

Your response

Volume 2: The causes and impacts of online harm

Ofcom's Register of Risks

Question 1:

- i) Do you have any comments on Ofcom's assessment of the causes and impacts of online harms?

The Wikimedia Foundation is the nonprofit that hosts and supports Wikipedia and other volunteer-led free knowledge projects. The Foundation strives to empower everyone, including people in the UK, to help build, self-govern, and participate in learning, culture and science. We support efforts to make the internet safer and richer. When people are harassed or feel otherwise unsafe communicating online, their ability to access, create or share knowledge is diminished. Online safety is essential to the right to participate in culture and science, and for decentralised decision-making by communities who collaborate to further the public interest. In our experience, online safety can only be achieved when adequate safeguards for privacy and freedom of expression are in place. The Foundation appreciates the opportunity to work with Ofcom and share our views about what it might take to ensure that the internet remains a place where people of all ages can safely and freely access information.

In response to this question, the Foundation values that Ofcom's approach recognizes that different kinds of services and features carry different risks for users, customers, and the public. The current proposal chiefly does this by trying to bring a platform's "commercial profile" and "business model" vaguely into the analysis. However, that proposed approach — and specific parts of guidance building on those concepts and their definitions — does not appear to incorporate one of the main causes governing platform and service provider behaviour and choices: purpose.

The purpose of a platform shapes its architecture and functionalities and, thereby, the potential causes and ultimate impacts of online harms. The purpose of the Wikimedia projects is to help everyone, everywhere, access and contribute to free knowledge. This purpose shapes the following aspects that we hope Ofcom will consider

- ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

A. Operator “capacity” restraints do not directly translate into risk; that can depend on the model (and purpose) of the platform.

Ofcom’s proposed conceptualisation of “commercial profile” and “business model” are narrow. “Commercial profile,” at present, is deemed to be “the size of the service in terms [of] capacity (i.e., revenue and/or number of employees), the stage of service maturity and rate of growth in relation to users or revenue.” The concept of “business model” then gets somewhat conflated with this concept (see paragraphs 6F.63, among others). Purpose and/or nature (e.g., nonprofit) and model (e.g., community-led) are given no attention here.

This narrow starting point then undermines the guidance that builds on those foundations. Ofcom could or should, for instance, look to deepen its consideration of these issues in paragraphs 6B.74-6B.77 (“commercial profile”), which talk about services with “limited technical and financial resources to moderate content, compared to more highly moderated mainstream services.” Ofcom’s approach seems to assume that a service operated by a platform operator with limited “capacity” (Ofcom’s definition, mentioned above, means low “revenue and/or number of employees”) inherently presents a higher risk.

In the case of the Foundation, our own limited capacity simply reflects our nature—a nonprofit principally funded by small donations—as well as the fact that the dominant source of moderating and/or governance “capacity” on the Wikimedia platforms does not come from the service provider at all: it is embodied in the community itself. This means that the Foundation plays two very important, but less commonly understood roles, neither of which are dependent on having the same “capacity”—in terms of staffing and revenues—as a large, for-profit, top-down moderated platform. Those roles are: (i) offering a moderatorial “backstop,” capable of intervening when community processes fail or on a small number of legal issues where both the law and the community has historically asked for and expected us to act directly; and, (ii) supporting community health and empowerment, that is to say, helping users feel empowered and motivated to make their own decisions and set up systems that work for them.

It should be straightforward for Ofcom’s proposed guidance to expand its foundational building blocks in this area so that it better acknowledges that possibility, rather than bake in an oversimplistic worldview where “low capacity provider” equates to risk.

B. When the relevant “capacity” comes from the user base, not the operator, there is a risk that measures promoted by Ofcom could have counterproductive safety outcomes; the guidance should warn about this.

The broader guidance in Volume 2 is heavily biassed in favour of increased interventions “from the top” by the platform operator; and an absence of such interventions plays out, time and again, as inherent evidence and/or cause of risks.

There are undoubtedly situations where that will hold true, but the guidance is missing a clear warning that there are cases where heavy top-down intervention might be counterproductive, leading to worsened safety objectives.

StackOverflow’s recent experience is clear evidence that unwise interventions by a service provider in community processes, without adequate community consensus, can lead to reduced moderation capacity.¹ Worsened safety outcomes are likely to follow—in addition to higher costs, if the operator is thus required to provide additional “backstop” resources to compensate for the reduced community moderation. Websites dependent on community health are complex ecosystems, and it is often the case that structures in which community moderators are the greater experts require the service provider to defer to those moderators to achieve the best outcomes. Consequently, it seems imperative for Ofcom’s guidance to warn about the risks of counterproductive “safety” interventions—no matter how well-intentioned—in community-led settings. An additional paragraph or two under 6B.77 could achieve this.

C. The purpose of a platform should be better recognised and embodied throughout Ofcom’s guidance, because—separate from “business model” and “commercial profile”—the pursuit of *public interest* and *user empowerment* are relevant risk minimising factors.

Besides the targeted recommendations offered above, we also note that the purpose of a platform is a critical factor for risk, and yet it receives insufficient attention in Volume 2. That manifests itself across several dimensions relevant to Ofcom’s regulatory outcomes here:

1. Governance and platform architecture: Nonprofit, public interest platforms including those hosted by the Foundation have made platform architecture choices that avoid most of the high risk vectors identified by Ofcom (like child sexual abuse material, or CSAM), and have developed robust, effective self-governance systems. These governance and platform architecture choices help limit harmful content on the Wikimedia projects, which have a far lower prevalence of such material than other large platforms (see page 155 of the OECD report “Transparency reporting on child sexual exploitation and abuse online 2023”). The same is true for other types of harmful content. The Foundation’s recently published Child Rights Impact Assessment affirms that the Wikimedia model does not present the same risks as for-profit platforms, whose business models aim to maximise advertising revenue by targeting users with highly engaging, but often inaccurate or unsuitable, content. Our principles around privacy and data minimization mean that Wikimedia projects do not collect and sell user data or use it to target users with paid advertisements. Furthermore, content on Wikimedia projects, which is freely available to anyone regardless of age, is mainly educational in nature. Community-led processes set and enforce rules and editorial standards for well-sourced educational content. As a result, both the scope of content and nature of the user experience differ significantly from what children might encounter on commercial online platforms.

2. Platform Provider Role and Control: The analysis appears to assume that all platform providers have total control of the platform, which is not true in the Foundation’s case. The Foundation partners closely with self-governing volunteer communities who exercise editorial control on the projects and have extensive technical access and input regarding the software and other technical infrastructure that operate the sites. Ofcom should

¹ See *Moderation Strike: Stack Overflow, Inc. cannot consistently ignore, mistreat, and malign its volunteers*: <https://meta.stackexchange.com/questions/389811/moderation-strike-stack-overflow-inc-cannot-consistently-ignore-mistreat-an>

consider the level of a platform provider's control as well as whom control over content is shared with by design.

3. Limited risk lens: The user-to-user (U2U) risk lens is not a good fit for public interest platforms, including Wikipedia. The purpose of such platforms is not to help visitors engage with other users on the platform, but to produce a public good for a broad audience (user-to-reader). Characterising Wikimedia projects as U2U services—even in their closest category (6A.7) information-sharing services as defined on page 20) misses their purpose and resultant low level of risk. Firstly, the U2U lens does not allow for consideration of the strong accompanying safeguards that the community-led self-governance model provides, such as the Universal Code of Conduct, the corresponding enforcement guidelines and the Universal Code of Conduct Coordinating Committee. Secondly, the perspective ignores the comparatively small number of U2U interactions on the platform, as well as the fact that the few spaces that exist for such interactions have been built by, or in close consultation with, the same self-governing communities who maintain the projects. They stringently enforce behavioural policies in these spaces that they create and adjudicate themselves to ensure that their interactions remains safe and serve their needs. One example of a self-governance mechanism that Wikimedians use to enforce policies on behaviour is the Arbitration Committee. This committee addresses misconduct by administrators and editors with access to advanced tools. Finally, the Child Rights Impact Assessment notes that with few exceptions, U2U communications on Wikimedia projects occur on publicly viewable "Talk" or discussion pages. These webpages document the interactions between editors as they collaborate, debate, and agree on how to improve Wikipedia articles or the webpages of other Wikimedia projects. "Talk" pages are public archives, since anyone can view interactions between other users. That means that the parent or guardian of a minor who is a user can view every one of their interactions as well. This considerably reduces the risk that Wikimedia projects can be used to engage in sustained harmful contact. Importantly, this independent Child Rights Impact Assessment notes that while such interactions could be initiated on these "Talk" pages, sustained harmful contact would be likely to migrate to other third-party messaging or platforms with messaging functionalities.

4. Jurisdictional lens: The framework does not appear to fully take into account that public purpose platforms benefiting the UK might not be directed specifically at UK audiences. Considering Wikipedia as an example, two reasons explain this. First, Wikipedia is organised by language, not by jurisdiction or geography; and, second, there are two key "users" of Wikipedia. These two types of users can be split into: 1) those who are passive readers of articles, doing so from anywhere in the world; and, 2) those who actively contribute content to articles, doing so from anywhere in the world. It would be contrary to the nature of the project's educational purpose to slice and dice facts, or U2U conversations, based on jurisdictional boundaries. The Foundation recognises that many for-profit providers—which are incentivised by commercial market considerations rather than strictly purpose—do tailor content and access to information based on jurisdictional considerations. For example, some commercial platforms do so when displaying geographic information like jurisdiction over the Crimean peninsula and other disputed territories. It is undesirable and would be catastrophic for Wikipedia to create an online information ecosystem in which this would be required of all information-providing platforms.

Adding purpose as an additional analytical pillar to the framework helps allocate limited regulatory resources to where they are most likely to have a positive impact in contributing to beneficial societal outcomes.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 2:

i) Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.

It seems that Ofcom proposes to expressly call out Wikipedia's editability as a risk factor for disinformation (6P.76 - 6P.80), but without any balanced mention of corollary risk mitigation and/or benefits that are inherent to peer review and editability. This is at odds with the balanced approach that Ofcom takes on other topics (e.g., 6D.19, 6D.30, 6D.35), sending a clear (albeit clearly mistaken) message about editability. Quite on the contrary, Ofcom's French counterpart (i.e., Arcom), which has been conducting a multiyear study of disinformation risks on major platforms, concluded in 2022 that "Wikipedia's model sets it apart, with information sourcing mandatory on the platform and verification at the heart of contributors' work."² Editability is a net risk-reducer because every Wikipedia edit is publicly viewable, and because malicious edits can generally be reversed by anyone at any time. This is particularly true in comparison to more traditional websites (e.g., forum, social media, video hosting, etc.), where users can only flag suspect content for more common, top-down moderation.

Besides this specific observation and insofar as the present question is concerned, we would also like to offer the following observations about Volume 2 more generally:

We believe that Ofcom's framework and its interpretation are much more useful for assessing risk factors and harms in the context of for-profit services and business models that typically depend upon targeted advertising or subscriptions, since these typically build on centralised content moderation by the platform provider. For the same reason, Ofcom's framework and its interpretation are less useful for nonprofit, open access, public interest projects that do not primarily feature centralised mechanisms for content moderation.

Ofcom's interpretation of the links between risk factors and types of illegal harm ignores the existence of public interest and nonprofit platforms, as well as those that have a combination of characteristics of multiple service types.

Ofcom has identified certain characteristics as "risk factors" with some link to illegal harms, including: service type; userbase; functionalities and recommender systems; and business model and commercial profile.

Wikipedia is not of a commercial nature

² See *Lutte contre la manipulation de l'information sur les plateformes en ligne (Bilan 2021)*: <https://www.arcom.fr/nos-ressources/etudes-et-donnees/mediatheque/lutte-contre-la-manipulation-de-linformation-sur-les-plateformes-en-ligne-bilan-2021> (in French)

First, Wikipedia is not of a commercial nature. The Foundation is a nonprofit organisation. Our primary source of revenue comes from individual donations that readers of Wikipedia make around the world. The average single donation is about £12.

Conversely, the annual revenue of the largest for-profit platform providers for 2022 was in many cases immeasurably larger: for Alphabet Inc., it was over US\$282 billion; that of Meta Platforms, Inc., over US\$116 billion; ByteDance Ltd., over US\$85 billion; LinkedIn Corporation, over US\$13 billion; Snap Inc., over US\$4.6 billion; Pinterest Inc., over US\$2.8 billion; and X Corp., over US\$4 billion, 90% of which came from advertising. We, however, are a nonprofit and have far less money available. Wikipedia is the only major online platform supported by a nonprofit organisation.

Wikimedia's revenue model is very different from those of the for-profit U2U and search platforms, which rely heavily on advertising, licensing, and subscriptions. Because there is no profit incentive behind Wikipedia and other Wikimedia platforms, there are no recommender systems or targeted advertising practices deployed that might amplify harmful content. Whistleblowers like Frances Haugen have clearly exposed how the algorithms driving profits for ad placements are at the root of the harmful content exposure that this Act is trying to address. Regulation that recognizes the economic factors that drive platform design is the only way towards a safer Internet. Community-led systems like ours allow nonprofits to run a safe, global website with a smaller staff. The decentralised nature of this work makes it possible to operate this independent source of free and open knowledge for the world with many fewer instances of harmful content, and while using far fewer resources than profit-oriented platforms.

Capacity

In terms of capacity and resources, the Wikimedia Foundation has only about 700 staff in total. This is smaller by orders of magnitude than other prominent online platforms. For example: Meta has over 66,000 employees (15,000 human moderators in the United States alone); LinkedIn has over 19,000; ByteDance has over 150,000 total employees (of which 7,000 are dedicated to TikTok); Snap Inc. has over 5,000; Pinterest has over 3,900; and, X Corp. has approximately 1,500.

At this time, the Foundation's legal department has only two, soon to be three, staff members located in the UK. Each of these individuals has a wide range of job functions, covering not only compliance and government relations, but other legal, policy, and community relations tasks as well.

Maturity

Wikipedia launched in 2001. The Foundation was founded in 2003. Wikipedia is funded by donations and has never been supported by advertising. Wikipedia's community-driven content moderation structure has been in-place for over two decades, and is the cornerstone of the encyclopaedia's thoroughness, reliability, and timeliness.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Volume 3: How should services assess the risk of online harms?

Governance and accountability

Question 3:

- i) Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice?

We wish to reiterate concerns that we have expressed in previous engagements around these principles, as well as in response to questions related to Volume 2 (comments on “Governance and Platform Architecture” and “Platform Provider Role and Control”). The Foundation is concerned that the proposals neither consider platforms that are governed by communities nor do they sufficiently factor in the resource limitations of nonprofits. These statements have been made in the past in our [Public Bill Committee submission in December 2022](#), public comments in [March 2022](#) and [November 2022](#), suggested amendments to the OSA when it was still a Bill in the House of Lords, additional public statements in [June 2023](#) and [September 2023](#), and other engagements since. The concerns are as follows:

- ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

1. The proposed Governance and Accountability framework does not appropriately account for the role of effective community self-governance, because it is modelled primarily on centralised approaches to trust and safety. The framework’s implicit focus on centralised platform governance is evident in Ofcom’s suggestion that all “multi risk services” must have “A Code of Conduct that sets standards and expectations for employees around protecting users from risks of illegal harm.” The obligation to “track evidence of new kinds of illegal content” is another example of an underlying assumption that services systematically monitor and remove content in a centralised manner. Community self-governance is essential for Wikipedia and the other Wikimedia projects to continue to function and provide accurate educational content that is free and open to use for everyone. Because the framework does not adequately consider community-led governance mechanisms, it does not benefit platforms like Wikipedia.

2. The Foundation and Ofcom are aligned in the belief that robust governance processes are an effective way of ensuring good risk management and can make a contribution to reducing online harm (Volume 3, page 6). For this reason, we ask Ofcom to consider the effectiveness of our longstanding community governance mechanisms. Wikimedia’s platform governance is based on a democratically created and publicly available [Universal Code of Conduct](#) (UCoC). The [Enforcement Guidelines](#) of this code of conduct are subject to a democratic mandate of the global base of volunteers who contribute to the projects. The guidelines [detail the](#)

[roles and responsibilities governing the ecosystem by type of violation](#), including the limits to where the Foundation can get involved. This power to determine limits to the Foundation's powers, considered alongside the fact that this decision is also made through a democratic process, sets Wikimedia projects apart from other platform providers—including those which collaborate in some areas on trust and safety policies with their community of users. Communities of different Wikimedia projects continue to independently develop and enforce their own policies within the UCoC's framework. Furthermore, the UCoC has been explicitly incorporated into the Foundation's [Terms of Use \(ToU\)](#) during its last collaborative reform of the ToU with its communities.

3. It is important to note that the Foundation was created two years after the launch of Wikipedia for the purpose of supporting and serving the volunteer community as the best way to achieve its free knowledge mission. The self-governing community-led Wikimedia model stands apart in that the Foundation plays a strictly limited, supporting role in enforcement, governance, or otherwise managing governance mechanisms like the [UCoC](#) or [Enforcement Guidelines](#) that the community has established. These governance structures continue to evolve, guided by a democratically elected community self-governance body, and based on the [Universal Code of Conduct Coordination Committee](#) framework as defined in the [Enforcement Guidelines](#). In large measure, Wikimedia projects are unique. They are one of the few online spaces that are self-governed by a community of users united around a goal: that is, to openly share knowledge for free with the rest of the world. This model only works if legislation around platform liability and governance enables the Foundation to respect the limitations of its own ability to intervene in community decisions.

4. The Foundation's Board of Trustees has a relative majority of members recommended to it for appointment by regular, democratic elections in which the global community of Wikimedia volunteers participate. These elections are transparently conducted under the guidance of a global volunteer [Elections Committee](#). Since its inception, the Board of Trustees has always had a majority of members with practical experience as active project community members.

5. The Foundation dedicates entire teams of staff to collaborate with and support the self-governance bodies of the Wikimedia volunteer community on trust and safety issues. Foundation leadership who manage these staff members are routinely available to the volunteer community and any interested member of the public during publicly accessible [Conversation Hours](#). These are held regularly on Zoom, live-streamed on Youtube, and include publicly posted minutes.

6. There are also practical limitations to the governance and accountability measures that Ofcom has outlined, especially for nonprofit organisations and smaller service providers. Imposing minimum risk assessment requirements that are not mandated in the Online Safety Act itself is one example of how Ofcom has opted for more resource intensive approaches for all platforms to implement, regardless of platform type or capacity (9.99 and 9.100).

7. The Foundation appreciates Ofcom's commitment to advancing safety, but this desire must include balanced approaches that accommodate the variety of service providers that exist online. We understand that Ofcom might want to stand firm in its attempt to go beyond what statute itself requires, but suggest that it consider

platform variety (including variety of capacity) or—at least—consider reducing the burden of complying with this approach as much as possible. Options to do this include: (i) minimising what the "minimum" (non-statutory) requirements are; (ii) avoiding prescriptive guidance like the suggestion to revisit the UK-tailored assessment every 12 months; (iii) articulating when departing from the guidance might be justified, such as if a platform can show evidence that the community using its services can sufficiently monitor and respond to risks. It could be counterproductive for a nonprofit with effective existing processes to have to dedicate resources to additional governance and accountability measures like extra training, internal tracking and reporting, internal audits, and more, especially when the funding for these measures would cut into resources meant to support publicly beneficial activities.

The Wikimedia communities have outlined a strong, public, and participatory governance framework over the course of 20 years. The Foundation's obligation is to effectively collaborate with these self-governing communities in order to work towards our shared educational, public interest mission.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 4:

i) Do you agree with the types of services that we propose the governance and accountability measures should apply to?

As noted in the replies to volume 2, the Foundation respectfully disagrees with the proposal, since it does not sufficiently take into account public interest projects delivering public goods, especially those with volunteer community self-governance models.

ii) Please explain your answer.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 5:

i) Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?

The Foundation and the Wikimedia projects are transparent (see our bi-annual [transparency report](#)). Anyone with internet access can audit the Foundation's activities, access the projects' [statistics](#) and [metrics](#), and monitor what content is

added, debated, edited, deleted, and improved across the projects. These steps of the content moderation process are publicly accessible to everyone, for example on ["Talk" pages](#), where editors collaborate, debate, and agree on improvements to Wikipedia articles or other Wikimedia projects' webpages. The Application Programming Interfaces (APIs) and developer environments are also open. Members of the Wikimedia communities, academic researchers, and other interested stakeholders can freely design and deploy tools and machine learning (ML) models to track the performance of the projects' ecosystem.

We are also concerned about the possibility of financing any external audit requirements. For the Foundation, such a mandate would be an example of where there is "potential for significant costs" (as Ofcom acknowledged on page 6 of Volume 3). If the proposed fees imposed by potential auditors in the context of the EU Digital Services Act (DSA) are any indication, then we can expect audit fees to be in the six-to-seven figure range. Payments to the auditors would go hand-in-hand with corresponding internal costs to support the audit. Investments would also be required to cover detailed training, design, and consultancy work required to make these UK-specific. Complying with multiple external audits across jurisdictions like the EU, UK, and others is an existential risk for a nonprofit, and diverts sizable resources away from their public interest mission.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 6:

i) Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?

The Foundation, as a nonprofit organisation, will not practise bonus-related performance and remuneration management for its staff. Moreover, it would oppose introducing such practices mandated for public interest platforms under government regulation. The reason is that introducing these incentives might undermine the very alignment with self-governing community partners, which is based on a shared educational and public interest mission that delivers the beneficial societal outcomes upon which the public has come to rely.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Service's risk assessment

Question 7:	
i)	Do you agree with our proposals?
<p>We respectfully disagree with Ofcom's proposals as we perceive them to be targeted only towards top-down, for-profit-oriented platforms and service providers.</p>	
ii)	Please provide the underlying arguments and evidence that support your views.
<p>The challenge for an organisation like the Wikimedia Foundation is not only that we do not have the infrastructure in place to meet top-down governance requirements, but also that such requirements threaten the very nature of our platform, which is self-governed by the global Wikimedia volunteer community.</p> <p>These proposals also impose disproportionate costs on public interest platforms when compared to for-profit companies, since they effectively reallocate very scarce resources away from the practical work of identifying threats and improving the safety of audiences and volunteer contributors towards documentation and reporting. Moreover, reallocating staff and resources away from the daily practical work of keeping contributors and audiences safe will ironically increase the safety and human rights risks to users on public interest platforms, as we have detailed with concrete examples in previous submissions and engagements with Ofcom.</p>	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Specifically, we would also appreciate evidence from regulated services on the following:

Question 8:	
i)	Do you think the four-step risk assessment process and the Risk Profiles are useful models to help services navigate and comply with their wider obligations under the Act?
<p>Given that the Wikimedia projects are public interest platforms self-governed by its volunteer communities, the Foundation is unclear how the proposal would enable us to better manage risks in our ecosystem's context compared to the status quo—which offers the same content and the same participatory options anywhere in the world. Our approach to risk assessments and mitigation already includes mandatory and voluntary assessments and reports, such as the work towards systemic risk assessment under the EU DSA, a comprehensive Human Rights Impact Assessment, and a Child Rights Impact Assessment.</p> <p>Regulatory fragmentation occurs when regulated entities have to comply with dif-</p>	

ferent regulations that are similar but not identical, or conflicting, in multiple jurisdictions or within the same jurisdiction. Fragmentation of this kind, as also noted by the [OECD](#) and other trusted actors in the evolving space, imposes disproportionate costs on public interest platforms.

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 9:

i) Are the Risk Profiles sufficiently clear?

NA

ii) Please provide the underlying arguments and evidence that support your views.

iii) Do you think the information provided on risk factors will help you understand the risks on your service?

iv) Please provide the underlying arguments and evidence that support your views.

v) Is this response confidential? (if yes, please specify which part(s) are confidential)

Record keeping and review guidance

Question 10:

i) Do you have any comments on our draft record keeping and review guidance?

NA

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 11:

i) Do you agree with our proposal not to exercise our power to exempt specified descriptions of services from the record keeping and review duty for the moment?

Since Ofcom is expanding the recordingkeeping regime by adding many requirements beyond those in statute (see, for example, paragraph 10.18), we respectfully suggest that Ofcom ought to have taken exemption much more seriously than it has. Paragraphs 10.26-10.30 do not even set out the evidence Ofcom considered in coming to its position that nobody should be exempt. We would encourage Ofcom to consider exemptions based on factors such as the nature or business model of a platform, and/or its profits or gross revenue. It would also seem less important for Ofcom to impose local, UK-specific recordkeeping and review obligations on platforms that are already complying with other requirements or voluntary

practices of this nature (for instance, DSA records), since those alternative records could provide an acceptable starting point for Ofcom's oversight.

Since these recordkeeping obligations are new, their benefits are unproven. It seems clear, however, that they would amount to significant burdens (especially for international platforms, should non-harmonized requirements continue to proliferate). Targeted exemptions would self-evidently enable public interest platforms to continue to operate cost-efficiently in the UK, without serious detriment to online safety.

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Volume 4: What should services do to mitigate the risk of online harms

Our approach to the Illegal content Codes of Practice

Question 12:

- i) Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?

The Foundation respectfully disagrees with the approach. Our reasons are the following:

- As written, the large services definition would require the Foundation to build new monitoring tools as well as change how we operate in relation to our users. We pride ourselves on not tracking users. For this reason, among others detailed in this consultation response, we do not know the size of our active UK userbase, and can only share an approximation of [6,000](#). Further caveats behind this sum include the fact that it refers to the amount of “active” contributors (defined as people who made five or more edits per month), and is limited to “active” editors on the English language version of Wikipedia. This is an issue that we also raised, and for which the European Commission provided us with accommodation, in its DSA implementation.
- The Foundation has raised substantial concerns about Ofcom’s proposed approach to risk profiles in response to the relevant volume as it relates to public interest platforms and, by implication, naturally has severe concerns about the proposed Code(s) that will be developed on that basis.
- The Foundation has concerns about the proposed approach to make all platforms subject to the Code(s) on a baseline irrespective of the actual, empirically-founded risks that they might entail. Doing so would be a disproportionate burden on small and medium enterprises as well as public interest platform providers with a low structural risk profile.

Ofcom’s proposed approach here—i.e., to impose measures on platforms simply because they are visited by many users, regardless of risk or actual resourcing—assumes that large services have greater capacity and/or capability. This is evident on page 4: “We consider larger services will tend to be better able to bear the costs of the more onerous measures than smaller services,” and also in paragraph 23.39, where Ofcom concedes the lack of evidence supporting related analysis.

In reality, there is no reliable logical connection between size of user base and ability to bear the cost of the measures Ofcom wishes to impose, since audience reach is at best an indirect proxy. A very small percentage of visitors from the UK are thought to be Wikipedia donors, and of those, few will donate more than £11.90 (the average donation is US \$15.00). The economics (and the operator's capacity) are entirely different compared to a platform such as YouTube, as previously mentioned.

If risk is ignored, as Ofcom proposes to do here, then proportionality could not possibly be ensured unless the platform operator's actual ability to bear Ofcom's burden is considered.

It also does not logically hold, as Ofcom suggests, that the wider a platform's reach, the more beneficial a measure will be:

- Firstly, the benefit of a measure depends on the platform's absolute level of risk, not its reach (as Ofcom seemed to acknowledge in footnote 150, and yet have disregarded overall).
- Secondly, the burden of a measure (e.g., interference with fundamental rights; cost of deploying more moderators; among others) can also scale with reach—which means that, in some cases, the cost-benefit ratio of a measure remains largely unchanged whether or not the platform is “large.”

The best approach is for Ofcom to always consider risk, not just reach. However, if Ofcom proceeds to consider reach alone when recommending some measures, then it would be imperative for “large” to be defined by reference to audience reach in addition to a more faithful indicator of ability to bear supplemental compliance costs. Suggestions are staffing and/or profitability.

Another factor could be nonprofit status, since diverting a nonprofit's resources towards administrative tasks—such as recordkeeping—necessarily takes them away from the pursuit of the nonprofit's core charitable purpose. In other words, the impact of measures being imposed by a regulator will be very different depending on whether they impact a nonprofit organisation or a for-profit business. Without accounting for this, Ofcom will struggle to ensure proportionality by looking just at numerically “large” status.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 13:

- i) Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk?

The Foundation respectfully disagrees with the proposal. Applying measures to platforms simply because they are "large" (as Ofcom currently proposes to define this) is both impractical and disproportionate—both in terms of the burden for some unfairly-captured platforms, and in terms of the privacy impact on users.

- ii) Please provide the underlying arguments and evidence that support your views.

More specifically:

- The numerical approach (7 million users) to defining large services does not appear to distinguish between users who are content contributors and users who are merely readers. As can be seen in our [publicly available statistics](#), even the largest public interest project that the Foundation hosts, English language Wikipedia, has only around [6,000](#) active contributors in the UK. This is due to the complexity of defining what it means to be a "user" of a Wikimedia platform—which we have explained, above, in this consultation response. "Users" of projects can be divided into two different types: 1) those who are passive readers; and, 2) those who are active content contributors. Another complicating factor is determining where these users are based: Wikipedia is language specific and, therefore, many readers and contributors of content in the English language will not necessarily be based in the UK.
- We also do not track the location of each reader of Wikipedia, or of our other platforms; instead, we track the (estimated) location of accessing devices. For example: If a person has four devices (i.e., personal laptop, tablet, phone, and work laptop), they would be counted four times. To distinguish specific users, including those who never log in or create an account, Ofcom's proposed numerical approach would, hence, require the Foundation to build and deploy surveillance infrastructure. Worse still, this would mean targeting the public globally specifically for the purpose of complying with the UK's national law. In addition to going against our values, building such surveillance infrastructure is not proportionate given our limited resources as a nonprofit organisation in terms of funds, staff, and technical infrastructure.
- The organisation has grave concerns that imposing such a requirement on a low-risk public interest platform, which has a strong record of effective self-governing communities, would pose a severe risk to these communities' continued engagement in a space that would now be required to surveil them—a strong, practical demonstration of distrust.

	<ul style="list-style-type: none"> It would also target the public of readers worldwide and, once such an infrastructure would be in place, malign state actors and state sponsored groups would most likely aim to get access to its capabilities.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 14:

i)	Do you agree with our definition of large services?
	<p>The Foundation respectfully disagrees with the proposal.</p>
ii)	Please provide the underlying arguments and evidence that support your views.
	<ul style="list-style-type: none"> The large services definition as written appears to require that the Foundation build and deploy surveillance infrastructure of its users exclusively for the purposes of complying with the Act, since the organisation is not currently tracking users. The numerical approach (7 million users) to defining large services does not appear to distinguish between users who are active contributors and users who are passive readers, despite the fact that these two types of users carry vastly differing risks: i.e., readers do not contribute to user-generated-content (UGC). This is significant because English language Wikipedia, for example, as seen in our publicly available statistics, has only around 6,000 active contributors in the UK. The difficulties associated with this approach are supported from our experience with the European Commission, which has had challenges with the EU DSA's user-counting approach. While the European Commission agrees with the estimates we have produced (its estimate was within 1-2% of our own), the Commission has a differing view of other platforms' reported numbers. If Ofcom perseveres with its proposal to rely on a userbase metric approach, then it may encounter the same methodological challenges. The reasoning put forward in 23.31 and 23.32 might be persuasive for the higher risk activities of many for-profit models, but does not apply to providers operating transparent, public interest platforms. <p>The user numbers methodology used by Ofcom neither accounts for everything on Wikipedia being publicly accessible nor for the fact that the vast majority of UK users access Wikipedia but do not interact with each other on the platform.</p>

The large for-profit social media platforms, without exception, paywall or require users to log-in to access their content. Therefore, there are few to no “visitors” who can access those sites without an account for any period of time, let alone interact with the content.

Wikipedia has always allowed visitors to freely access the encyclopaedia’s content without an account. There are approximately 6,000 UK users who actively interact on Wikipedia. These represent the number of active UK users of the U2U part of the service. There is a larger number of users (i.e, passive readers) that *do not* interact actively for lack of a user account.

In order to actively interact with other users, a reader would have to either create a registered account or their account would currently be identified by their IP address by default.

The vast majority of Wikipedia readers neither interact actively nor go on to make accounts or increase their participation on the projects. If Ofcom continues to combine these two user groups together, Wikipedia and its 6,000 active users will be shoehorned into the same risk category as services with over 7 million actively engaged UK users—who can and do interact actively on those platforms, and can and do have access to multiple ways of communicating and sharing content privately with other users from all over the world.

iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 15:	
i)	Do you agree with our definition of multi-risk services?
NA	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 16:	
i)	Do you have any comments on the draft Codes of Practice themselves?
<p>The Foundation has severe concerns about the Codes as written because:</p> <ul style="list-style-type: none"> • Their governance arrangements do not take into account public interest platforms with a strong track record of volunteer community self-governance that has consistently demonstrated effective risk mitigation benefiting the public. Instead, the Codes embrace an all-encompassing corporate top-down model of risk management. • The Codes apparently require active monitoring of ongoing UGC activity as well as intrusive age verification merely by virtue of the number of users. This would happen even in transparent, demonstratively low-risk public interest platforms, and impose unreasonable burdens on their providers as well as new harm risk vectors on their users. • They also require large services, of medium or high risk of one or more specified harms, and with at least one specified functionality, to offer enhanced user control measures. At least one of these required controls is completely unworkable in the case of Wikipedia. <ul style="list-style-type: none"> ▪ As a collaboratively maintained encyclopaedia, Wikipedia users can and do regularly comment on content posted by other users. The draft codes would require Wikipedia to offer user controls to disable comments, thereby undercutting the primary purpose and core functions of the platform. ▪ Because Wikipedia does not operate notable user or monetised subscription schemes, the codes would not require Wikipedia to implement user verification and labelling. However, draft codes related to age verification and preventing certain users from encountering content from “unverified” users would require some form of user verification and labelling. ▪ The Foundation’s recently published Child Rights Impact Assessment was carried out by independent experts, who consulted with a number of subject matter experts. Importantly, they did not recommend any age verification mechanisms to balance the need to protect children with the benefit 	

they receive from access to the encyclopaedic content hosted on Wikipedia. As noted by Article One Advisors, the specialised consulting firm that carried out this impact assessment: “[E]xperts advised against age verification and questioned [the] effectiveness of parental controls.”

- It is not evident to the Foundation what constitutes “run [...] on behalf of” in the context of 11.11 d).

The Foundation welcomes that Ofcom has not reached a view on the concept of trusted flaggers through dedicated reporting channels (DRCs) (11.15 d). We are concerned that this concept can undermine the community self-governance taking place on our platforms, which includes existing public editorial functions and dispute mechanisms that need to be transparent and accountable to these communities as well as the public. For this reason, trusted flaggers should not be a legal requirement.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Question 17:

- i) Do you have any comments on the costs assumptions set out in Annex 14, which we used for calculating the costs of various measures?

We are concerned that the proposed costs defined in 11.30(c) do not factor in opportunity costs. As the Foundation has outlined to Ofcom before, in the context of a nonprofit platform, the failure to adequately consider these opportunity costs can pose severe human rights risks to users and staff. The Foundation therefore remains unpersuaded by the reasoning put forward in 23.31 as potentially applied to public interest platforms.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Content moderation (User to User)

Question 18:

- i) Do you agree with our proposals?

The Foundation respectfully disagrees with the proposals for the reasons detailed in response to question 1) of volume 2). The proposal does not take into account public interest platforms that host educational content, freely available to everyone, which is created and curated by self-governing communities. It is important for the proposal to do so because the content in question is based on publicly available policies that these communities transparently enforce themselves.

ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Content moderation (Search)

Question 19:	
i)	Do you agree with our proposals?
NA	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Automated content moderation (User to User)

Question 20:	
i)	Do you agree with our proposals?
ii)	Please provide the underlying arguments and evidence that support your views.
<p>The Foundation favours consistent CSAM hash matching practices against established industry corpuses in all public UGC platform contexts. This practice works for an organisation like ours, which is perhaps unique in its wide reach-to-resourcing ratio as a nonprofit as well as particularly low risk profile (see the 2023 OECD report on the subject).</p> <p>While we remain unpersuaded that the proposals are practically feasible in the context of encrypted services without disproportionate negative rights impact imposed on the public, we want to be clear: Any requirements around keeping internet users safe from harm should protect end-to-end encrypted (E2EE) communications. These requirements should refrain from discouraging or prohibiting the use of E2EE communications, or from de-incentivizing platforms and other service providers from offering them to safeguard the privacy and safety of their users.</p> <p>The Foundation's recently published Child Rights Impact Assessment found no significant gaps in the existing policies and procedures used to identify and remove CSAM on Wikipedia. Nonetheless, because we take our commitment to combating CSAM seriously, we published a new Combating Online Child Exploitation Policy in January 2024, which seeks to further unify and codify many existing practices that serve to combat CSAM and protect minors on Wikipedia.</p>	

iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 21:

i)	Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated ‘publicly’ or ‘privately’?
<p>The conceptual distinction made in Annex 9 appears to be generally inapplicable in the context of the Foundation’s platforms and services because of the transparent and public interest nature of the Wikimedia projects and platform architecture—where content published by users on the platforms is necessarily accessible both publicly and globally.</p>	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Do you have any relevant evidence on:

Question 22:

i)	Accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services;
<p>In the context of CSAM, the Foundation favours consistent practices for public UGC platform providers regardless of service size. Economically viable options are available even to nonprofits like the Foundation. The Foundation has successfully operated such an infrastructure for ex-post facto identification of CSAM in-house for years.</p> <p>As outlined in answer 20), the Foundation remains unpersuaded that the proposal is technically feasible for encrypted service offerings without disproportionate negative rights impact imposed on the public. Hence, the economic aspects of the issue should not arise as a practical matter for that category of services.</p>	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 23:

i)	Ability of services in scope of the CSAM hash matching measure to access hash databases/services, with respect to access criteria or requirements set by database and/or hash matching service providers;
Refer to the Foundation's view on regulating that part of the vendor market outlined in answer 22.	
ii)	Please provide the underlying arguments and evidence that support your views.
As noted above, the Foundation's recently published Child Rights Impact Assessment found no significant gaps in the existing policies and procedures used to identify and remove CSAM on Wikipedia. The Foundation still takes its commitment to combating CSAM seriously and published a new Combating Online Child Exploitation Policy in January 2024, which aims to further unify and codify many existing practices that serve to combat CSAM and protect minors on Wikipedia.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 24:

i)	Costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy matching for CSAM URL detection;;
Answer 20 applies.	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 25:

i)	Costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services;
The Foundation favours a strict distinction between for-profit commercial service models (which have an inherent risk factor by virtue of selling products or services to consumers in exchange for monetary or other monetizable compensation) and nonprofit public interest platforms (which avoid fraud risks by virtue of not engaging in such activities in the first place). For this reason, we believe that nonprofit public interest platforms should not be subjected to measures being mandated.	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 26:

- i) An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around 'context' and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services.

The Foundation strongly favours alignment with the industry standard established by the relevant EU terrorism-related legislation. Fragmentation of regulation in this place makes enforcement less effective and less likely that practicable implementation will be consistent across both for-profit and nonprofit providers.

- ii) Please provide the underlying arguments and evidence that support your views.

- iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Automated content moderation (Search)

Question 27:

- i) Do you agree with our proposals?

NA

- ii) Please provide the underlying arguments and evidence that support your views.

- iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

User reporting and complaints (U2U and search)

Question 28:

- i) Do you agree with our proposals?

The Foundation is concerned that the guidance at 16.124 requires urgent attention, since—read in the light of paragraphs 12.80(a), 16.21, and 16.22—it is not workable. We worry that Ofcom's commitment to proportionality and international regulatory compatibility would be seriously undermined, and respectfully refer Ofcom to the EU DSA and other international laws, which show a way forward here.

- ii) Please provide the underlying arguments and evidence that support your views.

A. For purposes of clarity, below we cite the wording that causes concern in these paragraphs, some of them with added emphasis:

- At paragraph 16.124, Ofcom proposes to say: "Once a complaint has been received, it should enter the service's content moderation function. As set out in Chapter 12 – U2U Content Moderation, this means that *all services will need to handle the complaint as suspected illegal content under Measure 1* [emphasis added]."
- At paragraph 12.80, Ofcom also proposes to say: "[...] [W]hen a service *has reason to suspect* that content may be illegal content, it should either (a) make an illegal content judgement in relation to the content [...] [emphasis added]." Ofcom also sets out an alternative option, "(b)", which relates to setting harshly strict terms of use, and then assessing content against those rather than against UK law.
- At paragraphs 16.21 and 16.22, Ofcom discusses territoriality. 16.21 says: "We think that in the first instance, this means services have a choice. They may choose to do all the things the Act requires, for all their users no matter where in the world they are located. But they may instead choose to do those things only in relation to their UK users." Lastly, 16.22 says: "If a service wishes to comply with its duties around reporting and complaints in this narrower way, it will first of all need to know if the user who has submitted an illegal content complaint has been served this content in the UK."

Our understanding of Ofcom's proposed guidance here is that the Foundation would need to analyse each and every content complaint it receives—worldwide and regardless of language or project (from Wikipedia to Wikispecies to Wiktionary, to name but a few)—to determine whether the complaint relates to content that violates UK law. Alternatively, Ofcom suggests that we could amend our Terms of Use so that anything which could be infringing content under UK law is prohibited throughout the projects, and then apply this to the content complaints it receives. Whichever option is chosen, we would either need to impose this on users and content *worldwide*, or would need to change Wikipedia, the other Wikimedia projects, and our complaints handling processes in order to specifically discriminate between UK users and everybody else (where such discrimination could mean intrusive data collection about user location).

The financial costs of systematically conducting UK illegality assessments would be astronomical. Tailoring all our Terms of Use to be reliably stricter than UK law would also evidently be a disproportionate interference with the rights to access and share information in jurisdictions not subject to UK law. Finally, there would often be no real need to know if the user who has submitted an illegal content complaint has been served this content in the UK; that would seem like a disproportionate approach.

To exemplify why we find paragraph 16.124 (i.e., treating all reports as suspected content illegal in the UK) unreasonable, Ofcom should note that many of the content complaints we receive are along the following lines: a report that a minor, newly-incorporated business in Greece has tried to create a Wikipedia article about itself, that volunteer editors have decided does not belong in the global encyclopaedia. It would be senseless to have to “handle the complaint as suspected illegal content” (from a UK perspective), or to investigate whether they viewed the content from the UK.

By contrast, the EU DSA avoids all these problems, and arrives at a workable solution, by stating that content erasure should follow if user reports “are sufficiently precise and adequately substantiated to allow a diligent economic operator to reasonably identify, assess and, where appropriate, act against the allegedly illegal content.” The US Digital Millennium Copyright Act (DMCA) also places a minimum onus on the users who are making reports to give the operator some sense of whether a breach of US law is actually suspected. The DSA approach, which Ofcom can and should align with at paragraphs 16.124 and 16.21-22 (i.e., to avoid a whole host of problems), ensures a flexible, scalable approach. Once that is done, paragraph 12.80(a) becomes much more workable.

B. More general observations regarding content report handling, and the adverse effects of proposed guidance in this area

The self-governing volunteer communities that create and curate the freely available educational content on Wikimedia projects implement and operate their own user reporting and complaint mechanisms, and have done so successfully for more than two decades. The Foundation’s role is defined and constrained by the strong global governance structure outlined in response 3) for the governance and accountability volume.

The Foundation, therefore, respectfully disagrees with the proposals as written. They insufficiently enable the mechanisms essential to serving the public interest purpose of our organisation, that is, contributing beneficially to the online information ecosystem. Furthermore, our organisation is subject to—and not in charge of—the global, democratic [mandate](#) underlying the Universal Code of Conduct Enforcement Guidelines that the Wikimedia communities have created and approved.

In addition, the Enforcement Guidelines themselves mandate specific [enforcement pathway](#) development obligations for the organisation and [enforcement structure principles](#). The democratic mandate that underpins the Enforcement Guidelines binds

both the organisation and the communities contributing to and governing its platforms alike. It stands to reason that the Enforcement Guidelines align well with the public policy objectives that Ofcom aims to pursue with its proposals.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Terms of service and Publicly Available Statements

Question 29:

i) Do you agree with our proposals?

The Foundation remains concerned that the basic design vision underpinning the proposals excludes the reality of public interest platforms such as ours, which do not unilaterally set policy for their platform, but co-create foundational policies such as the Terms of Use (ToU) and Privacy Policy with the self-governing volunteer communities. This is done in transparent consultations that are open to anyone belonging to these communities. It is these self-governing communities who develop, implement, and operate local policies in pursuit of their specific educational aims and to complement the broad global framework agreed upon.

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 30:

i) Do you have any evidence, in particular on the use of prompts, to guide further work in this area?

The Foundation welcomes deepened dialogue about the nature of public interest platform governance; also, we welcome further discussion on the regulatory approach to the distinction between platform provider policies and local policies designed, implemented, and enforced by self-governing communities that pursue shared outcomes that are beneficial to society.

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Default settings and user support for child users (U2U)

Question 31:

i) Do you agree with our proposals?

The Foundation remains deeply concerned about the implied assumption underpinning the proposals: The requirement to build a surveillance infrastructure as

<p>outlined in responses 9) and 12) would mandate that public interest platforms such as the Foundation obtain user data they are not currently tracking. Given that the Terms of Use mandate hosting freely available educational content, and also that the vast majority of all visitors (i.e., passive readers) to our platform never create an account, it is not evident whom such settings would benefit or which kinds of risk it would mitigate in the context of Wikimedia projects.</p> <p>These proposals would require Wikimedia to collect and store more data on children who contribute to or access content made available by free and open knowledge projects like Wikipedia. This would diminish children's privacy on our projects and could diminish their ability to exercise their right to seek, receive, and impart information online, among other human rights.</p>	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 32:	
i)	Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around changing default settings?
See answer 31).	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 33:	
i)	Are there other points within the user journey where under 18s should be informed of the risk of illegal content?
See answer 31).	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Recommender system testing (U2U)

Question 34:	
i)	Do you agree with our proposals?
<p>The Foundation does not deploy any recommender systems on public interest platforms like Wikipedia.</p>	

ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 35:

i)	What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?
	NA
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

We are aware of design features and parameters that can be used in recommender system to minimise the distribution of illegal content, e.g. ensuring content/network balance and low/neutral weightings on content labelled as sensitive.

Question 36:

i)	Are you aware of any other design parameters and choices that are proven to improve user safety?
	As stated in our response to question 34, no recommender systems are used on Wikimedia platforms. As a result, we cannot provide much more in the way of design recommendations for these systems.
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Enhanced user control (U2U)

Question 37:

i)	Do you agree with our proposals?
	The Foundation respectfully and strongly opposes the proposals, since they do not take into account the collaborative approach to content creation and curation that the self-governing volunteer communities have successfully employed for more than two decades.
ii)	Please provide the underlying arguments and evidence that support your views.
	Ofcom's proposals are tailored to social media settings, in which they might well be both appropriate and proportionate, but do not take into account the back-and-

forths that take place publicly on Wikipedia “[Talk](#)” pages. These are public webpages where editors debate and improve the [reliability](#) of content based on [verifiable sources](#) in order to produce [trustworthy, neutral, and accurate](#) information. They are, however, not general purpose chat rooms or forums to discuss current events. On the “[Talk](#)” page for the English language article of [Queen Elizabeth II](#), for example, the volume and rapid work of editors collaborating to improve the article respectfully and accurately after the monarch’s death is well demonstrated.

Wikipedia’s “Talk” pages are effectively self-regulated volunteer environments. Debates there often take weeks to discuss and agree on the specific language to be included into the related Wikipedia articles serving the public. Such discussions cannot successfully occur if, for example, malign actors aiming to undermine the integrity of these self-governing editorial processes can prevent contributors from scrutinising and developing arguments and counter-arguments to proposals.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 38:

i) Do you think the first two proposed measures should include requirements for how these controls are made known to users?

See answer 37). The Foundation strongly opposes these proposals due to the significant risk they pose to the collaborative knowledge production processes playing out on Wikipedia “Talk” pages and comparable, effectively self-regulated collaborative environments.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 39:

i) Do you think there are situations where the labelling of accounts through voluntary schemes has particular value or risks?

See answer 37) once more. The Foundation strongly opposes these proposals due to the significant risk they pose to the collaborative knowledge production processes playing out on Wikipedia “Talk” pages as well as to the comparable, effectively self-regulated collaborative environments. Self-governing Wikipedia communities have traditionally developed, implemented, and enforced policies touching upon this topic on their own. Evaluative discussions have also found, repeatedly, that the best mitigation against risks of this type is the publicly accessible nature of all contributions and content, which is intentionally built into the existing platform architecture.

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

User access to services (U2U)

Question 40:

i) Do you agree with our proposals?

As outlined in answer 22, the Foundation supports consistent CSAM hash matching practices across public UGC platforms and consequently blocking by default all accounts uploading such material. However, the self-governing Wikimedia communities have the power to create and enforce specific, contextualised standards for their projects. These might build upon and go beyond the default approach if they deem these to help forward their mission of creating and curating educational content that is freely accessible to everyone.

The Foundation refers to answer 26) on terrorism issues while noting that the self-governing Wikimedia communities are, again, free to set and enforce standards themselves that build upon and go beyond the default approach.

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Do you have any supporting information and evidence to inform any recommendations we may make on blocking sharers of CSAM content? Specifically:

Question 41:

i) What are the options available to block and prevent a user from returning to a service (e.g. blocking by username, email or IP address, or a combination of factors)?

A user can be blocked based on all three factors identified in question 41). However, none of these blocks ultimately prevent a determined malign actor from returning to any platform, be it a regularly constructed for-profit service of any kind or a radically transparent public interest such as our own, which hosts freely accessible educational content.

The most effective manner to limit concerning activities is to construct platform architecture that limits incentives. For example, it is highly unlikely that users engage in storage or exchange of CSAM or terrorism-related content, recruiting or other concerning activities of such nature if the platform is transparent and collabora-

	tively governed. Making publicly transparent who is doing what and when while actively interacting with other users or specific content is highly dissuasive—especially when other users, the public, the press, and interested state institutions such as law enforcement agencies can view such interactions at any time. Ofcom recognizes in the introduction to this consultation that certain service offerings carry higher risk than others. The same is true for choices about platform architecture that providers make—or that, in the case of public interest platforms, providers make collaboratively and transparently together with their self-governing communities in pursuit of a specific public good.
ii)	What are the advantages and disadvantages of the different options, including any potential impact on other users?
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 42:

i)	How long should a user be blocked for sharing known CSAM, and should the period vary depending on the nature of the offence committed?
	<p>In the case of public interest platforms, indefinitely by default—as the self-governing English language Wikipedia volunteer community states in and enforces by means of their own publicly available policy.</p> <p>Banning user accounts from active participation on a public interest platform such as Wikipedia does not deprive the person or persons impacted from passively benefiting from the public interest good that the platform offers—that is, freely accessible educational content. The Foundation acknowledges that this fact is distinct from other types of platforms, involving more complex user account and platform services relationships.</p>
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

There is a risk that lawful content is erroneously classified as CSAM by automated systems, which may impact on the rights of law-abiding users.

Question 43:

i)	What steps can services take to manage this risk? For example, are there alternative options to immediate blocking (such as a strikes system) that might help mitigate some of the risks and impacts on user rights?
----	--

As outlined in answer 14), even very large public interest platforms have only a small number of active content contributors. As the OECD noted in its 2023 [“Transparency Reporting on child sexual exploitation and abuse online”](#) report, the National Center for Missing & Exploited Children (NCMEC) registered 29 reports by the Foundation in 2022 and eight the year prior. Final human adjudication of automated systems outputs were in line with appropriate rights impact assessments, and reporting to NCMEC was proportionate for such contexts. The Foundation opposes the line of reasoning outlined in 21.7 and 21.83 if applied to low-risk public interest platforms—consistent with our answers to 9), 13), and 41).

The Foundation acknowledges that large higher risk business model providers—such as social media companies and their huge volumes of concerning content (noted by the OECD in the same report)—cannot practically rely on such double-evaluation standards. However, our organisation has no first-hand experience with what is practically feasible for providers facing challenges of such magnitude and, hence, is in no position to take a stance on this matter.

ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Service design and user support (Search)

Question 44:	
i)	Do you agree with our proposals?
NA	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Cumulative Assessment

Question 45:	
i)	Do you agree that the overall burden of our measures on low risk small and micro businesses is proportionate?
The Foundation has several concerns about the proposals, which we have outlined in answers 9), 12), and 13).	
ii)	Please provide the underlying arguments and evidence that support your views.
Namely, our organisation finds the definition of “large” deeply problematic, since it does not take into account actual empirical risks due to platform architecture (see answer 41). Our organisation favours consistent CSAM hash matching practices for public UGC platform providers regardless of service size, as we have outlined in answer 20).	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 46:	
i)	Do you agree that the overall burden is proportionate for those small and micro businesses that find they have significant risks of illegal content and for whom we propose to recommend more measures?
See answer 45).	
ii)	Please provide the underlying arguments and evidence that support your views.
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 47:

- i) We are applying more measures to large services. Do you agree that the overall burden on large services proportionate?

See answers 14), 43), and 45). The proposals are neither appropriate nor proportionate for public interest platforms.

- ii) Please provide the underlying arguments and evidence that support your views.

- iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Statutory Tests

Question 48:

- i) Do you agree that Ofcom's proposed recommendations for the Codes are appropriate in the light of the matters to which Ofcom must have regard?

As applicable to public interest platforms, the proposals are not appropriate for the reasons outlined in answers 1) through 47). However, the Foundation welcomes the professionalism and thoughtfulness clearly visible in the proposals and their approaches. We consider that they establish a sound foundation on which to collaboratively explore the unique societal benefits and distinct regulatory challenges that public interest platforms represent.

- ii) Please provide the underlying arguments and evidence that support your views.

- iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Volume 5: How to judge whether content is illegal or not?

The Illegal Content Judgements Guidance (ICJG)

Question 49:

i) Do you agree with our proposals, including the detail of the drafting?

ii) What are the underlying arguments and evidence that inform your view?

A. Comments relating to proposed wording in Ofcom's sections discussing CSAM

The Foundation has a zero-tolerance approach to CSAM. For very good reasons, the underlying laws have been carefully designed and worded. However, we are concerned that Ofcom's proposed guidance may lead to over-censorship and banning of users, even on other platforms.

Due to how some of the current guidance is worded, even linking to Wikipedia articles could be problematic. There are several places where the proposed guidance could be amended.

1. Types of material

A drawing such as the encyclopaedic illustration on the "Lolicon" Wikipedia article—even assuming it depicts one or more children and is explicit—is not covered by the 1978 Protection of Children Act. It is not covered by the act because it is neither a photograph nor a pseudo-photograph, but a "manga"-style cartoon drawing. This level of nuance is important. Ofcom's current guidelines does not allow for the same distinction to be made.

At paragraph 26.144, Ofcom proposes to say: "Child sexual abuse material, or 'CSAM', refers to indecent or prohibited images of children (including still and animated images, and videos, and including photographs, pseudo-photographs and non-photographic images such as drawings)." The addition of "drawings" into this sentence impacts every single subsequent reference by Ofcom to "images" or "CSAM." As a result, even encyclopaedic articles like the one mentioned above would potentially qualify as CSAM under the OSA, which would be at odds with existing UK law like the afore-mentioned Protection of Children Act.

These problems are compounded by other proposed guidance.

2. Hyperlinking

We are further concerned by the fact that the guidance suggests that such content must always be treated as illegal (paragraph 26.145), and that Ofcom proposes to go even further by creating a stringent interpretation of UK law regarding hyperlinking.

At paragraph 26.154, Ofcom proposes: "We consider it particularly important to be clear that if the user of a U2U service posts a URL which leads to an indecent or prohibited image of a child, an obscene article or a paedophile manual, that is illegal content."

The cumulative effect of Ofcom's proposed guidance is to require platforms to treat any member of the public as distributing CSAM if they post a URL to the aforementioned encyclopaedia page or others like it.

In addition, proposed footnote 40 can also be read as suggesting that anyone merely clicking the link and, hence, accessing a copy of the Wikipedia article, is to be treated as "making" CSAM. We doubt that Ofcom had this intention in mind and appreciate the opportunity to surface these downstream effects.

3. Defences

Despite acknowledging the risks of over-censorship, Ofcom's proposal regarding CSAM does not offer guidance on how to defend against these over-censorship risks. The possibility might be all but dismissed due to the wording of paragraph 26.14, which says: "It is unlikely that a service will have reasonable grounds to infer that a defence is available."

On a system like Wikipedia, users encountering CSAM would most likely report it privately to the Foundation given both how easy this process is at the moment, as well as how awful and rare the nature of such content is. However, there is a plausible possibility, consistent with the platform's community-led model, that users would instead investigate, discuss, and moderate the content themselves. To discuss what to do with one or more uploads, users can very legitimately flag it to other users, and it can be deleted following community-led evaluation. However, community discussion and evaluation would be impossible without posting the hyperlink.

Anti-CSAM laws around the world either already have, or should have, a clear legal defence for such well-intentioned conduct. From a UK perspective, this could, for instance, be the "legitimate reason" provisions in the 1978 Protection of Children Act.

Ofcom's current guidance could prevent community-led content moderation processes from acting as the first line of defence against CSAM. At the same time, however, this guidance could encourage an overzealous platform moderator to delete community discussions that attempt to address CSAM, ban the users taking

on this challenging issue, and potentially even denounce those users to authorities. We are sure this was not Ofcom's intention, but nevertheless find it important to highlight these inconsistencies as a way to emphasise how important it is for Ofcom's guidance on CSAM to be explicit about statutory wording and design (including defences).

4. Global effect

Ofcom proposes that platforms must treat users in this way no matter where in the world they are (see paragraph 26.79). We urge Ofcom to consider more proportionate readings of the statute. The need for greater proportionality is highlighted by the ramifications of the paragraph in question, least of which are the conflict of laws problem it would cause in relation to legislations like the protections of EU DSA Articles 20-21 or European Media Freedom Act Article 17.

Moreover, "Any part of the [UK]" could be read to support the regional interpretation put forward by Ofcom at paragraphs 26.66-26.78 (namely, England vs Wales vs other regions). Hence, Ofcom does not necessarily need to add its globalised interpretation (i.e., anywhere in the world) at paragraphs 26.79.

B. More general observations regarding Volume 5

The Foundation respectfully disagrees with the proposal as it relates to public interest platforms. In order to achieve its nonprofit educational mission, the Foundation provides self-governing volunteer communities with online spaces to host freely accessible educational content as stated in its [Terms of Use](#). As also outlined in a [separate section](#) of the Terms, the Foundation has no editorial role since the content is global (see answer 1) and governed by a distinct global governance framework that the Foundation does not unilaterally control like other platform providers do (see answer 3).

Therefore, the Guidance as designed is not well-adapted to facilitating beneficial social goods and regulatory outcomes—even in those rare instances where legitimate concerns might arise, and even despite the platform architecture safeguards that cover all of the publicly accessible content on the platforms (see answer 41).

While the Foundation has no editorial role on the projects, the volunteer-run content governance ecosystem it supports has well-established and existing channels to surface concerns. Wikipedia is a global project that operates on the basis of language, not country or jurisdiction. German language Wikipedia (with articles in German) differs from English language Wikipedia articles in terms of their content, the community volunteers who add, monitor, and verify that content, and also those who police each others' behaviour on the project. This is a strength. It ensures that the community members who identify and remove illegal content have the contextual and cultural awareness required to best identify what is and is not illegal. Moreover, because of Wikipedia's transparent nature, regulators have ac-

cess to the same existing channels that Wikimedia community members and Wikimedia Foundation staff use to surface and monitor concerns related to illegal content. As it has done for other jurisdictions [such as India](#), our organisation looks forward to facilitating that process when the need might arise.

Within the draft codes and enacted regulations, there is a baked-in expectation of national versions of products or tailoring content on a per-country basis that simply does not apply here. Every French speaker, for example, accesses the same French language version of Wikipedia, regardless of whether they are located in France, Belgium, Canada, or Senegal. And this is precisely what allows people from around the world to collaborate to improve and enhance any given Wikipedia article. It is impossible for us to distinguish among Wikipedia content created by UK editors, content of non-UK origin that is visible in the UK, and content created by users whose location is unknown.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 50:

i) Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?

Putting aside concerns related to public interest platforms as detailed in answer 49), the Foundation agrees that the proposal is accessible.

ii) Please provide the underlying arguments and evidence that support your views.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Question 51:

i) What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?

The Foundation has concerns about the proposal as written. It insufficiently takes into account the jurisdiction-specific complexities arising for a reasonability standard confronted by services without any offerings specifically directed towards the UK. The issue extends beyond the challenges acknowledged in 26.13 because it requires considerable contextual knowledge that can be difficult to obtain for service providers without UK-directed services. In addition, it often requires extra expenditures. As written in 26.14, it also requires additional investment in staff capacity, time, and resources for US-residing services to address and balance these next to legal obligations in their domestic jurisdictions. See the additional context provided in footnote 5) on the same page.

ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Volume 6: Information gathering and enforcement powers, and approach to supervision.

Information powers

Question 52:	
i)	Do you have any comments on our proposed approach to information gathering powers under the Online Safety Act?
Given that the volume acknowledges that further information is forthcoming, the Foundation has no current view on the proposal. However, we welcome the opportunity to resurface key concerns.	
ii)	Please provide the underlying arguments and evidence that support your views.
<ul style="list-style-type: none">• The Foundation provides a public and transparent platform. Because it is freely accessible both through the default experience and open APIs, most relevant questions that stakeholders—including regulators—might have about activities on the platform can be answered without requiring our organisation’s active collaboration. We welcome the intention outlined in 28.55 a) in order to take advantage of that opportunity.• The Foundation has no services specifically directed at the UK, distinct from the regulator’s experience regulating television and other services with a specifically-directed nature.• The Foundation has no legal presence in the UK, but welcomes visits from Ofcom to its sole office in San Francisco, US.• The Foundation has deep concerns about approaches and protocols from Ofcom that are designed without considering how these might apply to public interest providers. We consider that they risk undermining the effective volunteer community self-governance that is essential to creating the educational content offered on the projects and, consequently, their beneficial societal outcomes. The same is true for the DRC concerns outlined in answer 16) and governance expectations concerns detailed in answer 3). The Foundation stands ready to collaborate with the regulator through the effective model noted in answer 49).• The Foundation has no profit. The European Commission also reviewed and acknowledged the issue when concluding to waive DSA supervisory fees for Wikipedia following its classification as a Very Large Online Platform (VLOP).• Ofcom, like any member of the public, is welcome to remotely view all publicly accessible systems and processes in production on the Foundation’s platforms.• In light of the second and third points of this answer, the Foundation remains concerned about the practicability of 28.27 through 28.30.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

--

Enforcement powers

Question 53:
i) Do you have any comments on our draft Online Safety Enforcement Guidance?
As noted in answer 52), the Foundation neither has services directed at the UK nor legal presence in the UK. We look forward to exploring together with Ofcom how to best collaborate on public interest platform provider issues should such a need arise, especially in relation to the educational content freely accessible on the Wiki-media platforms and supporting services.
ii) Please provide the underlying arguments and evidence that support your views.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Annex 13: Impact Assessments

Question 54:	
i)	Do you agree that our proposals as set out in Chapter 16 (reporting and complaints), and Chapter 10 and Annex 6 (record keeping) are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English?
<p>The Foundation's mission is to foster the sum of all human knowledge and make it available to every person everywhere on the planet. We welcome the opportunity to continue doing so in a manner that can positively impact Ofcom's activities related to the Welsh language. One of the organisation's partnering communities, Welsh language Wikipedia, currently consists in large part of users who appear to be located outside of the UK—based on the project's contributor statistics. It will be interesting to study whether an enhanced role of Welsh in UK internet use practices might foster more UK-based participation in the project.</p>	
ii)	If you disagree, please explain why, including how you consider these proposals could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.
NA	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

UNIVERSAL CODE OF CONDUCT (UCoC) INFORMATION FOR ANNEX

In 2021, the Wikimedia Foundation, the nonprofit that hosts Wikipedia, launched a first-of-its-kind [Universal Code of Conduct](#) (UCoC) for Wikipedia and other Wikimedia companion projects. The code expands on Wikipedia's existing policies to create a global set of volunteer editor community standards for addressing negative behaviour on the website. It was developed hand in hand with more than 1,500 Wikimedia project volunteers.

The code was created because the Foundation believes in empowering as many people as possible to actively participate in Wikimedia projects and spaces in order to reach our vision of a world in which everyone, everywhere can share in the sum of all human knowledge. We believe our communities of contributors should be as diverse, inclusive, and accessible as possible. We want these communities to be positive, safe, and healthy environments for anyone who joins (and wants to join) them. We are committed to ensuring that it remains so, including by embracing this UCoC and revisiting it for updates as needed. Also, we wish to protect our projects against those who damage or distort their content.

The UCoC provides a baseline of behaviour for collaboration on Wikimedia projects worldwide. Communities may add to it, maintaining the criteria listed here as a minimum standard, and develop policies that take account of local and cultural context. It applies to everyone who interacts and contributes to online and offline Wikimedia projects and spaces without any exceptions. Actions that contradict the UCoC can result in sanctions by designated community members, such as functionaries or as appropriate in various local contexts.

The Code's most distinguishing standards include:

- Clearly defining acceptable behaviour
- Delinating harassment on and off the projects for all Wikimedia participants
- Preventing the abuse of power and influence to intimidate others
- Combatting deliberate introduction of false or inaccurate content
- Providing consistent enforcement processes and shared responsibilities between the Foundation and volunteer communities

The [UCoC Enforcement Guidelines](#), on the other hand, describe how the community and the Foundation will be able to achieve the goals of the UCoC. This involves promoting understanding of the UCoC, engaging in proactive work to prevent violations, developing principles for responsive work to UCoC violations, and supporting local enforcement structures.

Through a series of online consultations, discussions, and stages of feedback, the final enforcement guidelines were created by and for volunteers from around the world who contribute to Wikipedia and other Wikimedia projects. Prior to ratification by the Board of Trustees,

the enforcement guidelines were ratified by a vote from the Wikimedia volunteer editor communities.

In line with the Wikimedia movement's principle of decentralisation, enforcing the UCoC will be a shared responsibility. Communities will enforce the Code and act as a first line of defence for code violations. A separate Universal Code of Conduct Coordinating Committee, composed of volunteers, will be responsible for reviewing escalated code violations as well as monitoring its enforcement for further improvements.