behavior* 26, no. 5 (September 1, 2010): 1089–97. https://doi.org/10.1016/j.chb.2010.03.012.

Carter, Ben, Philippa Rees, Lauren Hale, Darsharna Bhattacharjee, and Mandar S. Paradkar. "Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes: A Systematic Review and Meta-Analysis." *JAMA Pediatrics* 170, no. 12 (December 1, 2016): 1202–8. https://doi.org/10.1001/jamapediatrics.2016.2341.

Casale, Silvia, and Vanessa Banchi. "Narcissism and Problematic Social Media Use: A Systematic Literature Review." *Addictive Behaviors Reports* 11 (June 1, 2020). https://doi.org/10.1016/j.abrep.2020.100252.

Casey, B.j., Rebecca M. Jones, and Todd A. Hare. "The Adolescent Brain." *Annals of the New York Academy of Sciences* 1124, no. 1 (2008): 111–26. https://doi.org/10.1196/annals.1440.010.

Cataldo, Ilaria, Bruno Lepri, Michelle Jin Yee Neoh, and Gianluca Esposito. "Social Media Usage and Development of Psychiatric Disorders in Childhood and Adolescence: A Review." *Frontiers in Psychiatry* 11 (January 13, 2021). https://doi.org/10.3389/fpsyt.2020.508595.

Center for American Progress, Institute for Strategic Dialogue, Accountable Tech, et al. “CrowdTangle Letter,” May 2024. https://www.americanprogress.org/wp-content/uploads/sites/2/2024/05/CAP-ISD-letter2.pdf.

Center for Democracy and Technology. “Civil Society Responds to DSA Risk Assessment Reports: An Initial Feedback Brief.” Center for Democracy & Technology, March 17, 2025. https://cdt.org/insights/dsa-civil-society-coordination-group-publishes-an-initial-analysis-of-the-major-online-platforms-risks-analysis-reports/.

Center for Digital Health. “Stanford Youth Safety and Digital Wellbeing Report, 2025.” Stanford University, 2025. https://cdh.stanford.edu/sites/g/files/sbiybj29486/files/media/file/youth_safety_and_digital_wellbeing_report_2025.pdf.

Chapman, Peter. “Advancing Platform Accountability: The Promise and Perils of DSA Risk Assessments.” *Tech Policy Press*, January 9, 2025. https://techpolicy.press/advancing-platform-accountability-the-promise-and-perils-of-dsa-risk-assessments.

Chapman, Peter, and Zander Arnao. “KGI Comments to the European Commission on Guidelines to Enforce the Protection of Minors Online.” Knight-Georgetown Institute, October 29, 2024. https://kgi.georgetown.edu/research-and-commentary/kgi-comments-to-the-european-commission-on-guidelines-to-enforce-the-protection-of-minors-online/.

Chen, Yixin, Yue Fu, Zeya Chen, Jenny Radesky, and Alexis Hiniker. “The Engagement-Prolonging Designs Teens Encounter on Very Large Online Platforms.” January 29, 2025. https://doi.org/10.48550/arXiv.2411.12083.

Cho, Hyunsung, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. “Reflect, Not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis.” In *Proceedings of the ACM on Human-Computer Interaction*, 2021. https://doi.org/10.1145/3479600.

Cingel, Drew P., Alexis R. Lauricella, Lauren B. Taylor, Hannah R. Stevens, Sarah M. Coyne, and Ellen Wartella. “U.S. Adolescents’ Attitudes toward School, Social Connection, Media Use, and Mental Health during the COVID-19 Pandemic: Differences as a Function of Gender Identity and School Context.” *PLOS ONE* 17, no. 10 (October 27, 2022). https://doi.org/10.1371/journal.pone.0276737.

Coalition for Independent Technology Research, Citizens and Technology Lab, Knight First Amendment Institute at Columbia University, et al. “Letter: Imposing Fees to Access the Twitter API Threatens Public-Interest Research,” February 2023. https://independenttechresearch.org/letter-twitter-api-access-threatens-public-interest-research/.

Common Sense Media. “Constant Companion: A Week in the Life of a Young Person’s Smartphone Use,” September 26, 2023.

https://www.commonsensemedia.org/research/constant-companion-a-week-in-the-life-of-a-young-persons-smartphone-use.

Costello, Nancy, Rebecca Sutton, Madeline Jones, et al. "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, no. 2–3 (July 2023): 135–72. https://doi.org/10.1017/amj.2023.25.

Counts, Aisha. "TikTok's Rules Deter Researchers from Crunching Data on Users, Misinformation." *Bloomberg*, September 2023. https://www.bloomberg.com/news/articles/2023-09-21/tiktok-terms-of-service-strict-for-researchers.

Coyne, Sarah M., Jeffrey L. Hurst, W. Justin Dyer, et al. "Suicide Risk in Emerging Adulthood: Associations with Screen Time over 10 Years." *Journal of Youth and Adolescence* 50, no. 12 (December 1, 2021): 2324–38. https://doi.org/10.1007/s10964-020-01389-6.

Crone, Eveline A., and Elly A. Konijn. "Media Use and Brain Development during Adolescence." *Nature Communications* 9, no. 1 (February 21, 2018): 1–10. https://doi.org/10.1038/s41467-018-03126-x.

Cunningham, Tom, Sana Pandey, Leif Sigerson, et al. "What We Know About Using Non-Engagement Signals in Content Ranking." February 9, 2024. https://doi.org/10.48550/arXiv.2402.06831.

de Vries, Dian A., A. Marthe Möller, Marieke S. Wieringa, Anniek W. Eigenraam, and Kirsten Hamelink. "Social Comparison as the Thief of Joy: Emotional Consequences of Viewing Strangers' Instagram Posts." *Media Psychology* 21, no. 2 (April 3, 2018): 222–45. https://doi.org/10.1080/15213269.2016.1267647.

Dijkslag, I. R., L. Block Santos, G. Irene, and P. Ketelaar. "To Beautify or Uglify! The Effects of Augmented Reality Face Filters on Body Satisfaction Moderated by Self-Esteem and Self-Identification." *Computers in Human Behavior* 159 (October 1, 2024). https://doi.org/10.1016/j.chb.2024.108343.

Donovan, Nicola. "The Role of Experimentation at Booking.com." *Click Magazine*, September 2, 2019. https://partner.booking.com/en-us/click-magazine/industry-perspectives/role-experimentation-bookingcom.

Elsayed-Ali, Sherif, and Robin Berjon. "Algorithmic Pluralism: Towards Competitive & Innovative Information Ecosystems." Future of Technology Institute, March 2025. https://drive.google.com/file/d/1DuZsJ-IScqt4h_zpWaSBGffhIIr0V38q/view.

eSafety Commissioner. "Safety by Design Overview," May 2019. https://www.esafety.gov.au/sites/default/files/2019-10/SBD%20-%20Overview%20May19.pdf?v=1734548592263.

European Commission. "Commission Opens Formal Proceedings against AliExpress under the Digital Services Act," May 2024. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485.

———. “Commission Opens Formal Proceedings against Facebook and Instagram under the Digital Services Act,” April 2024. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373.

———. “Commission Opens Formal Proceedings against TikTok under the Digital Services Act,” February 2024. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926.

———. “Commission Sends Preliminary Findings to X for Breach of DSA,” July 2024. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761.

———. “Delegated Regulation on Data Access Provided for in the Digital Services Act,” May 2023. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/feedback_en?p_id=32045757.

———. “Status Report: Mechanisms for Researcher Access to Online Platform Data,” April 2024. https://digital-strategy.ec.europa.eu/en/library/status-report-mechanisms-researcher-access-online-platform-data.

European Union. Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), 2022. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

Fast, Nathanael, Juliana Schroeder, Ravi Iyer, and Matt Motyl. “Unveiling the Neely Ethics & Technology Indices.” *Designing Tomorrow*, June 22, 2023. https://psychoftech.substack.com/p/unveiling-the-neely-ethics-and-technology.

FBI San Francisco. “FBI San Francisco Warns of Increase in Sextortion Schemes Targeting Young Boys.” Federal Bureau of Investigation, May 2022. https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/press-releases/fbi-san-francisco-warns-of-increase-in-sextortion-schemes-targeting-young-boys.

Fitzsimmons-Craft, Ellen E., Melissa J. Krauss, Shaina J. Costello, Glennon M. Floyd, Denise E. Wilfley, and Patricia A. Cavazos-Rehg. “Adolescents and Young Adults Engaged with Pro-Eating Disorder Social Media: Eating Disorder and Comorbid Psychopathology, Health Care Utilization, Treatment Barriers, and Opinions on Harnessing Technology for Treatment.” *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity* 25, no. 6 (December 1, 2020): 1681–92. https://doi.org/10.1007/s40519-019-00808-3.

Gotfredsen, Sarah Grevy, and Kaitlyn Dowling. “Meta Is Getting Rid of CrowdTangle — and Its Replacement Isn’t As Transparent or Accessible.” *Proof*, July 9, 2024. https://www.proofnews.org/meta-is-getting-rid-of-crowdtangle-and-its-replacement-isnt-as-transparent-or-accessible/.

Griffiths, Scott, Emily A. Harris, Grace Whitehead, Felicity Angelopoulos, Ben Stone, Wesley Grey, and Simon Dennis. “Does TikTok Contribute to Eating Disorders? A Comparison of the TikTok Algorithms Belonging to Individuals with Eating Disorders versus Healthy Controls.” *Body Image* 51 (December 1, 2024). https://doi.org/10.1016/j.bodyim.2024.101807.

Hadjipanayis, Adamos, Elisavet Efstathiou, Peter Altorjai, et al. "Social Media and Children: What Is the Paediatrician's Role?" *European Journal of Pediatrics* 178, no. 10 (October 2019): 1605–12. https://doi.org/10.1007/s00431-019-03458-w.

HateAid. "Boundless Hate on the Internet – Dramatic Situation across Europe." HateAid, November 3, 2021. https://hateaid.org/en/eu-survey-boundless-hate-on-the-internet/.

Hickey, Cameron, Kaitlyn Dowling, Isabella Navia, and Claire Pershan. "Public Data Access Programs - A First Look." Mozilla Foundation, August 8, 2024. https://foundation.mozilla.org/en/blog/new-research-tech-platforms-data-access-initiatives-vary-widely/.

Holland, Grace, and Marika Tiggemann. "A Systematic Review of the Impact of the Use of Social Networking Sites on Body Image and Disordered Eating Outcomes." *Body Image* 17 (June 1, 2016): 100–110. https://doi.org/10.1016/j.bodyim.2016.02.008.

Horwitz, Jeff. "His Job Was to Make Instagram Safe for Teens. His 14-Year-Old Showed Him What the App Was Really Like." *Wall Street Journal*, November 3, 2023. https://www.wsj.com/tech/instagram-facebook-teens-harassment-safety-5d991be1.

Huszár, Ferenc, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. "Algorithmic Amplification of Politics on Twitter." *Proceedings of the National Academy of Sciences* 119, no. 1 (January 4, 2022). https://doi.org/10.1073/pnas.2025334119.

Information Commissioner's Office. "Age Appropriate Design: A Code of Practice for Online Services." Information Commissioner's Office. Accessed April 4, 2025. 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/.

Iyer, Ravi. "Introducing the Neely Center Design Code for Social Media." *USC Neely Center Newsletter*, October 18, 2023. https://uscneelycenter.substack.com/p/introducing-the-neely-center-design.

Jarman, Hannah, Sian McLean, Amy Slater, Matthew Marques, and Susan Paxton. "Direct and Indirect Relationships between Social Media Use and Body Satisfaction: A Prospective Study among Adolescent Boys and Girls." *New Media & Society* 26, no. 1 (November 2021). https://journals.sagepub.com/doi/10.1177/14614448211058468.

Jeff Horwitz and Deepa Seetharaman. "Facebook Executives Shut Down Efforts to Make the Site Less Divisive." *The Wall Street Journal*, May 26, 2020. https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499.

Knight-Georgetown Institute. "About." Knight-Georgetown Institute. Accessed April 7, 2025. https://kgi.georgetown.edu/about/.

———. “Better Feeds EU Policy Brief,” April 2025. https://kgi.georgetown.edu/wp-content/uploads/2025/03/Better-Feeds-EU-Policy-Brief-_Knight-Georgetown-Institute.pdf.

———. “The Gold Standard for Publicly Available Platform Data.” Accessed April 4, 2025. https://kgi.georgetown.edu/expert-working-groups/gold-standard-expert-working-group/gold-standard-faq/.

Knight-Georgetown Institute, USC Neely Center, and Tech Justice Law Project. “Taxonomy - Mapping Consumer Harm to Specific Social Media Design Elements,” October 2024. https://docs.google.com/spreadsheets/d/1GVO7sNuCNmNwqVK64PHQI7wxd8-Gmr9PqdkW12elmus/edit?gid=941162555#gid=941162555.

Kreski, Noah, Jonathan Platt, Caroline Rutherford, et al. “Social Media Use and Depressive Symptoms Among United States Adolescents.” *Journal of Adolescent Health* 68, no. 3 (March 1, 2021): 572–79. https://doi.org/10.1016/j.jadohealth.2020.07.006.

Kubin, Emily, and Christian von Sikorski. “The Role of (Social) Media in Political Polarization: A Systematic Review.” *Annals of the International Communication Association* 45, no. 3 (July 3, 2021): 188–206. https://doi.org/10.1080/23808985.2021.1976070.

Langvardt, Kyle. “Regulating Habit-Forming Technology.” *Fordham Law Review* 88, no. 1 (October 1, 2019).

LeBourgeois, Monique K., Lauren Hale, Anne-Marie Chang, Lameese D. Akacem, Hawley E. Montgomery-Downs, and Orfeu M. Buxton. “Digital Media and Sleep in Childhood and Adolescence.” *Pediatrics* (November 1, 2017): 92–96. https://doi.org/10.1542/peds.2016-1758J.

Lubin, Nathaniel, and Thomas Krendl Gilbert. “Accountability Infrastructure: How to Implement Limits on Platform Optimization to Protect Population Health,” June 12, 2023. https://www.platformaccountability.com/_files/ugd/424593_5af63341c56d48c4807afef21298f6f7.pdf.

Lubin, Nathaniel, and Ravi Iyer. “How Tech Regulation Can Leverage Product Experimentation Results.” *Lawfare*, July 11, 2023. https://www.lawfaremedia.org/article/how-tech-regulation-can-leverage-product-experimentation-results.

Lubin, Nathaniel, Yuning Liu, Amanda Yarnell, et al. “Social Media Harm Abatement: Mechanisms for Transparent Public Health Assessment.” *Annuls of the New York Academy of Science (Forthcoming)* (2025). https://arxiv.org/abs/2503.10458.

Lukoff, Kai, Ulrik Lyngs, Himanshu Zade, et al. “How the Design of YouTube Influences User Sense of Agency.” In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–17. Association for Computing Machinery, 2021. https://doi.org/10.1145/3411764.3445467.

McComb, Carly A., Vanman, Eric J., and Stephanie J. and Tobin. “A Meta-Analysis of the Effects of Social Media Exposure to Upward Comparison Targets on Self-Evaluations and Emotions.” *Media Psychology* 26, no. 5 (September 3, 2023): 612–35. https://doi.org/10.1080/15213269.2023.2180647.

McCrosky, Jesse, and Brandi Geurkink. “YouTube Regrets.” Mozilla Foundation, July 2021. https://foundation.mozilla.org/en/youtube/findings/.

McCrosky, Jesse, Tanya O’Carroll, Caroline Sinders, and Katarzyna Szymielewicz. “Prototyping User Empowerment: Towards DSA-Compliant Recommender Systems.” Panoptykon Foundation, August 2023. https://panoptykon.org/sites/default/files/2023-11/peoplevsbigtech_panoptykon_prototyping-empowerment_brief_20112023.pdf.

Mercer, Joe, and John Meakin. “The Power of Asymmetric Experiments @ Meta.” *Analytics @ Meta*, March 27, 2025. https://medium.com/@AnalyticsAtMeta/the-power-of-asymmetric-experiments-meta-8a8030d68c31.

Meta. “About Quality Ranking.” Meta Business Help Center. Accessed April 4, 2025. https://www.facebook.com/business/help/303639570334185.

———. “Facebook Feed AI System,” March 2025. https://transparency.meta.com/features/explaining-ranking/fb-feed/.

———. “Our Approach to Explaining Ranking.” Transparency Center, December 31, 2023. https://transparency.meta.com/features/explaining-ranking.

Milli, Smitha, Micah Carroll, Yike Wang, Sashrika Pandey, Sebastian Zhao, and Anca D. Dragan. “Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media.” December 7, 2024. https://doi.org/10.48550/arXiv.2305.16941.

Moehring, Alex. “Personalization, Engagement, and Content Quality on Social Media: An Evaluation of Reddit’s News Feed.” OSF Preprints, May 30, 2024. https://doi.org/10.31219/osf.io/8yuwe.

Moreno, Megan, Karyn Riddle, Marina C. Jenkins, Ajay Paul Singh, Qianqian Zhao, and Jens Eickhoff. “Measuring Problematic Internet Use, Internet Gaming Disorder, and Social Media Addiction in Young Adults: Cross-Sectional Survey Study.” *JMIR Public Health and Surveillance* 8, no. 1 (January 27, 2022). https://doi.org/10.2196/27719.

Mozilla Foundation, #Jesuislà, 7amleh, et al. “Open Letter To Meta: Support CrowdTangle Through 2024 and Maintain CrowdTangle Approach,” 2024. https://foundation.mozilla.org/en/campaigns/open-letter-to-meta-support-crowdtangle-through-2024-and-maintain-crowdtangle-approach/.

National Academies of Sciences, Engineering, and Medicine. *Social Media and Adolescent Health*. National Academies Press, 2024. https://doi.org/10.17226/27396.

———. "The Relation between Social Media and Health." In *Social Media and Adolescent Health*, 91–136. National Academies Press, 2024. https://nap.nationalacademies.org/read/27396/chapter/6.

Neechushtai, Efrat, Rodrigo Zamith, Seth C. Lewis. "More of the Same? Homogenization in News Recommendations When Users Search on Google, YouTube, Facebook, and Twitter." *Mass Communication and Society* 27, No. 6 (March 23, 2023): Pages 1309-1335. https://doi.org/10.1080/15205436.2023.2173609.

Nesi, Jacqueline, and Mitchell J. Prinstein. "Using Social Media for Social Comparison and Feedback-Seeking: Gender and Popularity Moderate Associations with Depressive Symptoms." *Journal of Abnormal Child Psychology* 43, no. 8 (November 1, 2015): 1427–38. https://doi.org/10.1007/s10802-015-0020-0.

New Mexico Department of Justice. "Attorney General Raúl Torrez Files Unredacted Complaint Against Snapchat, Exposing Internal Messages That Snap Knowingly Contributed to Harm Amongst Children," October 2024. https://nmdoj.gov/press-release/31302/.

Nobel, Safiya. *Algorithms of Oppression. How Search Engines Reinforce Racism*. NYU Press, 2018.

Office of the Surgeon General. "Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory." *Publications and Reports of the Surgeon General*, 2023. http://www.ncbi.nlm.nih.gov/books/NBK594761/.

Organisation for Economic Co-operation and Development. "Towards Digital Safety by Design for Children," June 2024. https://one.oecd.org/document/DSTI/CDEP(2023)13/FINAL/en/pdf.

Paakkari, Leena, Jorma Tynjälä, Henri Lahti, Kristiina Ojala, and Nelli Lyyra. "Problematic Social Media Use and Health among Adolescents." *International Journal of Environmental Research and Public Health* 18, no. 4 (January 2021). https://doi.org/10.3390/ijerph18041885.

Panoptykon Foundation. "What is Panoptykon." Panoptykon Foundation. Accessed April 7, 2025. https://en.panoptykon.org/about-us/what-is-panoptykon.

Park, Sora, Caroline Fisher, Terry Flew, and Uwe Dulleck. "Global Mistrust in News: The Impact of Social Media on Trust." *International Journal on Media Management* 22, no. 2 (April 2, 2020): 83–96. https://doi.org/10.1080/14241277.2020.1799794.

Paruthi, Shalini, Lee J. Brooks, 'Ambrosio Carolyn D, et al. "Consensus Statement of the American Academy of Sleep Medicine on the Recommended Amount of Sleep for Healthy Children: Methodology and Discussion." *Journal of Clinical Sleep Medicine* 12, no. 11 (November 2016): 1549–61. https://doi.org/10.5664/jcsm.6288.

Paschke, Kerstin, Maria Isabella Austermann, and Rainer Thomasius. "ICD-11-Based Assessment of Social Media Use Disorder in Adolescents: Development and Validation of the Social Media Use Disorder Scale for Adolescents." *Frontiers in Psychiatry* 12 (April 22, 2021). https://doi.org/10.3389/fpsyt.2021.661483.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2016.

Pescott, Claire Kathryn. "'I Wish I Was Wearing a Filter Right Now': An Exploration of Identity Formation and Subjectivity of 10- and 11-Year Olds' Social Media Use." *Social Media + Society* 6, no. 4 (October 1, 2020). https://doi.org/10.1177/2056305120965155.

Piccardi, Tiziano, Martin Saveski, Chenyan Jia, Jeffrey T. Hancock, Jeanne L. Tsai, and Michael Bernstein. "Social Media Algorithms Can Shape Affective Polarization via Exposure to Antidemocratic Attitudes and Partisan Animosity." November 22, 2024. https://doi.org/10.48550/arXiv.2411.14652.

Pinterest Engineering. "How Holdout Groups Drive Sustainable Growth." *Pinterest Engineering Blog*, February 17, 2017. https://medium.com/pinterest-engineering/how-holdout-groups-drive-sustainable-growth-35a4786c3801.

Pizzo Frey, Tracy, Jason Mills, Margaret Mitchell, et al. "Recommendation Systems in Social Media." Common Sense Media, October 7, 2024. https://www.commonsensemedia.org/ai-ratings/recommendation-systems-in-social-media.

Raffile, Paul, Alex Goldenberg, Cole McCann, and Joel Finkelstein. "A Digital Pandemic: Uncovering the Role of 'Yahoo Boys' in the Surge of Social Media-Enabled Financial Sextortion Targeting Minors." Network Contagion Research Institute, January 2024. https://networkcontagion.us/reports/yahoo-boys/.

Reddit. "Creating a Healthy Ecosystem for Reddit Data and Reddit Data API Access." *Upvoted*, April 2023. https://redditinc.com/blog/2023apiupdates.

Reid Chassiakos, Yolanda Linda, Jenny Radesky, Dimitri Christakis, Megan A. Moreno, and Corinn Cross. "Children and Adolescents and Digital Media." *Pediatrics* 138, no. 5 (November 2016). https://doi.org/10.1542/peds.2016-2593.

Ryan-Mosley, Tate. "Beauty Filters Are Changing the Way Young Girls See Themselves." MIT Technology Review, April 2021. https://www.technologyreview.com/2021/04/02/1021635/beauty-filters-young-girls-augmented-reality-social-media/.

Ryding, Francesca C., and Daria J. Kuss. "The Use of Social Networking Sites, Body Image Dissatisfaction, and Body Dysmorphic Disorder: A Systematic Review of Psychological Research." *Psychology of Popular Media* 9, no. 4 (2020): 412–35. https://doi.org/10.1037/ppm0000264.

Schroeder, Makenzie, and Elizabeth Behm-Morawitz. "Digitally Curated Beauty: The Impact of Slimming Beauty Filters on Body Image, Weight Loss Desire, Self-Objectification, and Anti-Fat Attitudes." *Computers in Human Behavior* 165 (April 1, 2025). https://doi.org/10.1016/j.chb.2024.108519.

Shannon, Holly, Katie Bush, Paul J. Villeneuve, Kim GC Hellemans, and Synthia Guimond. "Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-Analysis." *JMIR Mental Health* 9, no. 4 (April 14, 2022). https://doi.org/10.2196/33450.

Smith, Ben. "How TikTok Reads Your Mind." *The New York Times*, December 6, 2021. https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html.

Somerville, Leah H. "The Teenage Brain: Sensitivity to Social Evaluation." *Current Directions in Psychological Science* 22, no. 2 (April 1, 2013): 121–27. https://doi.org/10.1177/0963721413476512.

Stray, Jonathan. "Dependent Variables: The Outcomes We Will Measure." *The Prosocial Ranking Challenge*, February 2, 2024. https://rankingchallenge.substack.com/p/dependent-variables-the-outcomes.

Superior Court of California. "Neville v. Snap Inc.," January 2, 2024. https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=3828&context=historical.

Szymielewicz, Katarzyna. "Safe by Default: Moving Away from Engagement-Based Rankings towards Safe, Rights-Respecting, and Human Centric Recommender Systems." Panoptykon Foundation, March 2024. https://panoptykon.org/sites/default/files/2024-03/panoptykon_peoplevsbigtech_safe-by-default_briefing_03032024.pdf.

Szymielewicz, Katarzyna, and Dorota Głowacka. "Fixing Recommender Systems: From Identification of Risk Factors to Meaningful Transparency and Mitigation." Panoptykon Foundation, August 2023. https://panoptykon.org/sites/default/files/2023-08/Panoptykon_ICCL_PvsBT_Fixing-recommender-systems_Aug%202023.pdf.

Thiel, David, Renee DiResta, and Alex Stamos. "Addressing the Distribution of Illicit Sexual Content by Minors Online." Stanford Cyber Policy Center, June 6, 2023. https://cyber.fsi.stanford.edu/news/addressing-distribution-illicit-sexual-content-minors-online.

Turow, Joseph. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. Yale University Press, 2012.

United Nations Independent Investigative Mechanism for Myanmar. "Anti-Rohingya Hate Speech On Facebook: Content and Network Analysis." March 2024. https://iimm.un.org/wp-content/uploads/2024/03/Hate-Speech-Report_EN.pdf.

United States District Court for the Northern District of California. "In Re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation." Accessed February 13, 2025. https://cand.uscourts.gov/in-re-social-media-adolescent-addiction-personal-injury-products-liability-litigation-mdl-no-3047/.

University of Leiden and the Waag organisation. "Code for Children's Rights." Accessed April 6, 2025. https://codevoorkinderrechten.nl/wp-content/uploads/2022/02/Code-voor-Kinderrechten-EN.pdf.

Vandenbosch, Laura, Jasmine Fardouly, and Marika Tiggemann. "Social Media and Body Image: Recent Trends and Future Directions." *Current Opinion in Psychology* 45 (June 1, 2022). https://doi.org/10.1016/j.copsyc.2021.12.002.

Weinstein, Emily. "Adolescents' Differential Responses to Social Media Browsing: Exploring Causes and Consequences for Intervention." *Computers in Human Behavior* 76 (November 1, 2017): 396–405. https://doi.org/10.1016/j.chb.2017.07.038.

Wu, Tim. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Vintage, 2016.

X. "Heavy Ranker." GitHub, 2023. https://github.com/twitter/the-algorithm-ml/blob/main/projects/home/recap/README.md.

———. "Twitter's Recommendation Algorithm." *X Engineering*, March 31, 2023. https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm .

Yau, Joanna C., Stephanie M. Reich, and Tao-Yi Lee. "Coping With Stress Through Texting: An Experimental Study." *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine* 68, no. 3 (March 2021): 565–71. https://doi.org/10.1016/j.jadohealth.2020.07.004.

Yurtdaş-Depboylu, Gamze, Gülşah Kaner, and Semiha Özçakal. "The Association between Social Media Addiction and Orthorexia Nervosa, Eating Attitudes, and Body Image among Adolescents." *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity* 27, no. 8 (December 1, 2022): 3725–35. https://doi.org/10.1007/s40519-022-01521-4.

Zhang, Yun, Hongyan Wang, Chuan Luo, and Siyu Chen. "Ephemerality in Social Media: Unpacking the Personal and Social Characteristics of Time Limit Users on WeChat Moments." *Frontiers in Psychology* 12 (September 6, 2021). https://doi.org/10.3389/fpsyg.2021.712440.