Your response

Question	Your response
Question 1: How, and to what extent, are persons carrying out independent research into online safety-related issues currently able to obtain information from providers of regulated services to inform their research?	Confidential? – Y / N Social media and other digital technologies have transformed society, making it easier to find information, engage with politics, and connect with people globally. However, these digital tools can help fuel misinformation, enable harassment, and foment polarization, presenting urgent challenges to online safety and democratic governance. At New York University's Center for Social Media and Politics (CSMaP), we strengthen democracy by conducting rigorous studies on how the online environment impacts politics, policy, and democracy, making social media data foundational to our work. Social media data is the foundation of our research into online safety-related issues. But, since we began our work over a decade ago, we've constantly had to negotiate and creatively obtain platform data. There are three major pathways for doing so: ● First, we can access data from **platform APIs**. These include documented researcher APIs such as the Meta Content Library as well as undocumented APIs. While each platform's research programs and data access varies, what we've experienced is that our research is subject to platform decisions–oftentimes only sharing a portion of data or revoking access tools with minimal notice. ● Second, we can **collect data ourselves**. For example, we can develop data donation pipelines where survey respondents can share their social media data with us for analysis. While this access modality can provide insight into the "typical" timelines and content of users, it requires large financial and time investments from our research teams to ensure a robust sample. ● Last, we can obtain data through **web scraping**, a style of automated data collection that captures data that is rendered on a webpage. While scaping can enable access to otherwise inaccessible data, it can pose legal, ethical, and institutional challenges to researchers which must be carefully weighed.
Question 1a: What kinds of online safety research does the current level of access to information enable? ● What type of independent researchers are carrying out	Confidential? – Y / N Our center leverages big data and experimental methods to explore how various types of information—including but not limited to mis/disinformation—spreads online; how this information impacts people's attitudes, beliefs, and behaviors;

Question	Your response
research into online safety matters? ● What topics/issues they are researching?	what interventions—such as labeling on generative AI content—might make a difference to mitigate harms, and more. Our research has become increasingly vital as policymakers reshape the laws and regulations governing the online sphere. While updated policies are necessary to protect individuals from harm and ensure a healthy democracy, their success hinges on the evidence that informs them. For example, our work has challenged flawed assumptions about digital media in political conversations, such as the belief that many users share fake news online. Hence, without social media data access to produce high-quality empirical evidence, legislators risk operating in the dark. Through our research and data infrastructure, we directly engage with policymakers, civil society groups, industry professionals, and journalists working to tackle today's online safety problems.
Question 1b: Are there types of information that independent researchers are currently unable to access that may be relevant to the study of online safety matters? If so, what are they and what kind of research would they facilitate?	Confidential? – Y / N Under the current data access mechanisms, researchers have been able to study several critical issues to online safety. For example, research has provided insights into the recommendations of algorithmic systems, the patterns and effects of foreign influence, the prevalence of hate speech and harassment, and the efficacy of interventions. However, some platform data are limited or entirely inaccessible to researchers. The following are four areas where data is frequently hard to access and the subsequent research areas it impacts: ● **Video Data for Multimodal Research**: While social media platforms used to be largely text-based, new affordances and new platforms, such as TikTok, have contributed to the rise of video-dominant content which poses new questions about the impact of image and video-based content. ● **Non-English Language Content for Multilingual Research**: Despite the linguistic diversity of users on platforms, the majority of platform research focuses on English language content. Data access provisions can help close this gap by making multilingual content more readily accessible to research teams. ● **Standardized Metrics for Cross-Platform Research**: The social media ecosystem continues to fracture and users typically exist on more than one platform. There are both significant challenges such as identifying users across platforms and comparing content across platforms. Some

Question	Your response
	data can be standardized by using uniform column headers where possible (such as standard variable names for data types) and defining measures (such as a view, impression, or engagement), allowing for research that better captures a comprehensive picture of the information environment. • **Generative AI Data for Understanding Emerging Technology**: AI models are expected to significantly impact the information environment, for example, facilitating the creation and spread of misinformation and "slop" content. Mandated data access to datasets related to AI, such as content label warnings for AI, will help researchers better track and understand the impact of the burgeoning AI landscape on the information ecosystem.
Question 1c: What data governance models are currently used to allow access to online services' information for researchers? • This might include: open-access forms of information-sharing, such as publicly-accessible information libraries or databases; information-sharing models that rely on vetting or accreditation of individuals or organisations; and/or models that rely on the accreditation of the specific use cases for the information. • Please provide relevant examples of these governance models used in the online services industry.	Confidential? – Y / N Governance models for digital data have been proposed, piloted, and implemented in recent years. Below, we outline models that take a range of approaches—from centralized bodies to grassroots organizations to highlight their relative strengths and tradeoffs. What this data governance model shows is that close consultation with users (researchers) and sufficient capacity (personnel and funding) are vital to running and developing an effective and efficient data-sharing system. Some models are expansive, institutional mechanisms. For example, the data access provisions in the DSA promise broad access to platform data but require researchers to undergo an application process for approval and potential mediation for rejected claims, which may prove to increase timelines for data access. On the other hand, there are crowd-sourced models, such as the Media & Democracy Data Cooperative, where networks of researchers collaboratively collect digital data and share best practices. Here, researchers might be able to more quickly access and share data amongst themselves, these community networks are slow to scale, requiring sufficient funding and technical expertise. Data access regimes can also come from the platforms themselves. For example, the data analytic tool, CrowdTangle, and its replacement, the Meta Content Library (MCL), which is managed at the University of Michigan's Social Media Archive (SOMAR). While this latter model for provisioning access has undergone multiple iterations, it has faced criticism for data quality and application timelines. CrowdTangle and the MCL show platform-managed access can offer well-resourced data libraries. However, the data quality and

Question	Your response
	the accessibility of the tools themselves remain at the whims of tech companies to ensure ongoing access. Overall, there are tradeoffs inherent to each of these data governance regimes. Yet, these tradeoffs demonstrate that effectiveness and efficiency of any data-sharing system hinges on two critical factors: in-depth consultation with researchers and adequate resources, both in terms of staffing and financial support to establish a robust data exchange platform.
Question 1d: What technologies are typically used by providers of online services to facilitate existing information access?	Confidential? – Y / N In our experience, secure access modality technologies are a safeguard for ensuring that sensitive data is collected, shared, and stored responsibly. While this varies by providers, a widely used modality is virtual clean room environments, which provide highly controlled settings to minimize security and privacy risks. Clean room software offers security features like built-in de-identification capabilities, encryption protocols, secure file transfer, and access provisions to ensure data is shared with only approved users. Some examples of what a secure computing environment could look like is the Facebook Open Research and Transparency (FORT) Platform and the Social Media Archive (SOMAR), which provides a secure way for qualified users to access privacy-protected Facebook and Instagram data. Another example is the U.S. Federal Statistical Research Data Centers (FSRDCs). These environments typically comply with data security safeguards required by university Institutional Review Boards, thus reducing barriers to facilitating access.
Question 1e: Have services and/or researchers made use of privacy-enhancing technologies to enable access?	Confidential? – Y / N

Question	Your response
Question 2: What are the challenges that currently constrain the sharing of information for the purpose of research into online safety related issues?	Confidential? – Y / N Over the last year, researchers' access to social media data has dwindled, with platforms like Twitter/X and Reddit implementing changes that have made it more difficult for external researchers to access data, and Meta discontinuing its popular analytic tool, CrowdTangle. However, the challenges that limit researchers' access extend beyond platform data-sharing programs. Issues

Question	Your response
	including platforms' privacy concerns, minimizing risks associated with sensitive user data, and the lack of standardized security protocols complicate efforts to securely facilitate the sharing of sensitive personal data. Platforms vary in how they approach data privacy for sensitive personally identifiable data, often leaving researchers responsible for ensuring the data is properly structured and secure to prevent misuse or possible legal violations. These responsibilities may also be particularly challenging for less-resourced researchers and non-technical researchers. Without standardized security frameworks, platforms may be hesitant to share users' data, fearing reputational damage or legal liability. Policymakers can address this challenge by enacting policies mandating that platforms share data with vetted, public interest researchers. For example, Article 40 of the European Union's Digital Services Act, (DSA) offers a legislative framework granting vetted researchers access to very large online platforms (VLOPs) and search engines (VLOSEs). However, the DSA also has its limitations—the law needs to streamline the application and response processes, improve documentation of data inventories, clarify data security and protection measures, and address other gaps. Ofcom has a unique and important position to develop a comprehensive data-sharing law that addresses these privacy and security challenges head-on and supports researchers' ability to study online safety effectively.
Question 2a: What are the legal challenges/risks to sharing information from online services with independent researchers?	Confidential? – Y / N
Question 2b: What are the technical challenges relating to sharing information from online services with independent researchers?	Confidential? – Y / N
Question 2c: What are the security challenges relating to sharing information from online services with independent researchers? ● What are the security challenges relating to the potential sensitivity of information?	Confidential? – Y / N Security in sharing information is both a governance and technical issue. While online services have a responsibility to provide independent researchers access to information securely, they do not have a shared understanding or set standard for instituting security protocols.

Question	Your response
• What are the security protocols required to protect information from misuse? • To what extent do you view security as a governance issue compared to a technical infrastructure issue?	Therefore, regulators should clarify which security measures are sufficient under a potential data-sharing law, such as standardizing what constitutes a "secure" data-sharing environment. One method of doing the latter is clarifying what "security" means in a future data-sharing law, helping platforms develop a shared understanding of the privacy and security requirements when providing information to researchers, thereby reducing any security issues. Another way to standardize security is through a secure, government-run, computing environment for accessing sensitive data online. A model of such is the previously discussed United States Federal Statistical Research Data Centers (FSRDCs). These environments frequently comply with data security safeguards required by university Institutional Review Boards, allowing researchers to work with data in an environment that meets multiple data security compliance mechanisms.
Question 2d: What are the information quality challenges relating to online services sharing information with independent researchers?	Confidential? – Y / N As mentioned previously, data-sharing governance can help enable crucial cross-platform research while also ensuring that data quality is established and audited. To do so, the structure of data ought to be standardized where possible, for example through uniform column headers, standard variable names for data types, and defining measures such as a view or impression. Addressing this data quality challenge head-on would help direct analytical comparability and aid researchers studying online safety issues such as the direction and spread of misinformation. Secondly, researchers who access platform data must have some level of assurance that the data is accurate and complete. Data-sharing bodies must ensure mechanisms for auditing or otherwise ensuring the quality of platform data being shared with researchers.
Question 2e: What are the financial costs to online services relating to online services sharing information with independent researchers?	Confidential? – Y / N
Question 2f: What are the financial costs to researchers trying to make use of information shared by online services?	Confidential? – Y / N Platform-managed researcher application programming interfaces (APIs) allow researchers to analyze large amounts of real-time, structured public data and have been a crucial data access pathway for supporting online safety research. However, Researcher API programs can change suddenly, often at the

Question	Your response
	discretion of the platforms and with little notice, leaving researchers vulnerable to corporate interests. For example, X/Twitter used to provide researchers with university affiliations with free or heavily discounted API access. Now, the cheapest API package offers access to only 50 million tweets for **$42,000/£33,661.94 a month**, effectively blocking many researchers from accessing the platform's data and impacting hundreds of projects on critical online safety topics. As an academic research institute, we, like many others, are primarily funded by grants. Some researchers may find themselves in positions to align research with the interests of the grant, which may come from a platform, rather than pursuing other vital questions they may hope to explore. These financial constraints are particularly challenging for researchers outside of well-resourced institutions, potentially excluding smaller research institutions or labs in the Global Majority from conducting and publishing research on critical online safety topics. Moreover, reliance on Researcher APIs can limit research on online safety to those with advanced technical expertise. This restriction potentially narrows the scope of research from other disciplines, including those in the humanities and other interdisciplinary approaches, which may rely less on computational methods but still provide equally valuable analysis for understanding online safety and platforms' societal impacts. The challenges above emphasize the importance of governments mandating social media access to public interest researchers. In doing so, regulators should also regularly consult researchers—technical and non-technical— to help standardize a data-sharing environment that accommodates the diverse range of technical skills necessary for studying online safety.

Question	Your response
Question 3: How might greater access to information for the purpose of research into online safety issues be achieved?	Confidential? – Y / N Facilitating greater access to information for researching topics on online safety requires a multifaceted approach that leverages input and expertise from regulators, researchers, and platforms. Collaboration is especially vital as the online information environment continues to fragment, and research teams like ours at CSMaP work to study emerging platforms. Below, we outline ways to support this future research: What Data Should be Available

Question	Your response
	1. Public Data from Major Platforms: Given the significant user base, legacy platforms such as Twitter/X, Facebook, and Instagram remain critical sources of public data for our teams studying online safety-related issues, including the spread of mis/misinformation, political polarization and its potential impact on society, and others. At CSMaP, access to public data has enabled us to build a robust data infrastructure that contains billions of social media posts. Utilizing big data analysis and experimental methods, we challenge assumptions about social media and guide tech policy debates with empirical evidence. However, sudden data access policy changes imposed by platforms and the high costs of accessing certain platforms' APIs have hindered researchers' ability to conduct comprehensive analyses. Regulators can address this by mandating access to public data, ensuring that our research continues producing actionable insights that inform public policy. **2. Emerging and Smaller Platforms**: Access to user-specific data offers unique opportunities to study how certain individuals engage with online content, their unique risks, and the effectiveness of specific intervention methods. For example, a recent paper CSMaP published highlights that while misinformation reaches users across the political spectrum, those with extreme ideologies are more likely to encounter and believe false information. This analysis also showed that misinformation spreads quickly, often reaching the most susceptible users first. We recommend that to mitigate its impact, interventions must be targeted and deployed rapidly to those most at risk of believing falsehoods. Policymakers can help facilitate access to such data by drawing from competition laws in the UK and EU that empower users to control their data, making the process for those who wish to share their social media digests with researchers more accessible. Abiding by competition laws also ensures that research and platforms comply with security and privacy laws. **3. Platform-Specific Data**: While much of the current research on online safety is concentrated on legacy platforms like X/Twitter and Facebook, the social media landscape is no longer dominated by these sites. Newer, broadcast-style entertainment platforms, such as Twitch, private messaging apps like WhatsApp and Telegram, and other niche platforms, such as Gab and Rumble, have risen in their use over the last few years.

Question	Your response
	Hence, as the online information environment becomes more fragmented, potentially unique systemic risks may arise from these new platforms. At CSMaP, we are interested in exploring these online environments to understand the changing digital landscape and identify new patterns in the spread and impact of harmful content. However, imposing data-sharing mandates on smaller platforms could strain their limited resources. Regulators should then consider conducting periodic reviews of the information ecosystem to identify emerging platforms where risks may arise. Collaborating with researchers and platforms, regulators can help design platform-specific data-sharing requirements that balance research needs with platform capacities.
Question 3a: What models, arrangements or frameworks exist for allowing researchers access to sensitive information beyond the online services industry? What are the benefits and risks of those models, and how might they apply to the online services context?	Confidential? – Y / N
Question 3b: Are there any models or arrangements that exist in the online services industry already that might provide increased access to information for research purposes if applied more generally across the industry? If so, what are these and what are the benefits and disadvantages of these models/arrangements?	Confidential? – Y / N
Question 3c: What are some possible models for providing researchers with access to relevant information that may not exist or be widely used yet, but which might be implemented by industry?	Confidential? – Y / N A potential data access pathway for providing researchers with access to relevant information is through the industry widely implementing effective data donation mechanisms—a data collection method where users consent to donate their digital trace data for research. The process for enabling data donations is not complex for platforms since many already have portability mechanisms embedded into their interfaces as mandated by Article 20 of the European Union's General Data Protection Regulation (GDPR). Data portability provides users the legal right and technical ability to transfer their data from one digital service to themselves and/or to other digital services.

Question	Your response
	Given platforms' capabilities, regulators can make data donations a widely implemented feature by mandating platforms design systems supporting data donations accessible to users and researchers.
Question 3d: What are the advantages and disadvantages of this approach? ● These may include elements pertaining to financial, legal, security, technical or feasibility issues	Confidential? – Y / N There are several legal and scientific benefits to the industry implementing data donation mechanisms on platforms. First, data donations involve explicit, informed user consent, unlike data from Research APIs. If legally mandated, as seen in GDPR's Article 20, data donations also ensure researcher protection and remain resilient to platform policy changes, since users control their data. Scientifically, data donations allow researchers to collect data across multiple platforms, including emerging platforms—such as alt platforms (e.g., Gab or Parler), local platforms (e.g., Nextdoor), video game platforms (e.g., Twitch), and private messaging apps (e.g., Telegram)—that are becoming increasingly important. Additionally, data donations allow participants to donate data from multiple platforms, offering a richer and more comprehensive view of their online information diets. This capability is especially important as individuals, particularly younger users, engage with multiple platforms. However, key limitations exist that limit data donations' broad use to study the digital information environment. For example, the process of downloading and donating data to a study can be an arduous process, potentially leading to attrition from participants, and challenges for sufficiently-powered research. Additionally, there is no standardized format for downloading a data file across platforms, requiring researchers to provide significant assistance to participants. Lastly, donated data is not always structured appropriately for analysis, adding another level of complexity for researchers.
Question 3e: What role could third party organisations, such as regulatory bodies, civil society or public sector organisations have in facilitating researcher access to online safety information?	Confidential? – Y / N Third-party organizations, such as regulatory bodies, civil society groups, and public sector organizations, can play a pivotal role in facilitating researcher access to online safety information: ● **Research Consortiums**: These third-party organizations can convene policymakers, regulators, and companies to design and implement data-sharing systems aligned with

Question	Your response
	researcher needs. These consortiums engage the expertise of regulators in data rights, policy experts in platform transparency, and platform engineers to create systems that allow researchers to easily and securely access online safety information. • **Regulatory Access Programs**: A regulatory body can develop its own information access program for social media data to complement potential legislation mandating researcher access to online safety information. • **Researcher-Facilitated Infrastructures**: Researchers can facilitate access to online safety information by building and maintaining a shared infrastructure for data collection, analysis, and more. Civil society and public sector organizations can also collaborate with researchers to sustain and grow these shared infrastructures. • **External Collaboration on Experimental Research**: Regulators might also consider mandating collaboration with external researchers on implementing experimental studies to understand the causal impact of social media usage at regular intervals. One useful analogy here is the way that financial regulators mandate stress tests of financial institutions.
Question 3f: What could these third-party models look like, and what are some of the benefits and challenges associated with this approach?	Confidential? – Y / N Third-party models facilitating researcher access to online safety information can take several forms: • **Research Consortiums**: A research consortium facilitating researcher access to online safety information could establish mechanisms for data transfers from major platforms, and researchers could interact with that consortium to support their particular projects. The Social Media Archive (SOMAR) at the Inter-university Consortium for Political and Social Research (ICPSR) is an excellent model for a type of consortium that negotiates and facilitates data access between companies and researchers. SOMAR provides a centralized and reliable repository for social media research data, with its archive containing a range of data collected from large-scale social media platforms such as X/Twitter, Facebook, Instagram, and Reddit, and smaller, more specialized data sets focused on specific research topics. However, to facilitate other forms of data sharing, such as data donations, a consortium would need platform buy-in. • **Regulatory Access Programs**: Regulatory bodies can develop their own access programs for social media data

Question	Your response
	by looking to successful data-sharing models in the United States, such as the Food & Drug Administration (FDA)'s ClinicalTrials.gov. This platform ensures companies, universities, and other parties make clinical trial results available to researchers through established standards for data sharing. This standardization also allows the FDA to protect clinical trial data's quality, accuracy, and useability, making it easier for researchers to access and use protected data. A regulatory body could adopt similar mechanisms from the FDA's database to require that social media companies share certain data that is securely stored and made accessible to researchers. These programs can also help streamline access to social media data, standardize data file formats, and ensure researcher access complies with privacy laws. ● **Researcher-Facilitated Infrastructures**: While the aforementioned suggestions require time, resources, and multi-stakeholder collaboration, researchers can facilitate access to online safety information by building and maintaining a shared infrastructure. Projects such as the *Accelerator* infrastructure project, which is designed to power policy-relevant research, are a great model. By building shared infrastructure to support data collection, and analysis the *Accelerator*'s shared tools enable researchers to efficiently access and analyze data, save time and resources currently spent on duplicative engineering work, and more.
Question 3g: What categories of information should online service providers give researchers access for the study of online safety matters? Why would this information be valuable for the study of online safety matters?	Confidential? – Y / N The ability of academic researchers to independently study online safety-related issues on social media requires rich digital data about user and platform data. Although the following is by no means an exhaustive list, in general, the data we need for research falls into a few broad categories: ● **Exposure Data**: What posts appear in people's timelines when they're actually using platforms. ● **Engagement Data**: What posts people have clicked, shared, liked, etc. ● **Recommendation Data**: What content the platforms recommend that people view and which accounts, pages, and/or groups they recommend users follow. ● **Network Data**: Friend/follower relationships.

Question	Your response
	• **Content Moderation Data**: A catalog of actions taken against a post or account (if it's flagged for review or removed, for example). • **Access to Platforms for Running Experiments:** While descriptive data is enormously valuable, many of society's most pressing questions about the causal impact of social media usage can best be answered using experimental methods, many of which require the cooperation of platforms to both plan and implement. Altogether, this information is valuable for empirically studying the online environment and safety matters. Not only does it provide evidence for dynamics and safety concerns, but research informed by platform data can have wide-reaching impacts on public policy, civil society, and platforms themselves.