

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? – N It is incredibly hard to access <i>detailed</i> information from providers of regulated services. I personally do research on and engage in knowledge-exchange and feedback with social media platforms, and with Meta and TikTok in particular. Their opacity has brought my colleagues Tom Divon, Pam Briggs and I to write a paper titled “Gaslighting,” showing how communicating with platforms as a user, journalist, activist <i>or</i> researcher alike is an exercise in sifting through lies, omissions and half-admissions that are often a means to avoid accountability and prevent PR-damage. You can find the paper here: https://journals.sagepub.com/doi/10.1177/29768624241303109 Platforms award resources to ‘stakeholder engagement’ sparingly, when their budgets allow it or when they need spending at the end of the year. In my experience of speaking to colleagues, it is rare that platforms will reach out to researchers proactively. When they do or when, in my case, connections are made through events, press reports, activism campaigns and/or complaints, it is incredibly hard to reach product people. Most of us are only left to communicate with policy and stakeholder engagement teams, who are of course more knowledgeable ‘insiders’ than the outsourced corporate comms consultancies platforms employ to talk to the press, but who deploy the same strategy: add layers of PR to deny systemic issues within their governance, only to admit ‘errors’ in content moderation when the examples are mounting up and therefore become too difficult to deny. Platforms are also choosy when selecting which researchers to engage with. I was specifically told by Meta that they were not sure about whether to engage with me, because I looked like “too much of an activist,” which I think speaks volumes about the level of scrutiny and questioning they are prepared to receive. And even though researchers who engage with them are forced to sign an NDA, I was then shocked to be asked if I wanted to take a picture at the Meta office to post on my social media to show I had consulted on some policy for them – showing how platforms <i>want to be seen</i> to engage with researchers for brand alignment, but don’t so meaningfully because they reveal too little information. In several instances of my experience of communicating with platforms, detailed in the papers below, companies have deliberately withdrawn information or lied about the implementation of their governance processes, such as

Question	Your response
	shadowbanning and malicious flagging. However, when confronted with this fact in meetings or at events, they deny this ever happens. https://www.tandfonline.com/doi/full/10.1080/14680777.2021.1928259 https://journals.sagepub.com/doi/10.1177/14614448241228544 https://journals.sagepub.com/doi/10.1177/13548565231218629 In short, platforms demand trust from their users and researchers alike, but they do not create a level playing field where researchers and civil society can scrutinise them to attempt to gain that trust.
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 research into online safety matters? • What topics/issues they are researching?	Confidential? – N Particularly following the removal of access to APIs by several platforms, and their tendency to remove researcher access or even accounts during collaborations that got too critical, there is very little access to information about online safety and its surrounding areas. As such, researchers (like users and the press) are left to reverse engineer platform processes by examining ‘glitches,’ ‘errors’ or moderation trends to understand how platforms work. This is always done under the threat of potentially misusing platform data, although recent case law has protected researchers in feared cases of alleged criminal access to information: https://globalfreedomofexpression.columbia.edu/cases/sandvig-v-barr/. Still, if researchers alert platforms to the presence of harmful subcultures or content, platforms will remove them for PR reasons (and, of course, for user safety). This however means that researchers could not observe those subcultures and content to understand them and produce work on them, and that if they want to do so, they need to negotiate their visibility with their own safety through covert observations. Recruiting members of subcultures, or users at the margins such as sex workers and trans people (who are disproportionately affected by online abuse and censorship), puts researchers themselves at risk of censorship and abuse, making recruitment via platforms potentially dangerous and practically challenging, particularly when recruiting around frowned upon topics such as sexuality. I discuss all the challenges surrounding this in my paper on “Researching under the platform gaze,” where I argue that the way platform governance currently works harms and impedes the

Question	Your response
	production of knowledge: https://journals.sagepub.com/doi/10.1177/29768624241283912#bibr50-29768624241283912 .
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? – N • Case-by-case platform governance decision information to understand the inner working of platform processes • Information on the workforce that makes technologies to understand their inherent biases • The specific triggers resulting in content moderation decisions to interrogate the efficiency of the running of platforms • When something is moderated by a human and when by an algorithm, once again to understand platform processes • The reasons driving policy and enforcement change to understand the relationships linking platforms, governments, laws and scandals • Clear information about content moderators’ work to understand how their working conditions affect them and influence their decisions.
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

Question	Your response
Question 1d: What technologies are typically used by providers of online services to facilitate existing information access?	Confidential? – Y / N
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
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? What are the challenges relating to the scale and complexity of the information involved?	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? • What are the security protocols required to protect information from misuse?	Confidential? – N Answering just one question. 1) Releasing information about actions taken on victims of harm or on marginalised users can 'out' them to audiences they did not agree to be faced with if the information is published, and can result in the hosting and sharing of harmful material.

Question	Your response
To what extent do you view security as a governance issue compared to a technical infrastructure issue?	
Question 2d: What are the information quality challenges relating to online services sharing information with independent researchers?	Confidential? – Y / N
Question 2e: What are the financial costs to online services relating to online services sharing information with independent researchers?	Confidential? – N Sharing all this information will no doubt be costly to platforms, because it will reveal their operations potentially triggering requests to improve them, and therefore more investment as well as potential image and therefore financial damage. But revealing this information is the only way to allow the production of knowledge and move away from a solely platform-centred platform governance without social, academic and state oversight.
Question 2f: What are the financial costs to researcher trying to make use of information shared by online services?	Confidential? – N API purchases are incredibly costly, and potential lawsuit for information misuse can be even more so. However, the emotional and intensive free labour that is part of jumping through hoops to attempt to access governance information in an environment that doesn't facilitate scrutiny, reverse engineering decisions that rely on 'folk theories' by users and therefore face heightened scrutiny in peer review, is an added cost.

Question	Your response
Question 3: How might greater access to information for the purpose of research into online safety issues be achieved?	Confidential? – N I answered
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	Confidential? – Y / N

Question	Your response
risks of those models, and how might they apply to the online services context?	
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? – N Knowledge exchange sessions led by platforms where workers show how products are made, how policies are applied and enforced, or which kind of content is moderated and how can be greatly beneficial for researchers to understand platform governance in action.
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? – N Platforms will not want to divulge moderation information, which they sometimes see as a business secret, and they will for sure want to avoid sharing information about black boxes outside their inner circle for issues of PR and copyright. But knowing the process from the inside is the only way of understanding how platforms function apart from reverse-engineering their governance - although it's of course to be debated whether even by engaging in these activities platforms will be truthful and transparent.
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? – N Third-party organisations can build pervasive transparency into platform processes, administering fines and penalties for lacking transparency, and facilitating researcher access through knowledge exchange programs. The DSA is already a great example of how much information platforms can be compelled to release - see here: https://www.fastcompany.com/91203631/the-eu-has-helped-us-understand-exactly-how-horny-people-are . However, third-party orgs can also facilitate the publishing of better transparency reports, for platforms to share more than

Question	Your response
	mere figures of how much content they moderated and how, and provide actual examples of this moderation towards public understanding. Further, third-party organisations can ensure that users have access to immediate, easy-to-gather information about what happens to their content and data, compelling platforms to improve take-down notifications, provide case records of appeals and moderated content, and in general empower users about understanding what happens to their content. See user-centred design solutions created with the Web Foundation here: https://digitalcitizens.uk/blog/platform_governance_inequalities/
Question 3f: What could these third-party models look like, and what are some of the benefits and challenges associated with this approach?	Confidential? – N Society as a whole will benefit from more platform transparency: users will make more informed decisions about what they can post, and they will hopefully have access to fairer moderation as a result due to the sheer scrutiny platform processes will be under. But until platforms are compelled to release information or face penalties, they will not want to invest in improving transparency and in the reputational damage and costly governance changes this may take.
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?	Confidential? – N • Several examples of effective and ineffective automated and human moderation of specific content to understand how platforms enforce their policies, where automation falls short, and where policy can be improved • Several examples of the functioning of appeal processes, which should be a staple of any democratic system but that are, at present, ‘dysfunctional’ according to my participants: https://www.tandfonline.com/doi/full/10.1080/1369118X.2024.2396621#d1e562 • Access to platform moderators for interviews to understand their roles and working conditions • Access to platform policy makers to understand how policies are written and implemented, and how these respond to legislation

Question	Your response
	• Access to product developers within companies, to understand how they create technologies and affordances that implement policies.