I am a writing to submit a response to Ofcom's call for evidence related to researchers' access to information from regulated online services. I am a senior resident fellow at the Atlantic Council's Democracy + Tech Initiative, which resides within our Digital Forensic Research Lab.

That unit, created in 2016, is comprised of more than 50 researchers, spread across the United Kingdom, European Union, North America and Latin America. It has become a global resource for tracking harms associated with the information environment, and relies heavily on information access to social media platforms to conduct quantifiable research to support well-informed policymaking.

That includes extensive use of existing information access mechanisms made available by many of the regulated online services that fall under the United Kingdom's Online Safety Act. It also relies on other mechanisms, including third-party services, through which we have garnered extensive understanding of 1) how access to information at social media platforms work in practice; 2) what online harms are evident across these online services and how access to information can highlight such issues; and 3) what is required, going forward, to make such access to information for independent researchers more practical as the United Kingdom's government contemplates future legislative efforts.

It is worth remembering that any future mandated access to information within the United Kingdom should not be taken in a vacuum. Lessons from the European Union's separate data access regime, most notably via Articles 40.4 and 40.12 of the bloc's Digital Services Act, should be considered as the United Kingdom assesses the need, if necessary, for such a mandatory regime for information access.

That includes differences between the types of access and information required for different types of researchers. While not comprehensive, such differences can fall into two categories. 1) Real-time access to public information that allows organizations like civil society groups and journalists to find possible immediate harms, as outlined under the United Kingdom's Online Safety Act; and 2) longer-term access to private information that allows organizations like academic institutions to identify so-called 'systemic risks' associated with the United Kingdom's approach to online safety.

Both types of access to information require a thought-out approach that prioritizes individuals' security and privacy, above all. But they do necessitate different approaches to access to information, given the different priority areas for different types of researchers.

As the United Kingdom contemplates a potential mandatory regime for information access, policymakers, politicians and regulators should lay out a clear vision for why such access to information is required under the country's online safety regime.

For the Atlantic Council, that would include a prioritization of greater access to information to boost accountability and transparency for the regulated online services that fall within the United Kingdom's Online Safety Act. It should also focus on promoting such principles in a way that protects individuals' fundamental rights, including those related to privacy and free speech.

Access to information, in general, should not be envisioned as a mechanism for policing individual social media posts. At its core, such mandated access is best served to 1) support

Ofcom's existing powers related to transparency and accountability; 2) informing policymakers and regulators of potential harms via enabling independent researchers in their efforts to surface problematic areas; 3) supporting future legislative and regulatory responses that can be based on quantifiable evidence of harm, based on a systemic approach to access to information.

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?

Voluntary access to information regimes has existed for almost a decade. They grew out of platforms' responses to wider societal and regulatory calls for greater awareness of what took place on these online services and their effect on the wider society. Different regulated services, as defined under the United Kingdom's Online Safety Act, have approached the task in different ways. Without broad consensus, either between platforms or countries, there is mostly a free for all in terms of how access to information is granted. That includes the arbitrary cancellation of access to information regimes — or the imposition of high prices to what had been free-to-use information access regimes — as has recently occurred at X/Twitter and Reddit, respectively.

In general, current access to information regimes fall into four categories:

- **Platform-enabled online interfaces** that include online services providing user-friendly interfaces from which independent researchers can access information. Examples include Meta and Google's political advertising databases and Meta's Content Library, which is housed within the University of Michigan's Social Media Archive. While the latter example is technically housed in an independent body, it should be viewed through the lens of platform-enable online interfaces.
- **Application Programming Interfaces, or APIs,** that allow technically literate researchers to directly access a pipeline of allotted information that is provided directly from the online service. This option requires a high level of technical literacy, including knowledge of the likes of Python or R, but can allow researchers to more independent collect and analyse information compared to platform-enabled online interfaces. Repeated research, however, has demonstrated that such APIs do not accurately correspond to what social media users view within their own feeds. TikTok's research API, for instance, has shown significant disparities, including related to engagement statistics, compared to actual social media posts displayed within users' feeds.
- **Public-interest data scraping** that allows independent researchers to directly take public information from online services for analysis. Such provisions form the basis of how much of the internet currently works, for example how Google's search service "scrapes" websites to collate information for public consumption. Not all platforms, however, are supportive of this approach, and the United Kingdom's Information Commissioner's Office similarly has raised privacy concerns related to scraping. It can potentially violate the United Kingdom's existing data protection rules by permitting researchers to collect information that is either sensitive in nature or in a way that does not allow citizens the ability for their information to be deleted. A clear definition, mandated by legislation or regulatory guidance, of what represents "public information" is a requisite for such an approach.

— **Data donations** that allow individuals to voluntarily donate their social media information to researchers for analysis. This can be automated via browser plug-ins or via online forms from which citizens can download their information from regulated services (through existing forms made available by the companies) and then submitted to independent researchers for analysis.

Question 1a: What kinds of online safety research does the current level of access to information enable?

The current ecosystem of access to information regimes allows for widespread research related to the potential harms as outlined under the United Kingdom's Online Safety Act. It can provide significant value, for both Ofcom and policymakers, in identifying areas for concern and how best to respond to such issues. The main barriers for such online safety research, though, are both technical and qualitative.

Technical, in terms of researchers' ability to access often technical tools, see section above. And qualitative, in terms of researchers' access to employees at regulated services that can fast-track access. Currently, access to information via voluntary regimes is predicated on personal ties and networks to these employees. That has relegated research to a handful of primarily US academic institutions. The purpose of mandating widespread access to information regimes is to level the playing field so those researchers either within the United Kingdom or those elsewhere who are researching topics within the country have equal access to such access to information regimes.

- **What type of independent researchers are carrying out research into online safety matters?**

  As outlined above, researchers fall into multiple categories, including those related to academic institutions, civil society organizations and the media. While the current United Kingdom's Online Safety Act sets the bar relatively low for individuals who may be able to research such topics via access to information regimes, it is necessary to institute specific checks, most notably to ensure researchers have adequately addressed data protection issues, before they are able to access such information.

- **What topics/issues they are researching?**

  Just as there is a diversity of researchers, the topics that are researched is equally divergent. Many of these topics fall squarely into the list of potential harms, as outlined by the United Kingdom's Online Safety Regime. That ranges from tracking foreign interference and questions around election integrity to terrorist content and child online sexual material.

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?

There is a significant amount of information from regulated online services that remains off limits, or unknown, to independent researchers. Repeated internal documents made public by company whistleblowers, most notably Meta's Frances Haugen, detail extensive internal data that has not been made public to outsiders. Identifying that type of information, from the outside, is

next to impossible. As a first point, regulators should mandate the public disclosure from these companies about the types of information that they collect — and how it may interact with each other — so that independent researchers have a clearer understanding about what internal platform information may be available for independent research.

Question 1d: What technologies are typically used by providers of online services to facilitate existing information access?

See answer to Question 1. Currently, the primarily modalities used by providers of online services to facilitate information access are either via clean room (in the case of Meta); APIs (in the case of TikTok) or public-interest scraping (in the case of Alphabet and Amazon).

Question 1e: Have services and/or researchers made use of privacy-enhancing technologies to enable access?

As part of existing application process related to access to information, researchers must include specific data protection disclaimers, including the use of a data protection impact assess, articulation about how they will keep data secure and obligations for removal under existing national/regional privacy legislation.

The knowledge of privacy-enhancing technologies across the independent research community is varied. Government and the regulator can play a fundamental role in this area by facilitating capacity to make researchers aware of their data protection requirements. They can also support with cross-institutional legal documents, including data protection impact assessments, to boost community capacity and awareness.

Question 2: What are the challenges that currently constrain the sharing of information for the purpose of research into online safety related issues?

The challenges fall into three primary areas:

— Technical: most researchers who would benefit from any proposed access to information regime do not have the technical understanding or capacity. They do not know how to manage an API or use computer programs like R or Python to access the necessary data. This has limited independent research to a small number of primarily academic institutions with the requisite knowledge to conduct such research.

— Financial: Accessing social media information is costly. Researchers must spend limited resources of infrastructure, including cloud computing storage, to access and retain the information. They also must hire technical experts, see above point, whose salaries are significant. Third-party private data providers equally cost tens of thousands of pounds, annually, for subscriptions to access such information. Those within the civil society community, many of which have the sectoral knowledge to best ascertain potential harm under the United Kingdom's Online Safety Act, are equally constrained due to the lack of long-term core funding support, from philanthropic organizations of hands-off public funding.

— Relational: Currently, most wide-reaching access to information is conducted on a voluntary basis via interpersonal connections between regulated online services and those within the research community, many of whom have previously worked for these companies. That skews research topics to a small number of well-connected research institutions that inherently limits the equities in what type of research can be conducted.

Question 2a: What are the legal challenges/risks to sharing information from online services with independent researchers?

There is a significant litigation risk when access social media information via public-interest scraping. Many companies, most notably X, have taken an aggressive approach to enforcing their terms of service that includes provisions that forbid such public-interest data scraping. Others have imposed "anti-scraping" technical solutions on their services, even when their terms of service nominally allow such information gathering. That has reduced the ability for researchers to rely on public-interest scraping as a legitimate mechanism for information gathering.

The United Kingdom has some of the most stringent data protection rules in the world. This is a fundamental benefit to the country's citizens that should not be understated. Yet these rules also come with significant financial penalties for the mishandling of personal data — including in relation to independent research of regulated online services. Such privacy fines, especially for those in the civil society community with limited understanding of the UK's General Data Protection Regulation, are significant.

Question 2b: What are the technical challenges relating to sharing information from online services with independent researchers?

Companies rightfully worry that individual users' privacy and security rights may be infringed via sharing information within independent researchers. They also raised issues about proprietary intellectual property and business secrets that may be affected from such access to information regimes. So far, however, those concerns have not been demonstrated, via real-world research, to be as serious as the companies had first believed.

What are the challenges relating to the scale and complexity of the information involved?

Accessing and analysing information from online regulated services is complex and inherently incomplete. It remains difficult to quantify systemic risk due to the personalization inherent in how algorithmic recommender systems, as just one example, operate. On scale, researchers struggle to go beyond the anecdotal, in terms of accessing sufficient data to replicate potential findings across a statistically relevant dataset. On complexity, researchers can fall into the 'causation vs correlation' trap, when coming to conclusions, given the overlapping levers — some of which they may not be aware of — when it comes to how these online regulated services operate.

Question 2c: To what extent do you view security as a governance issue compared to a technical infrastructure issue?

It is both. Technically, researchers need to ensure that they approach such access to information in a way that relies on infrastructure that is sufficiently secure. Such a task is not as easy as it would appear. On governance, it is important that researchers consider both privacy and security questions when they first develop research proposals and then bake such governance structures into how they approaching data gathering and storage/analysis.

Question 2d: What are the information quality challenges relating to online services sharing information with independent researchers?

One primary area of concern is the ongoing differences between what data can be gathered via companies' APIs and what data is displayed in people's online feeds/accounts. Currently, those two datasets almost always do not align. There are various reasons for why this may occur, including internal content moderation policies that are applied to one of the datasets and not the other (re: removals of engagements, etc).

Researchers therefore can't rely on APIs as a like-for-like snapshot of what actual users are displayed on these regulated online services. To mitigate this issue, ongoing audits of APIs, coupled with the use of other information gathering techniques like public-interest data scraping, is necessary to ensure any information gathered is as accurately as possible.

Clean room also have their drawbacks. In Meta's Content Library, for instance, only specific accounts, deemed to be sufficiently public based on their following numbers, are included for research. That is a legitimate choice, based on balancing research needs and protecting individuals' privacy. But it inherently limits the information quality that is available for researchers.

Question 2f: What are the financial costs to researcher trying to make use of information shared by online services?

The financial costs related to accessing information relation to regulated online services varies from free to tens, if not hundreds of thousands of pounds each year. That includes Meta's free-to-use Content Library to X's so-called data access "fire hose" that can cost, for maximum access, an estimated $120,000 a month.

Accessing APIs and public-interest data scraping also have varying costs, depending on the types of storage and technical capacity required to carry out such research. That can vary between $10,000 to $100,000 a month, based on the level of information access required.

To mitigate these community-wide financial restraints, what is needed is significant collective investment in underlying information access infrastructure that can make such information from regulated online services available to the largest research community possible at the lowest overall cost. That would represent a public good, for researchers, regulators and the wider United Kingdom, that would equally prove attractive to non-UK-based researchers to investment their time and capacities in support of the country's online safety regime.

Question 3: How might greater access to information for the purpose of research into online safety issues be achieved?

Currently, the independent research community acts in a silo. Existing funding, research and inter-personal constraints mean it is difficult, if not impossible, for independent researchers to obtain the scale to conduct quantifiable and replicable work. That is particularly true for those working in, or targeting their research at, the United Kingdom where the research community is relatively small compared to international counterparts.

What is required is scale. And for that to happen — in other forms of access to information (clean rooms, APIs, public-interest scraping, data donation) — is for the creation of underlying

technical infrastructure to facility widespread research in a privacy- and security-preserving way. Currently, such infrastructure — akin to what Meta's CrowdTangle offered before its closure — does not exist at scale. It would reduce the technical difficulties associated with information access; democratize the ability for a wide variety of researchers to conduct systemic research; and provide the United Kingdom's regulatory body with a well-functioning and accessible community of independent researchers to support regulatory compliance.

What is missing from the access to information conversation, both within the United Kingdon and elsewhere, is a means to create such underlying technical infrastructure that can be used in an apolitical, cross-organizational way as a public good. The ability for the United Kingdom's government to support the creation of such infrastructure, as part of its discussions around broader access to information for independent researchers, is fundamental and existential for the success of any future regime.

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?

I have outlined the four primary mechanisms for access to information above. What is important to understand that it is not a 'one or the other' approach to which modalities are required for improved independent researchers' access to information. It is all of the above, and more. Different researchers need different forms of information access for different reasons.

The goal of any future United Kingdom legislation should be to avoid any technical 'lock-ins' to specific forms of technical responses that may become redundant in the years to come. Any new legislation should take a principles-based approach to access to information that avoids language that gives preference to one form of access modality over another.

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?

As mentioned above, the fundamental unanswered question for access to information is scale. Currently, researchers are working in silos and are often replicating modalities to access information in ways that are costly, overly technical and redundant.

Public sector organisations like UKRI could play a fundamental role in mitigating these problems by providing the underlying technical infrastructure for access to information on which researchers, both from academia and civil society, could rely. It would represent a fundamentally shift in global approaches to access to information that would put the United Kingdom at the forefront of a burgeoning regulatory environment that will only expand in the years to come.

Regulatory bodies, including Ofcom and the ICO, could also work to provide regulatory guidance over the types of access to information that is legally permissible, and then provide legal capacity and training to support such access in a privacy- and security-preserving way. The role of the ICO should not be understated. At their core, regimes that promote access to information are based on data protection norms, and should be viewed, primarily, via that lens, and not one of content moderation or online safety.

Question 3f: What could these third-party models look like, and what are some of the benefits and challenges associated with this approach?

See above response on underlying technical infrastructure.

Question 3e: 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?

That question is currently impossible to answer, given the ongoing opacity in the types of information that online service providers collect and generate on their users. As part of any potential future legislation or updates to the Online Safety Act, the United Kingdom's government should first mandate a public-facing audit of what categories of information is available from these companies; what formats they take and how such information can be inter-linked for greater insight into how these services function.

The European Union, under its Digital Services Act, is conducting some of this work via the implementation of Article 40.4 that would prove beneficial for the United Kingdom's separate policymaking in this area.