OPEN
TECHNOLOGY
INSTITUTE

Submission to the Consultation on the Roadmap for Digital Services Act: deepening the Internal Market and clarifying responsibilities for digital services

September 2020

I. Safety and Responsibilities

What good practices can you point to in handling the availability of illegal goods online since the start of the COVID-19 outbreak?

As discussed in OTI's report *How Internet Platforms Are Combating Disinformation and Misinformation in the Age of COVID-19*, many internet platforms have developed policies to reduce the spread of and remove misleading and inaccurate information related to the virus, but platforms need to do more to provide transparency and accountability around how these initiatives are being implemented and how they are impacting users and their online expression. One positive example from Amazon is that during the pandemic, they have expanded their policies regarding false claims about diseases, to enable them to take down harmful listings related to cures or treatments for the virus and handle the availability of illegal goods during the pandemic. The company's Prohibited Product Claims for Diseases policy states, "Amazon prohibits the sale of products that claim to cure, mitigate, treat, or prevent diseases in humans or animals without FDA approval." The list of examples of diseases that products cannot claim to cure includes "Coronavirus and/or COVID-19." In February, the company notified third-party merchants that it was taking down listings for items claiming to be a treatment, cure, or remedy for the coronavirus. After that notice, the company confirmed that it blocked or removed more than 1 million products for suspect or misleading claims. While Amazon has told sellers it would remove their listings for making unapproved medical marketing claims, the company has given sellers the opportunity to keep any valid product up without the prohibited medical claims.

However, Amazon and other companies can and should improve their efforts to connect users to authoritative information, moderate or reduce the spread of misleading content, alter and enforce advertising policies, and provide transparency around their efforts during the pandemic. Platforms should provide transparency around COVID-19 moderation and enforcement efforts by publishing a report and periodic public updates with data on the number of listings that the company has removed and the number of sellers the company has banned for violating its COVID-19 specific policies as well as its preexisting commerce policies.

When content is recommended to you - such as products to purchase on a platform, or videos to watch, articles to read, users to follow - are you able to obtain enough information on why such content has been recommended to you? Please explain.

As OTI outlined in our reports *Special Delivery: How Internet Platforms Use Artificial Intelligence to Target and Deliver Ads* and *Why Am I Seeing This: How Video and E-Commerce Platforms Use Recommendation Systems to Shape User Experiences*, internet platforms such as Facebook, Google, and Twitter, offer users a limited set of controls. To some extent, these tools allow users to understand why certain content has been recommended or delivered to them, but platforms do not provide enough such information or controls. For example, a user can visit the Google Accounts page to view information on how a user's ads were customized, including what some of the data sources that have contributed to these recommendations are. Users have the option to remove certain interest groups (e.g. interested in combat sports, gardening, etc.) from a list of active factors considered by Google's systems when making these recommendations. However, data from Google has suggested that although there were 2.5 billion visits to the Google Accounts page in 2018, only 20 million people per month visited the ad settings page. Generally, these internet platforms provide users with a limited amount of information around why they are seeing certain recommendations. Although some of this information is readily available (e.g. users on Facebook can click on a post in the News Feed and access a button that directs them to this information), other platforms like Google tend to nest this information under several drop down menus and pages, making it inaccessible to the regular user. These tools are often worded in complex language as well, making it difficult for users who do access these pages to actually understand how the different controls will influence and change their online experience. Further, these controls often do not allow users to opt-out of receiving algorithmically curated recommendations altogether. Research has indicated that promoting awareness of the use of algorithmic tools and enabling users to control their own experiences on a platform are fundamental steps in building trust with users. This lack of transparency and accountability is therefore concerning.

Are you aware of evidence on the scale and impact of erroneous removals of content, goods, services, or banning of accounts online? Are there particular experiences you could share?

Currently, internet platforms do not provide significant transparency around the scale and impact of erroneous removals of content, goods, etc. that occur on their platforms. In its quarterly Community Standards Enforcement Report (CSER), Facebook publishes data on the scope and scale of their content moderation efforts for 10 categories of content on Facebook and eight categories of content on Instagram. The report includes two relevant metrics: 1) how much of the content Facebook actioned did users appeal, and 2) how much of the actioned content was later restored. These metrics provide some insight into how often users appeal Facebook's content moderation decisions, and into how often Facebook restored content as a result of these appeals. The CSER also includes data on the amount of content that was restored as a result of proactive determinations by Facebook that an error was made. However, Facebook does not publish more granular and contextual data on factors such as how effective their algorithm-based content moderation systems are at reviewing and removing content, and how effective their human reviewers are. These missing data points would be important for understanding the larger landscape of erroneous removals at the company. Similarly, YouTube publishes a Community Guidelines Enforcement Report, which includes data points for the total number of videos removed, the total number of videos appealed, and the total number of videos reinstated. The company, however, does not break these numbers down by category of content. They also do not offer this data for channels and comments, which are content formats that they disclose data related to in other metrics in the report. Similar to Facebook, Google also does not publish any information related to the overall accuracy of their automated and human review process. Other platforms also report appeal data, but the metrics and data provided are similarly not granular enough to understand the true scale and impact of erroneous removals on these services.

Where automated tools are used to detect illegal content, goods or services, what opportunities and risks does their use present as regards different types of illegal activities and the particularities of the different types of tools?

Digital hash technology is commonly used to moderate illegal content. Generally the use of digital hash tools are more accepted when the content they are designed for is globally considered illegal and there are clear definitions delineating what content is included in these hash databases. For example, PhotoDNA, a digital hash-based tool that is used to detect child sexual abuse material (CSAM), operates by generating digital hashes from a database of thousands of existing illegal CSAM images, which are then used for image detection and removal. In response to user concerns around copyright-infringement, YouTube adapted PhotoDNA to create ContentID, enabling YouTube users to create digital hashes for their video content to help protect against copyright violations. Once these hashes have been created, all content subsequently uploaded to YouTube is screened against its database of audio and video files to identify potential copyright violations.

By contrast, when these tools have been used to address content that is not globally regarded as illegal, there are greater concerns. For example, internet platforms and the Global Internet Forum to Counter Terrorism (GIFCT) have adapted PhotoDNA to address extremist content online. This is concerning as the legality and definition of extremist content varies across jurisdictions, and as a result, there is no clear guidance on what content should be included in the hash database. In addition, most internet platforms' moderation efforts emphasize certain extremist groups (e.g. the Islamic State, al-Qaeda). As a result, their automated tools are less reliable when addressing the larger corpus of extremist groups that use their services. Internet platforms and the GIFCT do not provide adequate transparency around what kind of content is included in their extremist-content hash databases, how much content/how many accounts have been removed as a result of these hashes, and how many erroneous removals have occurred. This demonstrates a significant lack of accountability around the use of this automated technology. In addition, the moderation of extremist content often requires a nuanced understanding of varied regions and cultures, and an appreciation for the context in which a piece of content is posted. Automated tools cannot provide this contextual understanding and as a result it is important to have a human reviewer in the loop. This has resulted in the erroneous removal of content posted by journalists and human rights organizations seeking to raise awareness about terrorist atrocities, raising concerns related to overbroad takedowns and infringements of user expression. Without transparency and accountability around these processes and the volume of errors, users are often unable to receive remedy, and researchers are unable to understand the true scope of this problem.

See OTI's report: *Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*

Please rate the necessity of the following measures for addressing the spread of disinformation online. Please rate from 1 (not at all necessary) to 5 (essential) each option below.

Transparently inform consumers about political advertising and sponsored content, in particular during election periods	5
Provide users with tools to flag disinformation online and establishing transparent procedures for dealing with user complaints	5
Tackle the use of fake-accounts, fake engagements, bots and inauthentic users behaviour aimed at amplifying false or misleading narratives	5
Transparency tools and secure access to platform data for trusted researchers in order to monitor inappropriate behaviour and better understand the impact of disinformation and the policies designed to counter it	4
Transparency tools and secure access to platform data for authorities in order to monitor inappropriate behaviour and better understand the impact of disinformation and the policies designed to counter it	4
Adapted risk assessments and mitigation strategies undertaken by online platforms	4
Ensure effective access and visibility of a variety of authentic and professional journalistic sources	5
Auditing systems for platform actions and risk assessments	4

As we have discussed in our report *How Internet Platforms Are Combating Disinformation and Misinformation in the Age of COVID-19* and in our forthcoming report on election-related disinformation, internet platform efforts to combat misleading information must include four key components: 1) promoting legitimate and authoritative information and empowering informed user decision-making, 2) moderating and curating misleading information, 3) tackling misleading advertising, and 4) providing meaningful transparency and accountability.

In particular, internet platforms should partner with relevant entities and individuals to ensure that users have access to legitimate and authoritative information related to topics including COVID-19 and voting, including reliable news sources. In addition, internet platforms should provide adequate transparency around the identity of individuals or organizations who are posting and paying for content, both unpaid and paid, on their services. This information will likely affect users' assessments of how credible a source is.

Internet platforms should also create policies which clearly describe their prohibitions on misleading content and platform manipulation (e.g. bots, inauthentic behavior, etc.) and outline how the platform addresses this content on their services. Platforms should ensure that these policies are implemented consistently. In addition, companies should ensure that users have the ability to flag content that violates these policies, and should clearly outline how this content is reviewed, and how users will be notified of moderation decisions.

We also strongly encourage companies to establish programs which allow vetted researchers to examine internal company moderation and algorithmic data and systems in order to understand how platform policies, technologies, and practices related to disinformation can be improved. We also encourage platforms to proactively submit to independent audits of their algorithmic systems, which are often responsible for the promotion and distribution of misleading content online.

What would be effective measures service providers should take, in your view, for protecting the freedom of expression of their users? Please rate from 1 (not at all necessary) to 5 (essential).

High standards of transparency on their terms of service and removal decisions	5
Diligence in assessing the content notified to them for removal or blocking	5
Maintaining an effective complaint and redress mechanism	5
Diligence in informing users whose content/goods/services was removed or blocked or whose accounts are threatened to be suspended	5
High accuracy and diligent control mechanisms, including human oversight, when automated tools are deployed for detecting, removing or demoting content or suspending users' accounts	5
Enabling third party insight – e.g. by academics – of main content moderation systems	5

OTI is one of the original authors of the Santa Clara Principles on Transparency and Accountability in Content Moderation, which outlines minimum standards that internet platforms must meet in order to provide meaningful transparency and accountability around their content moderation practices. The Santa Clara Principles emphasize three key factors: 1) platforms should disclose granular numbers related to their content moderation efforts in regular transparency reports, 2) platforms should provide adequate notice to impacted users, and 3) platforms should offer timely and robust appeals processes to impacted users. OTI has been a longstanding advocate of the Principles and has been working with both small and large internet platforms to encourage adoption.

In addition, as previously outlined, internet platforms often tout automated tools as silver bullet solutions to their content moderation and curation problems. However, automated tools are unable to assess context and as a result companies should only use them to augment human review efforts. In addition, there is little transparency around how these tools are created, deployed, and refined, and how accurate they are. As a result, we strongly encourage that companies always maintain a human in the loop, especially when it comes to the moderation of categories of content that require context or subjective decision-making, as well as when it comes to the review and moderation of content that could have significant offline impacts, such as advertising related to politics, housing, employment, or credit.

Further, as previously outlined, internet platforms should establish programs which allow vetted researchers to access and audit their algorithmic systems, including their content moderation systems, in order to understand flaws, especially ones that could result in biased, discriminatory, or harmful outcomes, and make suggestions for improvement.

In your view, what information should online platforms make available in relation to their policy and measures taken with regard to content and goods offered by their users? Please elaborate, with regard to the identification of illegal content and goods, removal, blocking or demotion of content or goods offered, complaints mechanisms and reinstatement, the format and frequency of such information, and who can access the information.

As we have outlined in our Transparency Reporting Toolkit on Content Takedowns, the Santa Clara Principles, as well as our subsequent report series on how internet platforms use algorithmic decision-making for a range of content curation purposes, internet platforms should publish regular transparency reports that outline the scope and scale of their content moderation and curation efforts. Transparency reports should be published in an openly licensed, machine-readable format, with durable links. At a minimum, companies should disclose the number of accounts and pieces of content that were flagged and removed and break down this data by:

- How much of the content or how many of the accounts were flagged by automated tools, user flags, through Internet Referral Units, etc.
- What policies content or accounts that were removed or curated violated (e.g. hate speech, terror propaganda, disinformation)
- What product content or accounts were moderated on (e.g. where possible Facebook should report on Facebook, Instagram, WhatsApp, etc.)
- How many appeals were submitted for content moderation or curation practices
- How much content or how many accounts were reinstated as a result of appeals
- How much content was proactively restored by a platform in recognition of a moderation error

As curation processes such as labeling and downranking are increasingly used to moderate content, companies should also expand their transparency reports to include data on the scope and scale of these mechanisms across different categories of content. Further, as previously outlined, internet platforms should provide greater transparency around how accurate their algorithmic moderation and curation tools are.

In addition, platforms should ensure that they publish a detailed set of their content policies, which outline what content is permitted and what content is prohibited on their services. This information should also explain how the platform moderates violating content and addresses accounts that are repeat offenders. Further, as previously outlined, internet platforms should ensure that both users who have had their content or accounts flagged or removed, and users who have flagged content or accounts for removal, receive adequate notice informing them of moderation decisions. These users should also have the ability to appeal these decisions.

We also offer further recommendations on the moderation of disinformation and subsequent transparency reporting expectations in our COVID-19 disinformation report, *How Internet Platforms Are Combating Disinformation and Misinformation in the Age of COVID-19*.

Further, we have published a Transparency Report Tracking Tool, which outlines how different internet platforms are currently reporting on their content moderation practices, including a comparison of the metrics they report on.

In your view, what measures are necessary with regard to algorithmic recommender systems used by online platforms?

Internet platforms use algorithmic recommendation systems to identify and make personalized recommendations on content, products, and services that may be of interest to their users. These systems are incredibly pervasive, and are able to influence user interests, opinions, and behaviors,as well as their social group formation. Although many internet platforms assert that these systems enhance users' experiences through personalized and relevant recommendations, these systems also enable platforms to retain user attention on their services. This translates into significant financial benefits for the companies, as they can then target these users with advertisements and recommend further content to consume or items to purchase. In addition, researchers have found that recommender systems can also create a number of concerning outcomes. In particular, these systems can reinforce societal biases and augment harmful perspectives, such as those of extremists, conspiracy theorists, and mis/disinformation campaigns. Despite this, internet platforms that deploy these recommendation systems do not currently provide meaningful transparency and accountability around how these systems are created, how they operate, and how they make decisions.

In our report *Why Am I Seeing This: How Video and E-Commerce Platforms Use Recommendation Systems to Shape User Experiences*, OTI urges internet platforms to consider the following recommendations in order to promote greater fairness, accountability, and transparency around their algorithmic recommendation systems.

1. Disclose to users the situations in which the platform uses an algorithmically-curated recommendation system and provide comprehensive and meaningful explanations to users around how their recommendation systems work.

2. Explain to users why a recommendation was made to them. This explanation should at a minimum include information on the different signals and user characteristics the recommendation system considered to make the recommendation. It should also include an easy link to relevant user controls (per recommendation seven below) that could let the user change their recommendation preferences.

3. Disclose granular data around how the company trains its algorithmic recommendation systems. At a minimum, this should include information on the categories of users that a company's training data sets are trained on (e.g. which demographic groups).

4. Enable independent researchers to conduct audits to review and verify relevant internal models and data.

5. Hire independent auditors to conduct regular periodic audits of recommendation algorithms in order to identify potentially harmful outcomes and take steps to address findings of audits, including mitigating discrimination and bias.

6. Share granular data related to how the company tests its recommendation systems and how it determines how effective the company's systems are. At a minimum, this should include information on how well these systems predict the preferences of different demographic groups. In addition, this data should be continuously updated to indicate how various algorithmic changes have impacted the company's metrics and conclusions related to the overall effectiveness of the company's recommendation system.

7. Improve user controls so that users can easily manage whether and how their data is collected and inferred, how this data is used, and how it influences the recommendations that they see. These user controls should be easy to

access and understand. They should be available to all logged in users of a service. In addition, these controls should be accompanied with an explanation of how using these controls will impact a user's overall platform experience. Among other things, user controls should enable users to:

- Select and change the factors and personal data points that a recommendation system may consider when generating recommendations from them
- Opt-out entirely from receiving algorithmically curated recommendations or from using the autoplay feature
- Exclude certain videos, titles, channels, sellers, or items from factoring into their recommendations

8. Share the platform's Terms of Service Community Guidelines related to topics such as content and purchases, and how they are enforced.

9. Publish a transparency report outlining the scope and scale of Terms of Service enforcement actions in all of the regions in which it operates.

10. Explain how the company uses human evaluators to review and train its algorithmic and machine-learning models.

II. Liability regime

The E-commerce Directive also prohibits Member States from imposing on intermediary service providers general monitoring obligations or obligations to seek facts or circumstances of illegal activities conducted on their service by their users. In your view, is this approach, balancing risks to different rights and policy objectives, still appropriate today? Is there further clarity needed as to the parameters for 'general monitoring obligations'? Please explain.

Yes, the Directive should continue to prohibit Member States from imposing an obligation for companies to conduct general monitoring of their users' content. Companies should be permitted to offer services such as fully end-to-end encrypted messaging services that protect users' privacy and security. Encryption provides critical digital security to people around the globe ranging from journalists, to government employees, to domestic violence survivors, to ordinary consumers. It protects people against countless threats, including from computer criminals trying to defraud us, corporate spies trying to obtain our companies' most valuable trade secrets, or repressive governments trying to stifle dissent. Any general monitoring obligation would preclude companies from offering strong encryption, and therefore the Commission should not impose such an obligation.

In addition, the Commission should not impose a monitoring obligation even for unencrypted content that users post openly on platforms. Such a content monitoring obligation could force companies to screen user content before it is published, raising significant threats to freedom of expression, especially because platforms already play a significant role as gatekeepers of online speech. Further, in order to conduct such time-sensitive pre-publication reviews, companies would need to rely exclusively on automated screening tools, and such tools cannot assess context or provide the nuanced understanding of varied regions, cultures, and languages that is necessary for accurate content moderation. As a result, the DSA should reiterate that platforms are not required to use general monitoring approaches and tools, and that they should not institute broad policies that rely solely on automated filtering technology to address content removals.

III. Gatekeeper Platforms

To what extent do you agree with the following statements?

Consumers have sufficient choices and alternatives to the offerings from online platforms.	Fully disagree
It is easy for consumers to switch between services provided by online platform companies and use the same or similar services provider by other online platform companies ("multi-home").	Somewhat disagree
It is easy for individuals to port their data in a useful manner to alternative service providers outside of an online platform.	Somewhat disagree
There is sufficient level of interoperability between services of different online platform companies.	Fully disagree
There is an asymmetry of information between the knowledge of online platforms about consumers, which enables them to target them with commercial offers, and the knowledge of consumers about market conditions.	Fully agree
It is easy for innovative SME online platforms to expand or enter the market.	Fully disagree
Traditional businesses are increasingly dependent on a limited number of very large online platforms.	Fully agree
There are imbalances in the bargaining power between these online platforms and their business users.	Somewhat agree
Businesses and consumers interacting with these online platforms are often asked to accept unfavourable conditions and clauses in the terms of use/contract with the online platforms.	Fully agree
Large online platforms often leverage their assets from their primary activities (customer base, data, technological solutions, skills, financial capital) to expand into other activities.	Fully agree
When large online platform companies expand into such new activities, this often poses a risk of reducing innovation and deterring competition from smaller innovative market operators.	Fully agree

Which characteristics are relevant in determining the gatekeeper role of large online platform companies? Please rate each criterion identified below from 1 (not relevant) to 5 (very relevant).

Large user base	4
They capture a large share of total revenue of the market you are active/of a sector	5
Impact on a certain sector	5
They build on and exploit strong network effects	5
They leverage their assets for entering new areas of activity	5
They raise barriers to entry for competitors	4
They accumulate valuable and diverse data and information	5
There are very few, if any, alternative services available on the market	4
Lock-in of users/consumers	4

Do you believe that the integration of any or all of the following activities within a single company can strengthen the gatekeeper role of large online platform companies ('conglomerate effect')? Please select the activities you consider to strengthen the gatekeeper role:

- online intermediation services (i.e. consumer-facing online platforms such as e-commerce marketplaces, social media, mobile app stores, etc., as per Regulation (EU) 2019/1150 - see glossary)
- Search engines
- Operating systems for smart devices
- Consumer reviews on large online platforms
- Network and/or data infrastructure/cloud services
- Digital identity services
- Payment services (or other financial services)
- Physical logistics such as product fulfilment services
- Data management platforms
- Online advertising intermediation services
- Other. Please specify in the text box below.

Are there specific issues and unfair practices you perceive on large online platform companies?

Vertical integration and data consolidation enable large online platforms to capitalize on economies of scale and tap into network effects. Vertically integrated firms offer products that feed into one another along a single production vertical. In the absence of vertical integration, different companies usually produce a different product or service along a supply chain. When firms vertically integrate, however, they usually seek to tap into efficiencies gained from the supply chain integration, and give preference to their own supply chain components when designing products and services to the exclusion of other players in the ecosystem—in their API design, for instance. The more vertically integrated a platform is, the higher the risk that it may not offer APIs with sufficient data and functionality for other companies, particularly downstream businesses, to build products that are compatible with theirs.

Vertically integrated platforms have incentives to build their API design solely to their own needs, tailored to their own specific apps, features, and competitive strategy. Twitter, for instance, vertically integrated by purchasing apps like TweetDeck (a social media dashboard application for managing Twitter accounts) in 2011, Tweetie (then a leading iPhone Twitter client) in 2010, and Summize (a search engine built specifically for indexing Twitter posts) in 2008, and as a result was in a position to discourage developers from using Twitter's APIs to make apps that directly competed with their platform. Twitter rejected apps that relied on tweet feed via its API and revoked API access. As we have explained in the context of U.S. antitrust regulations, regulators should conduct in-depth reviews that assess these competition threats posed by vertical mergers.

See OTI and Public Knowledge's *Comments on the Draft Vertical Merger Guidelines*
See OTI's Report *Promoting Platform Interoperability*

In your view, what practices related to the use and sharing of data in the platforms' environment are raising particular challenges?

The open and diverse internet of the past has given way to concentrated power and data in the hands of a few large companies. Those companies now control most of the traffic on the internet. Online platforms such as social media are typically not interoperable. These platforms have become "walled gardens" within the larger context of the internet. In the past, interaction between people on the internet might have taken the form of links from one website to another, comments on blog posts, emails sent from one organization's server to another, or posts on a given newsgroup on Usenet (the first message board system, which operated without a single centralized server, and instead shared postings among many servers to which individual clients connect). Today, all of those actions are likely to take place entirely within the confines of a single company's services.

Interoperability decreases barriers to entry and facilitates greater competition by enabling new players to offer access to the users on, and at least some of the features of, the entrenched platforms. It also expands the overall market for a particular service or type of service by letting third parties fill in the gaps around the platform's feature set, as many games and other apps have done with Facebook's platform. Interoperability is a promising lever for regulators to use in their efforts to oversee and correct monopolistic abuses amongst the dominant online platforms. It has a unique ability to promote and incentivize competition—especially competition between platforms—and can also offer users greater privacy and better control over their personal data generally.

When platforms acquire smaller companies in adjacent markets, they often acquire user data that can be consolidated to give the platforms a unique competitive advantage. Usage data, information about how individuals use a product, is unique and cannot be easily replicated by competitors. OTI recently wrote to the European Commission to express our views that the data advantage Google would gain by acquiring Fitbit's data would not be remedied by creating a silo for the health and wellness data to prevent it from being used for targeting advertising. OTI explained that the merging of Google and Fitbit data would threaten both competition and user privacy.

See OTI's Report *Promoting Platform Interoperability*
See OTI's European Commission Submission *Re: Merger Case Number M.9660, Proposed Acquisition of Fitbit by Google*
See OTI's Article: *Pressure on Google Is Ramping Up. Could the Antitrust Probes Help Address Privacy Harms?*

Which are possible positive and negative societal (e.g. on freedom of expression, consumer protection, media plurality) and economic (e.g. on market contestability, innovation) effects, if any, of the gatekeeper role that large online platform companies exercise over whole platform ecosystem?

The gatekeeper role that large online platforms exercise can threaten freedom of expression, privacy, human rights, and civil rights. On the one hand, companies can democratize speech in some instances, by providing an online platform that can enable individuals to publicize their messages to wide audiences. In this way they have the ability to promote free expression. On the other hand, these companies also act as gatekeepers, and so their content policies and practices effectively determine who has a voice online and defines what speech is permissible. Further, large online platforms' reliance on algorithmic tools for content moderation can result in disproportionate harms to members of minority groups, such as through discriminatory targeting and delivery of ads for housing, employment, and credit. In addition, as described previously, market consolidation can have negative impacts on innovation and create significant barriers for new entrants.

IV. Advertising and Smart Contracts

From your perspective, what measures would lead to meaningful transparency in the ad placement process?

Internet platforms must provide greater transparency around their digital advertising operations. Algorithmic ad targeting and delivery systems enable advertisers to specify which categories of users they would like to target with their ads. This can result in users receiving relevant ads, but it can also result in the discriminatory exclusion of certain users, even when an advertiser sets non-discriminatory targeting parameters. This is because ad delivery algorithms make inferences based on engagement metrics and other data to identify users that are more likely to engage with an ad. Studies have shown that this can reinforce and exacerbate societal biases regarding race, gender, and socioeconomic status in housing, employment, and credit. Platforms have been slow to take action to redesign their ad algorithms to avoid perpetuating discrimination since digital advertising underpins their business models.

In order to provide greater transparency and accountability around digital advertising, internet platforms should:

1. Publish comprehensive descriptions of advertising *content* and *targeting* policies which outline what categories of ads, types of ad content, and accounts are prohibited on the platform, what information the platform and advertisers can use to target ads to users, which targeting parameters are prohibited on the platform, and what tools and processes the platform uses to identify ads and accounts that violate its ad targeting policies.

2. Prohibit targeting based on protected classes and sensitive characteristics that could result in discriminatory outcomes, including characteristics that have been shown to be proxies for protected characteristics.

3. Establish and disclose a comprehensive human review process for categories of ads that could have significant real-life consequences such as political, housing, education, employment, and financial services-related ads before they are permitted to run on a platform.

4. Hire independent auditors to conduct regular periodic audits of ad *targeting* and *delivery and optimization* algorithms in order to identify potentially harmful outcomes and take steps to eliminate or mitigate any harms identified through the audits.

5. Provide users with detailed explanations that help them understand how and why ads are targeted and delivered to them, why the platform collects, infers, and shares user data.

6. Improve user controls so that users can easily manage whether and how data is collected, inferred, and shared, how this data is used, and how it influences the content that they see. This should include the option to delete this data entirely and to opt-out of receiving targeted advertising entirely.

7. Provide clear labels for sponsored and paid content across all of the platform's products, services, and ad networks.

See OTI’s Report *Special Delivery: How Internet Platforms Use Artificial Intelligence to Target and Deliver Ads*

What information about online ads should be made publicly available?

Internet platforms should publish a transparency report that provides a granular overview of the platform’s advertising operations across all regions that it operates in.At a minimum, this transparency report should disclose: The total # of ads a platform ran, the total # of ads a platform ran in each country in which it operates, the total amount of ad spend across the platform, the total amount of ad spend in each country in which it operates, the top advertisers in each country, the top keywords in each country.

In addition, at a minimum, this transparency report should separately disclose the following information for ads that have been flagged or removed from the platform for every reporting period:

- The total # of ads flagged for violating the platform’s ad content policies
- The total # of ads removed for violating the platform’s ad content policies
- The total # of ads flagged for violating the platform’s ad targeting policies
- The total # of ads removed for violating the platform’s ad targeting policies
- A separate breakdown of the ads and accounts flagged and removed for violating the platform’s ad content policies by:
    - The ad content policy they violated
    - The format of the ad’s content (e.g. text, image)
    - The product or service on which the ad was run
    - The detection method used (e.g. user flag, automated tool)

- A separate breakdown of the ads and accounts flagged and removed for violating the platform's ad targeting policies by:
    - The ad targeting policy they violated
    - The format of the ad's content (e.g. text, image)
    - The country of the advertiser
    - The product or service on which the ad was run
    - The detection method used (e.g. user flag, automated tool).

Internet platforms should also create a publicly available online database of all the ads it has run on its platform. At a minimum, this database should disclose the following information about each of the ads in the database:

- The format of the ad (e.g. text, video)
- The name of the advertiser
- What region the ad was run in
- How much the ad spend for the ad was
- The time period during which an ad was active
- Granular engagement and interaction information, (e.g. how many users saw the ad, the number of likes, shares, and views an ad received)
- What targeting parameters the advertiser selected
- What categories of users the ad was delivered to (i.e. what targeting parameters did the ad delivery system select and optimize for)
- Whether the ad was delivered to a custom set of users or one generated by an automated system (e.g. Lookalike users)

See OTI's Report *Special Delivery: How Internet Platforms Use Artificial Intelligence to Target and Deliver Ads*

What information disclosure would meaningfully inform consumers in relation to political advertising? Are there other transparency standards and actions needed, in your opinion, for an accountable use of political advertising and political messaging?

As we have outlined in response to questions #15, #16, and #21, there are a range of methods which internet platforms can adopt to provide greater transparency and accountability around their digital advertising operations, including their political advertising operations. These recommendations are drawn from our report *Special Delivery: How Internet Platforms Use Artificial Intelligence to Target and Deliver Ads*.

Are there other emerging issues in the space of online advertising you would like to flag?

Internet platforms should provide meaningful notice to advertisers who have had their ads or accounts flagged or removed, as well as to users who have flagged ads or accounts. These notice procedures are particularly important where ads are run by individuals or civil society organizations, as erroneous removal or moderation of their ads could particularly infringe on freedom of expression. In addition, given the lack of clear definitions around categories of ads such as political and issue ads, such notice processes are important to protect freedom of expression. All notices should be available in a durable form that is accessible even if an advertiser’s account is suspended or terminated. In addition, users who flag ads should have a log of ads they have reported and the outcomes of the review process.

In addition, with regard to categories of ads that could have significant real-life consequences, such as political ads, housing ads, employment ads, and credit ads, internet platforms should offer advertisers who have had their ads or accounts flagged or removed as well as users who have flagged ads or accounts a robust appeals process.