Comments of Common Sense Media on the European Commission Consultation on the Digital Services Act

September 8, 2020

Executive Summary

Common Sense Media (Common Sense) is pleased to submit these comments in response to the European Commission's (Commission) Consultation on a proposed Digital Services Act package. Common Sense is an independent, nonpartisan voice for children, working to ensure that every child has the opportunity to thrive in the 21st century.

Online platforms take advantage of young people, exposing them to bullying, harassment, and hate as well as far too many inappropriate ads and unfair commercial practices.

We encourage the Commission to:

1. **Identify and limit the reach and influence of certain harmful users, ideologies, or trending topics--even if they are not "illegal".** The Digital Services Act consultation should be considered against the backdrop of existing regulatory efforts in the European Union and in member states to address hate speech and online terrorist content.[1] For example, the Terrorist Content Directive lays out a definition of content that solicits or supports terrorist offenses.[2] The UK government's White Paper on Online Harms lays out a non-exhaustive but detailed list of online harms with a special focus on children.[3] Legislation introduced in California would define political misinformation.[4] In addition, most of the major platforms announced firm policies to address the spread of false and misleading information about the COVID-19 pandemic.[5] The Commission should therefore look to existing rules and proposals to ensure standard definitions regarding problematic content. Such definitions should be expanded as needed by the Commission to address legal-but-harmful content for minors (such as certain commercial content) as well as other content which platforms have themselves identified as problematic.

[1] *See* European Commission, Summary report of the public consultation on measures to further improve the effectiveness of the fight against illegal content online (12 Sept. 2018); European Parliamentary Research Service, Terrorist Content Online: Tackling online terrorist propaganda (Mar. 2020). *See also* German Netzwerkdurchsetzungsgesetz (NetzDG) Network Enforcement Law

[2] European Parliament, Terrorist content online should be removed within one hour, says EP | News, (17 April 2019)

[3] Department for Digital, Culture, Media & Sport, Online Harms White Paper, (12 February 2020)

[4] AB 2885 False campaign speech and online platform disclosures, (21 February 2020)

[5] Joint industry statement from Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube (16 Mar. 2020).

2. **The Commission should ensure that platforms appropriately resource efforts to identify and limit amplification of problematic content,** for example by requiring that platforms spend a certain percentage of revenue on content moderation, and that moderation be conducted in-house with trained and supported staff and with the support of external experts. The Commission could also audit platforms on an ongoing basis to ensure these efforts are adequately resourced.
3. **Reduce reliance on algorithmically curated online feeds.** While even traditional news publishers seek to present the most relevant, topical, or engaging content to online users, opaque algorithmic curation has been repeatedly found to increase user engagement and the time spent on a platform by promoting salacious and emotionally-engaging material. Users are not provided sufficient choices in this regard. The Commission should require that algorithmically curated feeds are turned off by default. In addition, the Commission should ensure that platforms limit or cease algorithmic targeting of advertisements that can have significant effects--such as political ads. For example, in the U.S., lawmakers have introduced bills that would ban microtargeting of political ads. And the Commission should require that platforms amplify and share with caution, asking that platforms limit the speed and amount of content posted and shared to a level that can be responsibly overseen with appropriate human and/or automated review.
4. **Respect users.** The Commission should ensure that platforms do not manipulate users, especially children, into taking actions they do not want. Platforms should not subvert user-autonomy, decision-making, or choice. The Commission should look to the UK Age Appropriate Design Code, which offers an excellent example for how platforms can respect children's evolving capacities and prioritize children's best interests.[6] Similarly, as proposed in the U.S., the Commission should prohibit manipulative design features that keep kids glued to screens or that trick children into sharing data or making online purchases, such as rewards for watching or purchasing or autoplay.[7] The Commission should also support professional standards bodies that can provide guidance on design practices that undermine user autonomy and place limits on online behavioral and psychological experiments.[8]
5. **Working with relevant data protection authorities, the Commission should prioritize enforcing the GDPR,** whose principles, if actually followed, would cut down on profiling and targeting and undesired and inappropriate data processing. If users could use mapping apps for directions, or a smart home device for managing a home, without that data being collected, analyzed, and also used by a platform for unrelated purposes like advertising on different platform services, platforms would be less able to take advantage and profit from inappropriate market dominance. As the Commission has

[6] Information Commissioner's Office, Age appropriate design: a code of practice for online services, (12 August 2020).

[7] S.3411 Kids Internet Design and Safety (KIDS) Act, (5 March 2020)

[8] Deceptive Experiences To Online Users Reduction (DETOUR) Act (2019)

noted, the GDPR not only empowers EU residents but promotes trustworthy innovation via important principles such as data protection by design and by default.[9]

6. **Learn from the advertising ecosystem and limit advertising abuses.** The online ad space is rife with abuse. Ads profiting off of illegal content and political misinformation threaten democracy for all families. More transparency and safeguards in ad placement will help limit abuse and enable better understanding of platforms' data collection and targeting capabilities.[10] The Commission should require that platforms enable access to digital libraries of past advertisements, including any audiences targeted. And the Commission should ensure that independent experts are able to assess and understand targeting algorithms.
7. **Improve engagement among regulators, researchers, and other civil society stakeholders to identify harmful trends.** Platforms have increasingly sought to develop internal policies on how they choose to amplify individual accounts, content, and trends on their platform. External experts should be engaged in these discussions to ensure accountability and that the best interests of individuals--not just companies--are kept in mind. The Commission should require that platforms engage outside experts to assess moderation policies and the spread of misinformation, hate, and harmful content. Further, it should require that platforms make available an application programming interface or other technical capabilities to qualified, third party researchers to enable an independent analysis of any bias or unlawful discrimination in algorithms that support targeted advertising.

I. Introduction

Common Sense is pleased to submit these comments in response to the Commission's Consultation on a proposed Digital Services Act package.

Common Sense launched in the United States over 15 years ago, and established a presence in the United Kingdom in 2019. Common Sense has helped millions of families and kids think critically and make smart, responsible choices about the media they create and consume and the online experiences they participate in. We are the leading organization in the United States that parents, teachers, and policymakers go to for unbiased information, trusted advice, and innovative tools to harness the power of media and technology as a positive force in all children's lives. We have established the largest and most trusted library of age-appropriate family media ratings and reviews (30,000+ titles) covering all media types that reach 100+ million users. Additionally, Common Sense's innovative K-12 digital citizenship curriculum is currently being used in nearly 50% of U.S. schools.

In the UK we are also working with the Digital Learning Division at Education Wales to translate our curriculum for Welsh students, as well as several school groups in England.

[9] European Commission, Report: EU data protection rules empower citizens and are fit for the digital age, (24 June 2020).
[10] Joan Marsh, The Neutrality Debate We Need to Have, AT&T, (31 August 2020)

Common Sense continues to elevate the needs of children and families in public policy decision-making in the United States, United Kingdom, and European Union, advocating for stronger privacy protections for children and students, higher quality content, and closing connectivity gaps. Common Sense's research reports are helping fuel discussions of how media and technology are impacting kids and families today.

II. As designed, digital services and platforms often provide toxic and harmful online environments

Opaque algorithms are used by nearly every technology/social media platform to determine the content that individuals see online, including user-generated content and paid advertisements. Content-shaping algorithms determine the contours of a user's Facebook NewsFeed or what autoplay presents them on YouTube; they also dictate when, where, and what type of advertisements are shown. Both advertising and content personalization are only possible because of the vast troves of detailed information that the companies have accumulated about their users and their online behavior, often without specific, informed and unambiguous consent of the people being targeted.[11]

Targeted advertising encourages business practices that undermine user privacy and may have negative spillover effects. The underlying business model is premised on extensive data collection and sharing,[12] and it also encourages platforms to design algorithmic curation in a way that prioritizes sensational, controversial, and inappropriate content to maximize user engagement. This not only subjects children and young people to harmful and inappropriate material, but, as we have seen, it amplifies content that encourages the spread of conspiracy theories, undermines democracy, and can lead to the misinformation about vital public health matters such as the COVID-19 pandemic.[13] For example, Facebook permitted advertisers to profit from ads targeting people that the platform assumes are interested in "pseudoscience," which included more than 78 million people.[14] This sort of targeting facilitates the spread of misinformation and further indoctrinates users into harmful conspiracy theories -- and the platform directly profits from this process.[15]

This dynamic raises myriad issues that involve data protection, algorithmic governance, and free expression online. Many of these issues raise difficult questions about content moderation

[11] Nathalie Maréchal & Ellery Roberts Biddle, It's Not Just the Content, It's the Business Model at 13, Ranking Digital Rights (17 March 2020)

[12] *See, e.g.*, Johnny Ryan, A summary of the ICO report on RTB – and what happens next, Brave (26 June 2019), (Brave argues that online advertising leaks the habits of Internet users into the data broker ecosystem.)

[13] Nathalie Maréchal & Ellery Roberts Biddle, It's Not Just the Content, It's the Business Model, Ranking Digital Rights (17 March 2020)

[14] Aaron Sankin, Want to Find a Misinformed Public? Facebook's Already Done It, The Markup (23 April 2020)

[15] Chris Gilliard, Facebook Cannot Separate Itself from the Hate It Spreads, (6 July 2020)

by platforms. As Ranking Digital Rights[16] has explained, much of the current debate about content moderation is an acknowledgement that these systems are downstream efforts to clean up the mess caused upstream by platforms' own systems that are designed for automated amplification and audience targeting.[17]

In June 2020, Common Sense joined a coalition with leading U.S. civil rights organizations to pressure Facebook -- and encourage other online platforms -- to eliminate hateful content online. At a high level, we asked Facebook to adopt new policies that addressed (1) accountability, (2) decency on the platform, and (3) support for victims.[18] This included ten general demands that ranged from asks that Facebook remove all public and private groups focused on white supremacy, violent conspiracies, vaccine misinformation, and climate denialism and stop amplifying content associated with hate, misinformation, or conspiracies, to structural changes like establishing C-suite level civil rights infrastructure and submitting to regular audits of hate and misinformation on the platform. The demands included technical changes to how Facebook users could report harassing content and additional personnel to provide real time support to victims of severe harassment. We also asked Facebook to stop treating some public figures differently from average Facebook users -- a choice entirely dependent upon Facebook's leadership changing its mind.

Combatting these trends requires a mix of corporate accountability and outside auditing. Perhaps most important, it demands law and regulation. In the U.S., we have supported legislative proposals that include specific safeguards and requirements for technologies used by children.[19] Design mandates, limiting monetary incentives, and controls on children's advertising and content may all be appropriate topics for regulation. As discussed below, we recommend that the European Commission consider similar regulatory safeguards and other actions to mandate platform accountability.

III. Digital platforms and services are not taking sufficient actions to protect their users and especially kids

Currently, the wellbeing of children is either minimized or a secondary consideration for too many online platforms. A number of the Commission's consultation questions are responsive to

[16] Ranking Digital Rights (RDR) works to promote freedom of expression and privacy on the internet by creating global standards and incentives for companies to respect and protect users' rights.

[17] Nathalie Maréchal, Rebecca MacKinnon, & Jessica Dheere, Getting to the Source of Infodemics: It's the Business Model, A Report from Ranking Digital Rights, New America, (May 2020)

[18] Stop Hate For Profit Campaign, Product Recommendations, (June 2020)

[19] Common Sense has, for example, supported limits to children's exposure to unhealthy online content via social media and other algorithmically curated platforms, and limiting incentives to push inappropriate ads and disturbing and illegal content onto kids, as well as controls on algorithmic amplification and user interface design that subverts user choice and amplifies harmful content, S.3411 Kids Internet Design and Safety (KIDS) Act, (5 March 2020); S. 1084, Deceptive Experiences To Online Users Reduction Act (2019).

this reality, and our recommendations below identify which of the Commission's questions are relevant.

Platforms overly expose young people to inappropriate ads and unfair commercial practices. (Sec. I(1)(A) Qs. 20/21/22). Platforms offer near frictionless opportunities for endless consumption, be it of videos, posts, apps, or physical goods. Age verification efforts are inadequate (such as simply stating in a Terms of Use that a user must be 18) or nonexistent, and have, for example, allowed children to purchase weapons from Amazon.[20] Even products and content that is allegedly for children is rife with commercial manipulation. A third of children under age eight report sometimes or regularly watching unboxing videos -- essentially program-length advertisements -- on YouTube.[21] Popular children's apps use adult ad-networks, inappropriately commercially profiling and behaviorally targeting kids.[22] They offer children rewards for viewing advertisements.[23] Or they feature cartoon characters who berate preschool players for not spending money.[24] Platforms make these apps widely available and easily accessible. And both platforms and games disguise in-app purchases so children don't realize they are spending real money.[25]

Platforms take insufficient steps to protect children from bullying, sexism, racism, and hate online. (Sec. I(1)(C), Qs. 1/3). A 2018 Common Sense survey found that 64% of teen social media users say they come across hateful content on social media; one in five report they "often" see inappropriate content.[26] And about as many parents whose younger kids watch YouTube say their child has encountered content they felt was unsuitable for children.[27] A recent 2020 survey from the ADL found that nearly half of respondents report harassment online; more problematic, 28% of respondents experienced severe online harassment, which includes sexual harassment, stalking, physical threats, swatting, doxing and sustained harassment.[28] And a majority of teens (59 percent) have experienced "some sort of cyberbullying."[29]

In addition to harassment, online platforms have become a conduit for facilitating and sharing images of child abuse. The pandemic underscores how the spread of exploitative material about

[20] The Parent's Accountability and Child Protection Act, A California 14 year old purchased a BB gun, throwing knives, and a hunting knife on Amazon without his parent's knowledge. (September 2018)
[21] Common Sense Media, Common Sense Census: Media Use by Kids Age Zero to Eight, (19 October 2017)
[22] Fangwei Zhao; Serge Egelman; Jenny S. Radesky, et al., Data Collection, Sharing Practices of Apps Played by Young Children JAMA Pediatrics (2020) (on file with author)
[23] S.3411 Kids Internet Design and Safety (KIDS) Act, (5 March 2020)
[24] Marisa Meyer, Victoria Adkins , Nalingna Yuan , Heidi M Weeks , Yung-Ju Chang , Jenny Radesky, Advertising in Young Children's Apps: A Content Analysis (January 2019)
[25] Letter from Common Sense Media et al. to Federal Trade Commission, (21 February 2019).
[26] Common Sense Media, Social Media, Social Life: Teens Reveal Their Experiences (10 September 2018)
[27] Common Sense Census: Media Use by Kids Age Zero to Eight, (19 October 2017)
[28] Anti-Defamation League, Online Hate and Harassment Report: The American Experience 2020, (2020)
[29] Common Sense Media, Teaching Digital Citizens in Today's World: Research and Insights Behind the Common Sense K–12 Digital Citizenship Curriculum, (2019)

children, often termed child sexual abuse materials or CSAM, is a rampant and growing problem. Research surveys suggest the number of CSAM crimes is increasing,[30] and a New York Times report discovered that, in just one year (from 2017 to 2018), tech companies reported *double* the number of online photos and videos of children being sexually abused.[31]

IV. The Commission should ensure platforms are protecting young users from illegal content, as well as legal-but-harmful content, by appropriately complying with the law

Existing regulation, as well as many digital services' current community guidelines and moderation practices, create a potential foundation for action. The Commission can investigate, legislate, and ensure appropriate incentives for platforms to protect their online communities and children.

First, the Commission must mandate additional and appropriate resources for platforms' moderation practices. (Sec. I(2), Qs. 1/2/3 and 9/10/11). Enforcing platform rules and moderating content raises many challenges and costs. While most platforms have dedicated trust and safety teams, the scope and volume of personnel and monetary resources dedicated to trust and safety policies varies.[32] For example, Facebook employs approximately 15,000 content moderators, the vast majority of whom are employed as contract workers to third-party vendors.[33] Facebook often highlights the successes of its teams, but errors are common. Recently, Facebook acknowledged "an operational mistake" when moderation contractors permitted self-proclaimed white militia groups to organize on the platform, including soliciting armed attendees that resulted in the murder of protestors in Kenosha, Wisconsin.[34] Facebook's approach to content moderation has resulted in negative mental health consequences for its contractors, as well.[35]

One recent study from the NYU Stern Center for Business and Human Rights has called for ending this outsourcing and bringing content moderation in-house. Additional recommendations include (1) *doubling* the number of moderators to improve review quality, (2) hiring someone to oversee content and fact-checking who reports directly to C-suite, (3) and provide moderators with access to mental healthcare and sponsor research into the health risks of content

[30] Internet Watch Foundation,Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse, (May 2018)

[31] Michael H. Keller and Gabriel J.X. Dance, The Internet Is Overrun With Images of Child Sexual Abuse. What Went Wrong?, The New York Times, (29 September 2019)

[32] *See* Carrie Melissa Jones, The 2018 State of Platform Trust & Safety Standards, (13 September 2018); CSO Online, Trust and safety 101, (7 July 2013)

[33] Charlotte Jee, Facebook needs 30,000 of its own content moderators, says a new report, MIT Technology Review, (8 June 2020)

[34] Russell Brandom Mark Zuckerberg says Kenosha Guard rulings were 'an operational mistake' (28 August 2020)

[35] Zack Whittaker, Facebook to pay $52 million to content moderators suffering from PTSD, (12 May 2020)

moderation.[36] In addition to requiring in-house, trained, supported moderation staff, the Commission should ensure that platforms appropriately resource efforts to identify and limit amplification of problematic content, for example requiring commensurate spending on moderation as advertising or user engagement, or a certain percentage of revenue. The Commission could audit platforms on an ongoing basis to ensure moderation is adequately resourced.

Second, the Commission must ensure that civil society has a voice in industry's efforts to address illegal content. (Sec. I(2), Qs. 1/2/3 and 9/10/11). Regulators, academics, and outside experts are often essentially working in the dark when they attempt to understand or combat hatred and misinformation on platforms. The Commission should ensure that independent outside experts have the tools, both administratively and technically, to look "under the hood" (and can do so in a way that does not expose trade secrets); society and platforms themselves will benefit from this collaboration. In the United States, the Stop Hate for Profit coalition of civil rights and technology advocacy groups have sought to increase their direct collaboration with the teams at major platforms that are responsible for these decisions. The Commission may wish to explore how it can encourage more formalization of these efforts at relationship-building.

Third, the Commission should put forward, to the extent possible, a common set of standards or rules for identifying illegal and inappropriate content and advertisements. (Sec. I(2), Qs. 1/2/3 and 9/10/11). With respect to illegal and harmful content, for example, the UK government's White Paper on Online Harms lays out a non-exhaustive but detailed list of online harms -- illegal and unacceptable alike -- with a special focus on children's exposure to such harms.[37] The Terrorist Content Directive lays out an understanding of content that solicits or supports terrorist offenses.[38] With respect to advertisements, some of the most harmful ads are political misinformation. Legislation introduced in California would define such political misinformation and prohibit the distribution, with actual malice, of false material statements of fact with the intent to deceive a voter into voting for or against a candidate or measure.[39] And author amendments contemplate extending this prohibition to the amplification of such political disinformation as well.[40]

V. The Commission should address algorithmically-generated content and platform manipulation, including considering how existing data protection rules under the GDPR can aid these efforts

[36] Paul Barrett, Who Moderates the Social Media Giants? A Call to End Outsourcing, NYU Stern Center for Business and Human Rights (June 2020)
[37] Department for Digital, Culture, Media & Sport, Online Harms White Paper, (12 February 2020)
[38] European Parliamentary Research Service, Terrorist Content Online: Tackling online terrorist propaganda (Mar. 2020)
[39] AB 2885 False campaign speech and online platform disclosures, (21 February 2020)
[40] Amendments to AB 2885, False campaign speech and online platform disclosures, (17 June 2020)

First, the Commission should consider how existing privacy and data protection laws in the EU such as the General Data Protection Regulation could be used to address online harms ranging from user manipulation, online content amplification of harmful content, and opaque algorithmic profiling. (Sec. I(2), Qs. 10/11; Sec. IV, Qs. 18/19). Specific rules requiring purpose specification for data processing under Article 5, limiting how special categories of data may be processed under Article 9, controls on automated profiling under Article 22, and user rights to access and port data under Article 15 and Article 20 could prove useful tools to limit online harms.[41] Unfortunately, enforcement of the GDPR remains inconsistent,[42] and regulators have yet to take firm positions on how common data-driven practices and advertising activities by major platforms are implicated by the Regulation.[43] A simple focus on enforcing the already existing requirements of the GDPR -- that companies use data only for specified purposes, that companies limit processing of special category data, that companies respect controls on automated profiling and that users can easily move from service to service -- could limit large platforms' ability to target and manipulate users. It would also improve users' relative power compared to platforms, providing more opportunity for them to leave a platform (though "social network" effects may still provide a limitation) particularly if the user disagrees with that platform's stance on misinformation, violence, or other harmful content.

Article 22's right to avoid decisions based solely on automated processing should be used to ensure individuals can avoid online experiences that are largely the basis of automated processing and profiling, which can lead to significant impacts such as radicalizing children.[44] The Commission should ensure that individuals are protected by requiring algorithmically curated feeds be used only with consent and be turned off by default.

Second, the Commission should ensure technology design does not manipulate users, particularly children. (Sec. I(2), Qs. 10/11). Online apps, platforms, and services are also rife with manipulative user interfaces and design. So-called "dark patterns" are design tactics used to nudge, manipulate, or push data subjects towards activities that benefit the company. Design elements including hard-to-find buttons and confusing menus encourage children to spend more money, share more data, engage more with the platform, or discourage them from taking control of their experience.[45]

In the United States, the KIDS Act would prohibit manipulative design features that keep kids glued to the screen or that trick children into sharing data or making online purchases, such as

[41] *See* Council on Foreign Relations, Could Europe's New Data Protection Regulation Curb Online Disinformation?, (20 February 2018)

[42] Vincent Manancourt & Mark Scott, Two years into new EU privacy regime, questions hang over enforcement, PoliticoEU (25 May 2020); Nicholas Vinocur, 'We have a huge problem': European regulator despairs over lack of enforcement, Politico EU (27 December 2019)

[43] *See* UK ICO, Our work on adtech

[44] *See* Kevin Roose, The Making of a YouTube Radical, The New York Times (8 June 2019).

[45] Dark Patterns; see also Katie McInnis, How might we evaluate dark patterns?, Digital Lab at Consumer Reports (15 October 2019)

rewards for watching or purchasing or autoplay.[46] Also, the bipartisan Deceptive Experiences To Online Users Reduction (DETOUR) Act is a legislative response to address dark patterns on large online platforms.[47] In addition to provisions that curb design interfaces that create compulsive usage among children under the age of 13 years old, the DETOUR Act encourages the creation of professional standards bodies to provide guidance on design practices that undermine user autonomy, places limits on online behavioral and psychological experiments,[48] and promotes the development of independent review boards.[49] Other ideas include the promotion of "light patterns" which are interface designs that empower users, offering transparency, information, and options that are accessible and intuitive.[50] Transatlantic civil society organizations have called for additional transparency in the form of "nutrition labels for digital content" that could be considered for trusted information, misinformation, algorithmic curation, and accounts (such as distinguishing among verified and other bot accounts). Adding friction that can slow down content sharing can also empower users.

VI. The Commission must make digital platforms and services accountable for the actions they take and promises they make

First, the Commission should reject the tech company mantra of "move fast and break things." There are methods to build safeguards and friction that protect users, which the Commission should encourage or mandate. (Sec. I(2), Qs. 10/11). While platforms have been incentivized to encourage user engagement, leading to a toxic race to the bottom, the Commission should devote resources and attention to ensuring children and families have access to quality, trustworthy content. Policymakers should take action to limit algorithmic and other targeting of advertisements that can have detrimental societal effects. For example, political misinformation ads can pose a clear threat to democracy. In the United States, lawmakers have introduced bills that would ban microtargeting of political ads -- with exceptions only for location[51] or, in another bill, for location, age, and gender.[52] These are practices some platforms already endorse.[53]

[46] S.3411 Kids Internet Design and Safety (KIDS) Act, (5 March 2020)

[47] Deceptive Experiences To Online Users Reduction (DETOUR) Act (2019)

[48] Research Review at Facebook, Evan Selinger and Woodrow Hartzog, Facebook's emotional contagion study and the ethical problem of co-opted identity in mediated environments where users lack control (13 May 2015)

[49] Ryan Calo, Consumer Subject Review Boards (2013).

[50] Karen Kornbluh, Ellen Goodman & Eli Weiner, Safeguarding Democracy Against Disinformation, German Marshall Fund (2020).

[51] H.R.7014, Banning Microtargeted Political Ads Act (25 May 2020)

[52] H.R.7012, Protecting Democracy from Disinformation Act (26 May 2020).

[53] Google purports to limit targeting and Twitter has put in place a ban on political ads. *See* Kate Cox, Proposed bill would ban microtargeting of political advertisements, Ars Technica (26 May 2020). Facebook meanwhile allows users to turn them off--less effective given it requires user action. *See* Alison DeNisco Rayome, Don't want political ads in your Facebook or Instagram feed? Turn them off now, CNET (26 August 2020).

It is no longer appropriate -- if it ever was -- for platforms to “move fast and break things.”[54] Platforms have suggested that content moderation is impossible to do well at scale.[55] If this is so, platforms should have a duty and legal responsibility to consider limiting the speed and amount of content posted and shared to a level that can be responsibly overseen with appropriate human and/or automated review. There are many strategies the Commission could look at as it encourages digital services to build friction into how content is posted on their platforms and encourage more responsible information sharing. For example, Twitter has recently announced it will explore using warning interstitials when users attempt to share weblinks they have not actually read.[56] YouTube uses machine learning to identify potentially inappropriate comments on videos and can hold these comments for review by video creators, putting hateful and harmful comments into a holding pen by default.[57] The reality is that the environment online for user-generated content is closer to spam-filled email inboxes than the image of informed discussion and constructive self-expression that platforms portray. Platforms could take lessons learned from email mailbox hosting services, service providers, and marketers to control, report, and police spam. A combination of filtering and reporting have been useful mechanisms to establish feedback loops that control the flood of spam.[58]

Second, the Commission should require additional platform transparency, especially with respect to advertisements, and hold platforms responsible for the promises and commitments they make. (Sec. I(2), Qs. 18/19/20; Sec. IV, Qs. 14/15). Platforms operate in near opacity. Algorithms that push content, offers, and opportunities and that, in essence, define children’s and other individuals’ online realities must be made visible to understanding and critique -- at the very least by expert third parties and competent regulators.

This call for transparency is coming not just from consumer advocates but from global corporations as well.[59] Some U.S. legislative proposals to improve transparency of advertisement targeting include the bipartisan HONEST ADS Act which would, among other things, require large platforms to maintain a public file of electioneering communications including “a digital copy of the advertisement, a description of the audience the advertisement targets, the number of views generated, the dates and times of publication, the rates charged, and the contact information of the purchaser.”[60] A proposal in California would require platforms to make available an application programming interface or other technical capabilities to qualified, third party researchers to enable an independent analysis of any bias or unlawful

[54] Even FB ostensibly moved on from this years ago. *See* Samantha Murphy, Facebook Changes its ‘Move Fast and Break Things’ Motto, Mashable (30 April 2014).
[55] Mike Masnick, Masnick’s Impossibility Theorem: Content Moderation At Scale is Impossible to Do Well, TechDirect (20 November 2019)
[56] Twitter Support (20 June 2020), https://twitter.com/twittersupport/status/1270783537667551233
[57] Potentially inappropriate comments now automatically held for creators to review, YouTube Help (8 June 2020), https://support.google.com/youtube/thread/8830320?hl=en
[58] Messaging Anti-Abuse Working Group (MAAWG), Complaint Feedback Loop Best Current Practices, (April 2010)
[59] Joan Marsh, The Neutrality Debate We Need to Have, AT&T (31 August 2020)
[60] S.1356, The Honest Ads Act (7 May 2019)

discrimination in algorithms that support targeted advertising.[61] Understanding ad placement and targeting will not only assist in identifying ads placed near illegal and inappropriate content, it will also assist in better understanding platforms data collection and targeting capability in general.

Platforms often make overbroad promises about how violent, hate-filled content is not allowed. And then they do little to back up those promises.[62] Promises platforms make to users in their Terms of Service or in Community Standards should be enforceable against the platform, and no “purchase” or login should be required. In the U.S., failing to keep up with any promises in terms or standards can be enforceable by Attorney Generals under UDAP laws (Unfair and Deceptive Acts and Practices), but regulators struggle in proving cases in court given the lack of specificity in promises. The Commission should not allow platforms to use vague language to avoid obligations, and should take platforms to task when they have a pattern and practice of falling short in meeting promises to users.

[61] Amendments to AB 2885, False campaign speech and online platform disclosures, (17 June 2020)
[62] Common Sense Media, 2020 Social Media Voter Scorecard, (forthcoming 9 September 2020)