BILETA's Response to Ofcom's Call for Evidence on Researcher Access to Regulated Online Services Information; submitted on 17th January 2025.

This response is submitted on behalf of BILETA (British and Irish Law, Education and Technology Association).

Established in April 1986, BILETA aims to advance and disseminate high-quality research and knowledge on technology law and policy to diverse audiences, including organisations, governments, professionals, students, and the public. BILETA also champions the use of and research into technology across all levels of education.

This response has been prepared by Reader Dr Edina Harbinja and Principal Lecturer Dr Felipe Romero-Moreno.

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** Independent researchers face significant hurdles in accessing data from regulated online service providers for online safety research. While some avenues exist, they are often limited and inconsistent. Platforms publish transparency reports, such as those from Google (https://transparencyreport.google.com/), Meta (https://transparency.fb.com/), and X (formerly known as Twitter) (https://transparency.twitter.com/), detailing content removals and government requests, but these often lack granular data. Some platforms offer APIs, but access can be restricted, as evidenced by recent changes to Twitter's API - see e.g., https://arxiv.org/abs/2404.07340#:~:text=Twitter%20data%20studies%20have%20increased,by%2013%25%20compared%20to%202022. Academic partnerships offer another route, but these are limited in scope - see e.g., https://link.springer.com/article/10.1007/s11301-023-

	00349-1. In some jurisdictions, Freedom of Information requests can be used, but this process can be lengthy - see e.g.,https://www.muckrock.com/news/archives/2023/oct/25/foiaonline-backup-pogo-muckrock-archive/. Emerging initiatives like data trusts, discussed by organizations like the Open Data Institute (https://theodi.org/) and the Global Partnership on AI (https://gpai.ai/), aim to facilitate responsible data sharing, but are still developing. These avenues are hampered by a lack of standardization, privacy concerns, commercial sensitivity, and varying legal frameworks. Recent legislation like the UK's Online Safety Act 2023 and the EU's Digital Services Act aim to improve data access, but their effectiveness depends on implementation. For example, the Online Safety Act 2023 acknowledges the importance of independent research into online safety but stops short of directly granting researchers access to data from regulated online service providers. While Section 100 empowers OFCOM to compel information disclosure from providers "for the purpose of exercising, or deciding whether to exercise, any of their online safety functions" (Section 100(1)), this power is discretionary and doesn't guarantee researcher access. Furthermore, the focus on "information demonstrating in real time the operation of systems, processes or features" (Section 100(3)(a)) may not serve broader research needs. However, the Act offers some indirect benefits. Schedule 2 mandates providers to publish annual reports detailing their content moderation practices, potentially providing some insights, though the level of detail is yet to be determined. Crucially, Section 162 requires OFCOM to produce a report examining current researcher access to information and exploring the legal and other issues surrounding information sharing for research, which could inform future data access initiatives. Therefore, while the Act doesn't provide direct data access for independent researchers, it recognizes the importance of research and sets the stage for potential future improvements, contingent on the findings of the report mandated by Section 162 and subsequent regulations. In conclusion, while some access points exist,

	researchers face substantial challenges due to these limitations.
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** The current level of information access, though restricted, allows for certain types of online safety research. Researchers can analyse publicly available information like terms of service, community guidelines, and transparency reports to understand platform approaches to online safety, including content moderation and reporting mechanisms. Aggregate data from these reports allows for the identification of trends in online harms, though granular data limitations restrict in-depth analysis. Where accessible, APIs enable studies on online behaviours like misinformation spread and hate speech prevalence, but access is often restricted. For example, researchers could use X (previously more open) Twitter APIs to track misinformation via hashtags and keywords, analyse content sharing patterns, and identify bot activity - see e.g., https://www.cambridge.org/core/services/aop-cambridge-core/content/view/6B31D18C5E2F9B8F9C0301BFB05F1C27/S0730938421000198a.pdf/div-class-title-twitter-as-research-data-div.pdf. Similarly, hate speech could be studied via NLP on API-collected text , mapping hate networks , and analysing content moderation impact - see e.g., https://www.sciencedirect.com/science/article/pii/S0925231223003557. However, rate limits restrict data collection, and data availability can be limited, and API access policies can change, hindering long-term studies and replication - see e.g., https://www.nature.com/articles/s41562-023-01750-2. For instance, COVID-19 misinformation studies were impacted by Twitter's 2023 API changes, and online harassment research is hampered by limited access to

	interaction data - see e.g., .https://www.sciencedirect.com/science/article/pii/S2468696420300458. These restrictions limit research scope and hinder understanding of online harms. Qualitative research, using public data like online discussions or interviews with those who have experienced online harms, provides valuable insights into lived experiences - see e.g., information available and often published on https://www.eff.org/ and https://www.accessnow.org/.This research is conducted by various independent researchers, including academics studying cyberbullying, harassment, misinformation, and extremism; non-profit organizations focused on digital rights and online safety; investigative journalists holding platforms accountable; and independent data scientists analysing available data. These researchers explore topics such as the prevalence and impact of online harms, the effectiveness of platform policies, the spread of misinformation and disinformation, online radicalization and extremism, and the impact of new technologies like AI on online safety - see e.g., https://www.tandfonline.com/doi/full/10.1080/13600869.2024.2324540 and https://committees.parliament.uk/writtenevidence/39317/pdf/. Furthermore, for instance, Ofcom's Online Safety Research Agenda https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/using-research-to-guide-our-online-safety-work/ exemplifies the types of research being conducted. However, it is crucial to acknowledge that limited data access significantly constrains the scope and depth of research, often forcing reliance on indirect methods or limited datasets, potentially impacting the reliability and generalizability of findings.
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,	**Confidential? – N** Independent researchers are often unable to access crucial information relevant to online safety studies, hindering comprehensive understanding of online harms. Granular user data, encompassing detailed individual user information like demographics, interests,

what are they and what kind of research would they facilitate?	and network connections, is often unavailable, limiting research into targeted harassment, echo chambers, and the impact of personalization algorithms, as highlighted by the ODI's discussion on the need for such data (https://theodi.org/news-and-events/blog/data-access-and-the-online-safety-bill/). Data on content moderation decisions, detailing platform moderation processes, criteria, and algorithms, is also frequently inaccessible, preventing effective evaluation of moderation policies and the identification of potential biases, a point emphasized by the Centre for Countering Digital Hate (https://counterhate.com/blog/data-access-for-researchers-the-key-to-online-safety-regulation/). Access to data on platform algorithms themselves, crucial for understanding their role in spreading misinformation and hate speech, is also restricted, a concern raised by ISD in their discussion of algorithmic transparency (https://www.isdglobal.org/digital_dispatches/why-the-online-safety-bill-must-include-provisions-for-data-access/). Finally, data from internal platform research, which could allow for independent verification of platform claims and the uncovering of hidden harms, is typically unavailable. This lack of access significantly constrains research, hindering the development of evidence-based solutions to online safety issues.

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? – N** Several data governance models facilitate researcher access to online service information, ranging from open access to highly controlled use-case accredited systems. Open-access models include publicly accessible information libraries and databases, such as transparency reports published by Google (https://transparencyreport.google.com/), Meta (https://transparency.fb.com/), and X (https://transparency.twitter.com/), which provide data on government requests and content removals but often lack granular detail. The Internet Archive https://archive.org/about/ also offers a form of open access by archiving web pages and other digital content. Vetted access models include API access with restrictions, where platforms provide data access subject to rate limits, data limitations, and application processes, often requiring researchers to demonstrate credentials and ethical considerations (see e.g., Meta Research https://research.facebook.com/). Data collaboratives or partnerships represent another vetted access model, involving formal collaborations between platforms and research institutions with data shared under specific agreements, though these are often limited in scope - see e.g., https://www3.weforum.org/docs/WEF_Data_Collaboration_for_the_Common_Good.pdf. Use-case accredited access models include data trusts, relatively new structures holding data on behalf of beneficiaries (e.g., researchers) with defined access and usage rules. The Open Data Institute (ODI) is a key proponent of data trusts (https://theodi.org/), which, while not widely used for online safety research in the online services industry, represent a potential future model. Secure data enclaves or clean rooms, secure computing environments allowing researchers to analyse sensitive data without direct downloading, are also emerging as a use-case accredited model. It's important to note that these models can be combined, with platforms potentially using transparency reports (open access), restricted APIs (vetted access), and data collaboratives (use-case accredited access). The evolving regulatory

	landscape, particularly with legislation like the EU's Digital Services Act (DSA), influences these models, pushing for increased transparency and researcher access. For instance, the DSA addresses transparency and researcher access primarily through obligations on Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs). Article 40(12) is the most direct provision for researcher access, mandating data sharing with vetted researchers for studying systemic online risks. Other articles, such as those on risk assessment, mitigation measures, independent audits, and transparency reporting, contribute to a more transparent online environment and indirectly support research efforts.
Question 1d: What technologies are typically used by providers of online services to facilitate existing information access?	**Confidential? – N** Providers of online services employ various technologies to facilitate information access for researchers, each offering different levels of control and data granularity. APIs (Application Programming Interfaces) are a key tool, allowing researchers to programmatically retrieve data. For example, X's API (in its prior, more open form) enabled data collection on tweets and user information (see e.g., https://www.tandfonline.com/doi/full/10.1080/19312458.2023.2181319), though access has since become more restricted. Platforms like Reddit and YouTube also offer APIs with varying access levels (https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethics-Using-internet-and-Social-media-data-v8.pdf). Some platforms provide data dumps or bulk downloads of datasets, often anonymized or aggregated, enabling offline analysis, though these are often one-off rather than ongoing access mechanisms (see e.g., https://developers.google.com/freebase). Secure data enclaves or clean rooms offer a more controlled approach, providing secure computing environments where researchers can analyse sensitive data without

	direct downloading, though public examples are limited due to data sensitivity (see e.g., https://dssc.eu/space/BBE/178422141/Data+Sharing+Governance). Transparency reports, published by companies like Google (https://transparencyreport.google.com/), Meta (https://transparency.fb.com/), and X (formerly Twitter) (https://transparency.twitter.com/), offer aggregate data on content removals and government requests, serving as primary sources (see e.g.,https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/looking-ahead-to-online-regulation-transparency-reporting/). Finally, web scraping, including genAI scraping for dataset training, while not platform-provided, involves automated data collection from publicly accessible websites, raising ethical considerations regarding privacy and terms of service (see e.g., https://onlinelibrary.wiley.com/doi/10.1111/jwip.12331?af=R). The availability and accessibility of these technologies differ significantly between platforms and can change over time, influenced by the evolving regulatory landscape, such as the UK OSA and the EU DSA.
Question 1e: Have services and/or researchers made use of privacy-enhancing technologies to enable access?	**Confidential? – N** Yes, both online service providers and researchers are increasingly using privacy-enhancing technologies (PETs) to facilitate data access while protecting privacy. Differential privacy, a technique that adds calibrated noise to datasets to allow analysis of trends without revealing individual information, is one such example, with Apple using it for aggregate user behaviour data collection (see e.g., https://link.springer.com/chapter/10.1007/978-3-540-79228-4_1). Federated learning, which trains machine learning models on decentralized data without direct access to raw information, is another example, with Google employing it to improve keyboard predictions on Android devices https://research.google/blog/distributed-differential-privacy-for-federated-learning/ (see also e.g.,

	https://www.edps.europa.eu/press-publications/publications/techsonar/federated-learning_en). Homomorphic encryption, allowing computations on encrypted data without decryption, is being explored for sensitive data analysis in areas like healthcare (see e.g., https://crypto.stanford.edu/craig/craig-thesis.pdf?PHPSESSID=af675c6c533141591dc910d383262de5), while secure multi-party computation (MPC) enables joint data analysis between multiple parties without revealing private inputs (see e.g., https://research.cs.wisc.edu/areas/sec/yao1982-ocr.pdf). These PETs enable research by allowing access to sensitive data, facilitating collaboration, and building trust through privacy protection. For instance, the Royal Society, discusses the potential of PETs for safe data use, including in online safety research in some of its reports (see https://royalsociety.org/news-resources/projects/privacy-enhancing-technologies/). The ICO's guidance on anonymization, pseudonymization, and PETs (https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf) also highlights their role in complying with data protection principles. While challenges related to implementation, scalability, and the trade-offs between privacy and data utility remain, PETs represent a significant advancement towards responsible and privacy-preserving data access for online safety research.

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? - N** Researchers face significant challenges in accessing data from online service providers for studying online safety-related issues. While platforms like Google, Meta, and X publish transparency reports as noted above,

	these often lack granular data, limiting detailed analysis of content moderation or algorithmic effects. Changes to access mechanisms, such as X's recent API restrictions, further hinder long-term and reproducible studies. Although some platforms offer academic partnerships or vetted APIs, these are selective and often inaccessible to independent researchers, journalists, or non-profits. For example, the reliance on restrictive APIs limited studies on COVID-19 misinformation and hate speech analysis. Privacy and data protection concerns, and regulatory constraints exacerbate these challenges. Data sharing is complicated by legal frameworks like GDPR, which require robust safeguards against re-identification risks. Emerging technologies like differential privacy and federated learning offer solutions but are not widely adopted. Additionally, commercial sensitivities deter platforms from sharing data that could expose proprietary algorithms or moderation practices. This dynamic creates significant inequities, as demonstrated by Ofcom's findings that much of the research currently depends on publicly available data, interviews, or limited datasets from organizations like the Internet Archive. While legislation like the UK's Online Safety Act and the EU's Digital Services Act recognise the need for research, their implementation remains complex, leaving researchers reliant on inconsistent and fragmented access models.
Question 2a: What are the legal challenges/risks to sharing information from online services with independent researchers?	**Confidential? - N** Sharing information from online services with independent researchers poses several legal challenges and risks. A primary concern is compliance with data protection laws such as the (UK) GDPR and DPA, which regulate the processing of personal data, including anonymised datasets that could potentially be re-identified. Platforms must ensure that shared data meets stringent privacy requirements, balancing transparency with the protection of users' rights. Additionally, intellectual property and commercial confidentiality laws may limit the ability of platforms to

	disclose proprietary algorithms or sensitive operational details, particularly if such disclosures could undermine their competitive position. Liability risks also arise if shared data is misused or leads to unintended harms. For instance, platforms could face legal action if researchers inadvertently publish findings that expose individual users or sensitive information. Jurisdictional differences add complexity, as platforms operating globally must navigate varying national laws on data sharing, privacy, and research ethics. Recent regulatory frameworks, such as the UK's Online Safety Act and the EU's Digital Services Act, attempt to address these issues by mandating certain data-sharing mechanisms, but they also introduce ambiguities around implementation and enforcement, leaving platforms cautious about engaging in broader data-sharing initiatives.
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? – N** Sharing information from online services with independent researchers involves significant technical hurdles. Platforms often use proprietary data formats, making it difficult to standardise and share data in a manner that is interoperable across different research tools and institutions. APIs, a common mechanism for data access, are frequently restricted in terms of rate limits, data granularity, and availability, limiting their utility for large-scale or nuanced research. Additionally, ensuring the security of shared data presents a challenge, as platforms must prevent unauthorised access or data breaches while enabling researchers to access sensitive information. Privacy-enhancing technologies like differential privacy, federated learning, and secure multi-party computation can help balance access and user privacy, but these methods require substantial computational resources and expertise, which may not be universally available. Implementing such solutions at scale often involves high costs and infrastructure demands, particularly for smaller platforms.

	The scale and complexity of data generated by online services further complicate information sharing. Platforms handle vast quantities of user-generated content, behavioural data, and interaction metadata, making it challenging to extract, anonymise, and share relevant datasets without compromising user privacy or data integrity. For instance, datasets from platforms like X or Meta can include billions of interactions, requiring robust mechanisms for filtering and processing data. Moreover, the dynamic nature of online environments means that data is constantly changing. This raises challenges for researchers who require stable datasets for longitudinal studies or reproducible results. Complexities in platform algorithms, which often determine the spread and visibility of content, add another layer of difficulty. Without transparency into these algorithms, researchers may struggle to interpret data accurately or understand the context of the information provided. Together, these challenges demand advanced technical solutions, clear regulatory guidelines, and collaborative efforts between platforms and researchers.
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? To what extent do you view security as a governance issue compared to a technical infrastructure issue?	**Confidential? – N** Sharing information from online services with independent researchers involves significant security risks, particularly when handling sensitive or user-related data. The potential for re-identification of anonymised data remains a persistent challenge, as advanced data analytics techniques can cross-reference datasets to uncover personal identities. Additionally, platforms must safeguard against unauthorised access, data breaches, and malicious actors who could exploit vulnerabilities in data-sharing mechanisms. Sensitive information, such as internal moderation practices, platform algorithms, or detailed user interactions, requires heightened protection to prevent its misuse for competitive, political, or harmful purposes. Security protocols must ensure the integrity and confidentiality of shared data. This includes employing

	encryption for data in transit and at rest, implementing secure access controls (e.g., multi-factor authentication), and monitoring data usage to detect any anomalies or unauthorized activities. Privacy-enhancing technologies like differential privacy and secure multi-party computation can further mitigate risks by enabling analysis without exposing raw data. However, these measures demand substantial technical expertise and infrastructure, which smaller platforms may lack. Security in data sharing is both a governance and a technical infrastructure issue. Governance concerns include establishing clear policies on who can access data, under what conditions, and for what purposes. Transparent governance frameworks can reduce the risk of misuse by requiring researchers to comply with ethical guidelines, data use agreements, and accountability mechanisms. For example, platforms may adopt a "Five Safes" approach, evaluating the safety of data projects, people, settings, data, and outputs to ensure responsible sharing. https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/ From a technical perspective, robust infrastructure is essential for implementing these governance measures. However, even with advanced technologies, weak governance structures can lead to misuse or non-compliance with security protocols. Conversely, strong governance frameworks without adequate technical capabilities may fail to enforce safeguards effectively. Therefore, security in data sharing requires an integrated approach that combines governance policies with scalable technical solutions, fostering trust and minimizing risks in sharing sensitive online service information.

Question 2d: What are the information quality challenges relating to online services sharing information with independent researchers?	**Confidential? – N** Information quality challenges significantly impact the sharing of data from online services with independent researchers. Platforms often provide aggregated or incomplete datasets that lack the granularity needed for in-depth research, limiting the ability to draw meaningful insights. Data inconsistencies, such as missing metadata or unclear documentation, further hinder usability and replicability. Additionally, biases in the data—stemming from algorithms, user behaviour, or moderation practices—can skew findings and reduce generalisability. The dynamic nature of online platforms, where data evolves rapidly, also complicates longitudinal studies and creates challenges in maintaining dataset accuracy and relevance. Without standardised formats, clear contextual information, and mechanisms to address biases, shared data may fall short of supporting robust, reliable research outcomes.
Question 2e: What are the financial costs to online services relating to online services sharing information with independent researchers?	**Confidential? – N** Sharing information with independent researchers imposes financial costs on online services. Developing and maintaining secure data-sharing infrastructure, such as APIs or secure data enclaves, requires substantial investment in technology and personnel. Privacy-preserving measures like differential privacy, data anonymisation, and federated learning add further costs, particularly as these techniques demand advanced expertise and computational resources. Platforms must also allocate resources for compliance with legal and regulatory requirements, including vetting researchers, monitoring data usage, and conducting audits to ensure ethical use. Smaller platforms, in particular, face challenges in balancing these costs with their operational budgets, making it difficult to provide meaningful access without compromising financial stability. These expenses, combined with the potential legal and reputational risks of data misuse, illustrate the

	financial burden n these smaller platforms of enabling independent research.
Question 2f: What are the financial costs to researcher trying to make use of information shared by online services?	**Confidential? – N** Researchers face considerable financial costs when utilizing information shared by online services. Access to data often requires payment for API usage, which can escalate with high data volume or long-term research needs. Advanced technical infrastructure is frequently necessary to process and analyse large datasets, including high-performance computing resources and specialised software. Additionally, researchers may need to invest in training or hiring skilled staff to navigate complex data-sharing mechanisms, privacy-preserving technologies, or platform-specific formats. Legal and ethical compliance, such as securing institutional ethical approvals or meeting data protection requirements, can also add administrative costs. For smaller research teams or independent scholars with limited funding, these financial burdens can limit their ability to engage with or fully utilize available data.

Question	Your response
Question 3: How might greater access to information for the purpose of research into online safety issues be achieved?	**Confidential? – N** Achieving greater access to information for online safety research requires a multi-faceted approach combining technical and regulatory solutions. Enhanced API access is crucial, including providing more granular data (while respecting privacy through techniques like

	anonymization), increasing rate limits to enable larger datasets, and establishing stable and predictable API access policies, as highlighted by Bruns (see e.g., https://www.nature.com/articles/s41562-023-01750-2). Establishing standardized data sharing frameworks with common data formats and metrics and interoperable APIs would facilitate cross-platform data comparison and meta-analyses, a point advocated for by the ODI (https://theodi.org/). Independent oversight and auditing, through an independent data access body and mandatory data sharing with vetted researchers under specific conditions, as suggested by the CCDH (https://counterhate.com/blog/data-access-for-researchers-the-key-to-online-safety-regulation/), could further ensure appropriate access. Promoting wider adoption and continued development of privacy-enhancing technologies (PETs) like differential privacy, federated learning, and secure multi-party computation can enable data sharing while mitigating privacy risks, as discussed in the Royal Society report (https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/from-privacy-to-partnership.pdf). Finally, supporting the development of data trusts and exploring the potential of data cooperatives, as also discussed by the ODI (https://theodi.org/), can provide legal, ethical, and user-centric frameworks for responsible data sharing. Balancing research needs with user privacy and platform interests remains a key challenge in implementing these solutions.
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? – N** Several models and frameworks outside the online services industry offer valuable insights for granting researchers access to sensitive information while balancing research utility with privacy. Trusted Research Environments (TREs) or secure data enclaves provide secure computing environments where researchers analyse sensitive data without it leaving the controlled environment, like the UK Biobank (https://www.ukbiobank.ac.uk/) which offers access to anonymized health data, or the UK's Office for National

	Statistics (ONS) Secure Research Service (https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice) which provides access to government data. These offer enhanced security and privacy but can be expensive and limit analysis types. Furthermore, for instance in Scotland, the Scottish Public Benefit and Privacy Panel (PBPP) is a key body responsible for overseeing access to government data for research, ensuring public benefit, privacy, and ethical data use. This includes data relevant to online safety research, such as health and social care data that might reveal online harms like cyberbullying or the spread of misinformation. Researchers wanting to use this data must apply to the PBPP, demonstrating their project's value and how they will protect individual privacy. https://www.gov.scot/binaries/content/documents/govscot/publications/factsheet/2020/12/scottish-government-statistics-request-our-data/documents/statistics-public-benefit-and-privacy-panel-terms-of-reference/statistics-public-benefit-and-privacy-panel-terms-of-reference/govscot%3Adocument/terms%2Bof%2Breference.pdf. Applying this to online services could allow researchers access to user data within a platform-controlled environment for sensitive research. Data trusts, as championed by the Open Data Institute (ODI) (https://theodi.org/), are legal structures providing independent data stewardship with defined access rules, balancing stakeholder interests and promoting transparency, though they involve complex setup and governance. Such trusts could govern access to platform data for online safety research. Synthetic data, artificially generated data mimicking real data's statistical properties, offers another approach used in healthcare and finance, allowing analysis without compromising privacy, though it may not perfectly capture real data's complexity - see e.g.,https://acrobat.adobe.com/id/urn:aaid:sc:EU:7c09413f-af5e-4fc4-abd0-ddfa60b3cf4d. Differential privacy, used by Apple and the US Census Bureau, adds noise to datasets to protect individual privacy while preserving aggregate trends - see

	https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf and https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html. This could be used to release aggregated statistics on online harms. The "Five Safes" framework (https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/), emphasizing safe projects, people, data, settings, and outputs, provides a useful lens for evaluating data sharing. By adapting these models, the online services industry can create stronger, privacy-preserving frameworks for online safety research.
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? – N** Several existing models within the online services industry could be more broadly applied to enhance research access to information, although each presents its own advantages and disadvantages. Platform-specific research programs or portals, such as Meta Research (https://research.facebook.com/) offering various research initiatives and tools, or Google Scholar (https://scholar.google.com/) indexing scholarly literature including research on online platforms, can facilitate targeted research and provide access to unique datasets. However, access is often limited to specific areas or institutions, potentially raising concerns about platform influence. Data challenges and competitions, like the Netflix Prize https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data or Kaggle competitions https://www.kaggle.com/competitions often using platform data, can stimulate innovation and generate valuable insights but often involve data limited to a specific timeframe or purpose, potentially discouraging broader research or collaboration. Standardized API access with varying tiers, offering different access levels based on researcher credentials and data sensitivity, provides a more structured and transparent approach, balancing research needs with privacy but requiring significant development effort and clear criteria for access tiers - see

	e.g.,https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006; https://www.amazon.co.uk/Privacy-Context-Technology-Policy-Integrity/dp/0804752370 and .https://royalsociety.org/news-resources/projects/privacy-enhancing-technologies/. Industry-wide data sharing initiatives, such as the Data Transfer Project (https://dtinit.org/) aiming to facilitate data portability, could provide a more comprehensive view across platforms and reduce individual platform burden, but require significant inter-platform cooperation and raise complex data governance and competition law issues. The success of these models hinges on addressing key challenges like protecting user privacy through robust anonymization, ensuring research independence by preventing platform influence, and promoting transparency and accountability through clear guidelines and oversight. Learning from these existing models and addressing these challenges can facilitate a more open and collaborative approach to data sharing for online safety research.
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** Several promising models, not yet widely adopted in the online services industry, could enhance research access to information while addressing evolving online safety challenges. Combining differential privacy with synthetic data offers a robust approach: differential privacy generates summary statistics from real data, which then informs the creation of synthetic data, allowing researchers to work with realistic data without compromising individual privacy (see e.g.,https://www.nist.gov/blogs/cybersecurity-insights/differentially-private-synthetic-data). Distributed data analysis platforms, employing techniques like federated learning or secure multi-party computation, would enable researchers to analyse data held by multiple platforms without direct data transfer, facilitating large-scale, cross-platform research while preserving privacy and addressing competitive concerns (see e.g., https://arxiv.org/abs/2404.12273). Enhanced data clean rooms, building upon existing concepts, could offer more sophisticated analytical tools and incorporate

	PETs for greater privacy within secure, controlled research environments (see e.g., https://www.amazon.co.uk/Privacy-Age-Big-Data-Recognizing/dp/1442225459). Finally, "privacy-preserving record linkage" techniques could allow researchers to link records across different datasets without revealing identifying information, enabling crucial cross-platform research in a multi-platform online environment (see e.g., https://www.amazon.co.uk/Data-Matching-Techniques-Data-Centric-Applications/dp/3642430015). Implementing these models requires collaboration between researchers, platforms, policymakers, and privacy experts to balance research needs with robust privacy safeguards.

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** Adopting advanced data-sharing frameworks for online safety research offers several advantages and disadvantages across financial, legal, security, technical, and feasibility dimensions. On the positive side, approaches like Trusted Research Environments, data trusts, and privacy-preserving technologies enable secure access to sensitive data while protecting user privacy and ensuring compliance with legal standards such as GDPR. These frameworks foster trust between platforms and researchers, promoting transparency and collaboration. However, they also entail significant financial costs, including the development and maintenance of secure infrastructures and the expertise required for their implementation. Legally, such approaches introduce complexities in balancing regulatory compliance with intellectual property and confidentiality concerns. From a technical perspective, the infrastructure required for secure data sharing can be resource-intensive, and achieving interoperability across platforms remains a challenge. Furthermore, the feasibility of these models depends on the willingness of platforms to adopt them and collaborate openly, which may be hindered by concerns about exposing proprietary practices or operational risks. While these approaches hold great potential for improving data

	sharing, their success relies on addressing these multifaceted challenges effectively.
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 organizations, including regulatory bodies, civil society groups (some of which we mention above), and public sector organisations, can play a crucial role in facilitating researcher access to online safety information by acting as intermediaries, advocates, and enforcers. Regulatory bodies, such as Ofcom or the European Commission under the Digital Services Act, can mandate platforms to establish standardised data-sharing mechanisms, define access criteria, and ensure compliance with privacy and security regulations. They can also oversee the vetting of researchers, monitor data usage, and enforce penalties for non-compliance, fostering a more transparent and accountable framework. Civil society organisations and public sector entities can advocate for equitable access to data and highlight underexplored issues in online safety research, ensuring diverse perspectives are considered. They can also facilitate collaboration by creating partnerships between platforms and researchers or managing data trusts that provide secure and independent stewardship of sensitive information. Additionally, these organisations can contribute to capacity building by providing technical and financial support to researchers, particularly those from smaller institutions or underfunded areas. By coordinating efforts across stakeholders, these third-party organisations can help balance the competing interests of platforms, researchers, and users, ensuring that data sharing advances online safety research while upholding ethical and legal standards.

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** Third-party models to facilitate researcher access to online safety information could take various forms, including Trusted Research Environments (TREs), data trusts, public-private partnerships, and independent accreditation bodies. TREs, such as those used by the UK Biobank, provide secure platforms where researchers can analyse sensitive data without removing it from controlled environments, ensuring privacy compliance and security. Data trusts, overseen by neutral entities, manage data on behalf of stakeholders, balancing transparency, accountability, and user protection. Public-private partnerships and third-party accreditation bodies can also support data sharing by fostering collaboration, certifying researchers, and establishing standardised access protocols. These models offer significant benefits by promoting transparency, equitable access, and legal and ethical standards compliance. They enhance security through robust governance frameworks and support cross-platform research by standardising data-sharing processes. However, their implementation involves challenges, including high costs, legal complexities, and potential resistance from platforms concerned about privacy risks, data misuse, or loss of control over proprietary information. Public-private partnerships, for instance, risk being influenced by commercial interests, while data trusts require intricate legal and governance structures to succeed. Despite these challenges, third-party models hold promise for bridging the gap between platforms and researchers. By fostering trust and collaboration, they can enable robust and ethical online safety research, supported by transparent governance, stakeholder

	commitment, and sufficient financial and technical resources. These models could form the foundation for a more transparent and accountable digital environment.
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** Online service providers should grant researchers access to several key categories of information to facilitate the study of online safety matters, each offering unique insights into online harms and platform behaviour: 1. **Content moderation data**: This includes information on the types of content flagged, the rationale behind content removal, and how moderation policies are applied. Access to such data enables researchers to assess the effectiveness, fairness, and potential biases in moderation practices, helping to identify areas for improvement in addressing harmful content like misinformation, hate speech, and extremism. 2. **Algorithmic data**: Details about platform algorithms, including how they recommend, rank, or promote content, are critical for understanding their role in amplifying harmful content or creating echo chambers. Studying this data allows researchers to evaluate the systemic risks posed by algorithmic curation and suggest measures to mitigate harm while preserving user engagement. 3. **User interaction and behavioural data**: Information about user interactions, such as comments, likes, shares, and network connections, can reveal patterns in the spread of harmful content, cyberbullying, or coordinated disinformation campaigns. Analysing this data helps identify behavioural trends and intervention points to promote safer online spaces. 4. **Transparency reports and compliance data**: Aggregated data on content removal, government requests, and platform policy enforcement provides a broader understanding of how platforms handle safety issues. This helps

	researchers evaluate platform accountability and compliance with legal and regulatory frameworks. 5. **Demographic data (anonymised)**: Anonymised demographic information, such as age, gender, and geographic location, helps researchers study the differential impact of online harms on various user groups. For instance, understanding how cyberbullying affects younger users or how misinformation targets specific demographics can inform tailored interventions. Providing access to these categories of information is valuable as it enables a comprehensive analysis of the factors contributing to online harms, the effectiveness of current mitigation efforts, and the development of evidence-based strategies to improve online safety. While safeguarding user privacy and addressing security concerns remain vital, structured access to these datasets empowers researchers to uncover actionable insights that benefit users, platforms, and policymakers.