Your response

Please refer to the sub-questions or prompts in the annex to our call for evidence.

Question	Your response
Question 1: Please provide a description introducing your organisation, service or interest in protection of children online.	*Is this response confidential? – N* The Wikimedia Foundation (**Foundation**) submits these comments in response to Ofcom's Online Safety Call For Evidence (**CFE**). The Foundation appreciates the opportunity to offer input to some of the questions posed by the CFE. The Foundation hosts several free knowledge projects, the largest of which is Wikipedia. Wikipedia, the free encyclopaedia, is a collaborative project created and maintained in over 300 languages by volunteers across the globe. The community of volunteers, who comprise the global Wikimedia Movement, collaboratively write and edit the content of the encyclopaedia. This community also creates and enforces rules regarding content and behaviour on the platform. Since Wikipedia is organised around a singular goal, the construction and maintenance of an online encyclopaedia, the types of potential harm on the platform are different than on most large social media platforms. The Wikimedia Movement's approach to addressing potentially harmful content has been tailored over years of community and organisational practice to promote fairness and minimise harm, and involves close collaboration between volunteer moderators and professional trust and safety staff. Wikipedia is first and foremost, an encyclopaedia, and encyclopaedias in the physical world are not age-restricted or censored based on the age of the person holding the volume, though they may contain material that could be considered disturbing for younger readers. We believe that access to knowledge is an important right for everyone of any age. We recognise and promote young persons' access to information and education; their ability to express themselves; their participation in projects and communities; their involvement in decisions that affect them; their access to reliable media on matters that interest them; and their privacy and

	dignity. These are all protected rights under the UN Convention on the Rights of the Child, which the UK ratified in 1991.
Question 3: What information do services have about the age of users on different platforms (including children)?	*Is this response confidential? – N* Wikipedia collects no mandatory demographic information on its editors and readers in accordance with our Privacy Policy. Because we collect so little information, requiring birthdate or age information for all editors, readers, and others who may access the site would run counter to our commitments to data minimisation principles and to upholding our readers' right to privacy. We deliberately do not collect any information about the age of our users, and should not have to do so, because of the educational nature of the Wikimedia projects, such as Wikipedia, which is curated and edited by volunteer editors according to detailed rules for neutral and factual sourcing. Many of the threats that younger users face on social media platforms are also less prevalent on Wikipedia due to the nature of the platform. Conversations on the platform are open to all and tend to revolve around the building of the encyclopaedia. There are no private messaging capabilities for users. Any exchanges between users on the Wikimedia projects are publicly posted and visible, meaning that private predatory behaviours cannot take place on the projects.
Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?	*Is this response confidential? – N* Content moderation on Wikipedia, and other volunteer-run free knowledge projects that the Foundation hosts and supports, is largely conducted by a community of nearly 300,000 global volunteer contributors. In addition to editing Wikipedia, volunteers also collaborate to create and enforce policies as well as adjudicate disputes that arise under those policies. Many Wikimedia projects have boards dedicated to proposing new policies for the projects which are discussed and voted on by other volunteer community members until consensus is reached, not simply a majority vote. On English Wikipedia, for example, proposals to introduce or change policies must be announced on the "Village Pump" noticeboard and "require discussion and a high level of consensus from the entire community for promotion to guideline or policy."

	Wikipedia is collaboratively edited, which means that almost every change to articles, even small grammatical edits, are based on community-determined standards and could be considered an act of content moderation. Every article has a "history" section, which indicates what changes have been made and who has made those changes, and a "discussion" section, where users can discuss changes they want to make before hitting "edit." These basic safeguards build accountability into the editing process and put content moderation tools and processes in the hands of the entire community. More experienced volunteers within the movement are given greater enforcement powers through a community selection process. These "administrators" and "bureaucrats" have the ability to block or unblock accounts, temporarily protect pages from being edited, and delete pages entirely. These volunteers have typically engaged extensively with the projects by contributing hundreds of edits when they are selected as administrators, and much of the proactive work to prevent vandalism and non-relevant content is done at this intermediary level of volunteer enforcement. On English Wikipedia, our largest project, there is also an elected Arbitration Committee which handles disputes over content and conduct on the projects. These cases involve formal hearings, which can be private or public, as well as a formal appeals process. Once a dispute is settled, the Arbitration Committee will publicly publish its decision along with any consequences which have been taken. While much of this dispute resolution is processed wholly within the volunteer community, the Foundation's trust & safety and legal teams regularly engage in dialogue with users and community members, providing community members with opportunities to ask staff about policy decisions or other issues of concern. This close collaboration has led to initiatives like the Universal Code of Conduct, a policy developed with the community that offers new levels of protection for volunteers on Wikimedia projects when it comes to conduct disputes. Finally, there are certain situations which cannot be handled by volunteers and are escalated to the Wikimedia Foundation trust & safety emergency response team to address. This includes situations where there is a threat of serious harm to someone's physical safety, as well as some higher

	level conduct issues which require a full, confidential investigation. This type of escalation is possible because of the trusted relationship between the Foundation and the volunteer administrators who maintain the Wikimedia projects.
Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?	*Is this response confidential? – N* The Wikimedia Movement's approach to addressing potentially harmful or illegal content has been tailored over years of community and organisational practice to promote fairness and minimise harm. This necessarily involves close collaboration between volunteer moderators and professional trust and safety staff. The Wikimedia community is already highly effective at removing illegal and harmful content on the projects. Researchers at the Berkman Klein Center for Internet and Society at Harvard University found that the median amount of time harmful content remained on English language Wikipedia was 61 seconds. However, improvements can always be made to further protect the human rights of readers and editors. Recognizing that many histories and perspectives have been excluded by structures of power and privilege, the Wikimedia Foundation envisions a key role that free knowledge projects can play in achieving inclusive, equitable, quality education, and in realising the human right to non-discrimination.We are committed to improving knowledge equity for women, LGBTQ+ communities, historically underrepresented racial and ethnic groups, people with disabilities, and communities in underserved regions, and in people's native languages. Furthermore, we recognize that not all content on our platforms may be appropriate for all audiences, including children. We strive to support our volunteers with training to ensure that content is handled sensitively and appropriately where it may be deeply disturbing or result in harm. The Foundation has commissioned a Child Rights Impact Assessment, scheduled to conclude this year, which will inform further steps that the Foundation and volunteer community might take. This, together with a wider program of improvements (for instance, our Anti Harassment Program) will also help us meet our obligations under the EU DSA, including the child-specific obligations in EU DSA Article 28. To ensure

	efficiency and focus, the OSB should be drafted and implemented in a way that dovetails with those requirements as far as possible. Age verification requirements would restrict user activity, create privacy and security risks for readers and contributors alike, and interfere with our commitment to data minimization. Age-gating Wikipedia would increase friction for readers, leading them to prefer sites not subject to the same requirements (or which are disregarding the law). It would further intrude on users' privacy, scaring away both privacy-sensitive users and users whose use of Wikipedia places them at high personal risk, (e.g. those that were arrested in Belarus or who want to read about LGBTQ+ issues in the Middle East). This means that some of the most important usage for our projects would be most likely to be harmed by mandatory age-gating. The Wikimedia Foundation believes that harms can be mitigated by supporting children, not barring them from all risky experiences and that what is "risky" varies substantially by age. The OSB threatens to define persons about to turn 18 as "children." Our policy–to look to mitigate harm for vulnerable users without employing age-based discrimination—allows us to hold the data minimisation approach that has served us and society well, for over 20 years.
Question 22: How are human moderators used to identify and assess content that is harmful to children?	*Is this response confidential? – N* Wikimedia volunteers do the majority of the assessment, moderation, and removal of content on the Wikimedia projects according to rules developed by both the Wikimedia Foundation and most importantly, the community itself. First, the terms of use [ToU] for the Wikimedia projects prohibit a broad range of harmful activities, and explicitly prohibit the misuse of the service for illegal purposes or activities. Our ToU are officially translated into 29 different languages, and we maintain a "Governance Wiki" where we maintain documentation related to policies and governance of the projects. The Wikimedia volunteer community also enforce project-specific policies which address illegal content, like these from English Wikipedia. The Foundation's ToU describe the rights and responsibilities of users and the Foundation, but each Wikimedia project also has its own set of policies and guidelines. These include speedy deletion policies, which allow administrators to

	immediately delete pages or media without going through the formal deletion procedures. Criteria that make articles or pages subject to speedy deletion include pure vandalism and blatant hoaxes, as well as attack pages that "disparage, threaten, intimidate, or harass their subject or some other entity, and serve no other purpose." On a more granular level, most edits, contributions, and other actions taken on Wikipedia are documented and publicly displayed. There are edit histories for articles and contribution lists for users—including anonymous users that are identified by their IP address. This policy is a safeguard and means that no one can upload, add, or edit content without leaving a (limited but important) footprint that allows their conduct on the site to be policed by the wider community, without this undermining the more general guarantee of privacy that allows legitimate users to confidently engage with the site. As described above, the Foundation's Trust & Safety team has processes in place and removes harmful content (e.g., CSAM, TVEC) if and when it is reported to us. Outside of those circumstances, the Foundation believes that the open, participatory content governance on sites like Wikipedia guarantees that what is on the project serves socially useful purposes. Changing that balance, by forcing the platform to dictate policy, then on a day-to-day basis monitor, assess, categorise, and selectively or wholly deny access to content, fundamentally changes that dynamic, leads to editor attrition, and thus harms the very thing that makes these projects functional, relevant, and socially useful. And further undermines the primary purpose and core function of Wikipedia: a freely available and widely-accessible online encyclopaedia that is not age-gated or censored based on the age of the person holding the volume.
Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?	*Is this response confidential? – N* Editors on Wikipedia employ a multi-layered approach to discovering and removing harmful speech on the projects. The Foundation seeks to empower users to participate in content and user moderation processes by, for example, providing them access to machine learning tools which they can use to improve or quickly remove content. While the Foundation may assist developers with building tools, they are used and maintained by community members.

	One of the tools editors can use is ClueBot NG, an automated tool which uses a combination of different machine learning detection methods and requires a high confidence level to automatically remove vandalism on the projects. ClueBot NG, like other automated tools deployed on our projects, is open source and subject to extensive public oversight, control, and policymaking. Another open source tool is a machine learning tool called Objective Revision Evaluation Service (ORES) which assigns scores to edits and articles in order to help human editors improve articles. Additionally, users with special privileges have access to the AbuseFilter extensions, which allows them to set specific controls and create automated reactions for certain behaviours. While automated tools are used to support existing community moderation processes, the bulk of the work is still done manually. Wikimedia uses select automated tools to scan for CSAM and works closely with law enforcement to report such content. Independent of community moderation processes and automated tools maintained by community members, the Foundation does not use other tools to reduce the visibility and impact of specific content on the projects. This is because algorithmic highlighting or amplification are not deployed on the projects. Unlike some other commercial platforms, the Wikimedia projects do not amplify or target content to maximise reader engagement or attention. To the contrary, the projects are structured in a way that does not allow content to spread virally on the projects, limiting the threat of illegal content being widely viewed.
Question 26: What other mitigations do services currently have to protect children from harmful content?	*Is this response confidential? – N* One additional step the Wikimedia Foundation has taken to mitigate risks and harms from content on our projects is to conduct a Human Rights Impact Assessment (HRIA) to identify human rights risks related to the Wikimedia projects as well as opportunities to address and mitigate those risks. Our inaugural HRIA report identified several steps which could be taken to reduce harmful content and mitigate risks to child rights specifically. As discussed in question 11, we are now conducting a Child Rights Impact Assessment, which will allow us to gain greater insight to the risks to child rights

	on the platform and additional opportunities for mitigation. We plan to follow this up by implementing a framework and set of processes for human rights due diligence across the organisation. These frameworks and processes will support the systemic risk assessment, mitigation and auditing program required by the EU DSA. Additionally, peer-to-peer education and training can help the volunteer community to be better prepared to address harmful content if they encounter it, and can even improve digital literacy skills overall. Wikimedia UK, a local chapter of the Wikimedia Movement, regularly works with schools and universities to put on classroom education activities, teaching students how to contribute to Wikipedia and educating them about how information is shared and spread online. These programs were designed with digital literacy skills development in mind, and help students to better exercise their writing, research, and critical thinking skills while navigating content online.