Call for evidence response form

Please complete this form in full and return to os-cfe@ofcom.org.uk

Contact phone number

Representing (select as appropriate)

Organisation name
Google

Email address
✂

Confidentiality

We ask for your contact details along with your response so that we can engage with you on this consultation. For further information about how Ofcom handles your personal information and your corresponding rights, see Ofcom's General Privacy Statement.

Your details: We will keep your contact number and email address confidential. Is there anything else you want to keep confidential? (select as appropriate)
No

Your response: Please indicate how much of your response you want to keep confidential (select as appropriate)
None

For confidential responses, can Ofcom publish a reference to the contents of your response? (select as appropriate)
N/A

Your response

See here for more detail on what to include in responses.

Preliminary question

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.
Is this a confidential response? No

About Google

Google's mission is to organise the world's information and to make it universally accessible and useful. Google achieves this by providing users with a range of services to exchange information and ideas. When it comes to children, we believe deeply in technology's ability to contribute to their education, development, and engagement with their peers. At the same time, we recognise that children also face a particular set of risks online, and we fully understand the responsibility we have to keep our child users safe.

We put the safety of children at the heart of how we develop our services — allowing them to fully and safely derive the benefits provided by our products. We recognise that new risks are continually emerging and we are always thinking about what more can be done to protect children. We also believe that protecting children should not mean shutting off access to valuable services when they have sufficient maturity to use them, but ensuring that they can access the service with appropriate safeguards. We look forward to working with Ofcom as it develops its approach.

Google's services[1]

Each of the products and services that Google offers has a different purpose, and their users have different expectations of the kind of content they will encounter on each, and whether they will encounter other users. Google builds products to be useful to all our users, including children. Furthermore, we have a minimum age requirement of 13 years old in the UK for most Google products. The only exceptions to this are parent-supervised experiences, such as Family Link and

[1] Most of our services (including Search and YouTube) are provided to users in the UK by Google LLC. This response relates to those services as provided by Google LLC and is focused primarily on YouTube and Google Search because of the reach of these services as well as the types and range of content that can be accessed through them.

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.

YouTube Kids (as explained further below). If we become aware that a user is younger than 13, we give them 14 days in which to verify their age or set up parental supervision, or the account will be disabled. For this reason, and save where otherwise indicated, where we refer to "child" or "children" in this response, we are generally referring to 13-17 year olds.

By way of summary of our products:

- **YouTube**: YouTube is an online video sharing platform that allows users to create, share and view user-generated content. Users can search for and watch videos, create a personal YouTube channel, subscribe to follow other YouTube channels, and create playlists to organise videos.

- **Search**: Google Search processes billions of searches per day. Google uses automation to discover content from across the web and other sources. Search engines are essentially indexes of the web. Hundreds of new web pages are published every second, and search services help users find the most relevant, authoritative information they are actively seeking. Search services are about access to information, not user engagement. They are inherently different from user-to-user services, both in terms of what they offer and how they are used. Search services are not primarily designed to host user-generated content, nor facilitate user-to-user interaction, nor amplify content through individual and personalized feeds. The benefits and risks presented by search services differ from those presented by user-to-user services.

- **Maps**: Google Maps is an online-based consumer map and navigation service. The core feature and primary content is the underlying maps, images and data facilitating navigation. To enhance the map and navigation service, Maps also allows users to share reviews, photos, and videos on places (including businesses).

- **Workspace**: This broad product category includes Google Drive (where files can be stored and shared via link), user-to-user video products (like Google Meet) and user-to-user messaging products (like Google Chat).

- **Photos**: Photos is a service that allows users to manage, edit, store, and share their photos and videos. Photos and videos can be shared through the service or via link.

- **Assistant**: Google Assistant allows users to use voice commands to search the web on mobile and home automation devices.

Please note that the majority of this submission largely concentrates on Google Search and YouTube.

Google revenues and business model

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.

In terms of our global and UK revenues:

- Alphabet Group revenue was USD 257.6 bn for calendar year ended 2021[2].
- Google UK Limited[3] turnover for the 18 months to the end of December 2021 was GBP 3.4bn.

Across the company, Google's main source of revenue is advertising – mostly from ads on the sites and apps of our own products and services. By serving ads, we can keep many Google services open to anyone with an internet connection, no matter where they live or what their background is, free of charge (see more detail here).

Google's approach to online child safety

We're committed to maintaining a safe and positive online experience for kids around the world. We have a three-pronged approach to kids & families:

- PROTECT: We offer default protections, robust child safety policies, and educational programs protect kids & teens from harm online and keep their data private and secure.
- RESPECT: We offer tools that offer choice and flexibility to respect each family's unique relationship with technology.
- EMPOWER: High-quality, delightful experiences empower kids and teens to safely explore, grow, learn and play online.

This means we do not take a one-size-fits all approach and we build products that are flexible and meet the developmental stages of children as they grow. What works for a child 0 - 12 should not necessarily be copied and pasted for teens that are 13-17 years old. And what works for one family may not be what others want for theirs. We believe every child and every family are different, but all children deserve to be able to access and use services in ways that respect their unique vulnerabilities and protect them from harm and abuse.

Over the years, we have been significantly investing in the policies, products and practices to help us protect children and their privacy. We have a team of researchers with expertise in child development, education, and privacy and safety. We know that shared expertise in these areas will lead to better products and experiences for kids and teens.

---

[2] As Alphabet is the parent entity of Google's broader corporate group, Alphabet's total revenues are submitted. The revenue data is sourced from Google's accounting systems. These represent booked revenues which are invoiced to customers (advertisers, developers and publishers), and reflect certain manual accounting adjustments (such as sales incentives and invoicing adjustments due to, for example, invalid clicks for Search (i.e., spam)). Revenue data is converted from local currency to USD based on a monthly average rate.

[3] Google UK Limited is a wholly-owned subsidiary of Google LLC. Google UK Limited does not provide any online services and is not the contracting entity to the terms of service governing the use of Google's consumer services.

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.

We've worked to hire those with expertise across a variety of disciplines and specialties, such as those with backgrounds in child development, cross-cultural parenting, children and media, as well as those with expertise regarding child safety and protection — all to ensure that our work reflects the evolving needs of kids, teens and families.

Google's approach in practice

We have strict content and privacy policies in place to protect our young users across our products, including for the ads young users see. We regularly review and update these policies and roll out product improvements. For example:

- Google Privacy Guides provides parents, teens and kids with access to age appropriate, clear and accessible information about our privacy practices, so they are able to make meaningful choices for accessing and controlling their data.
- YouTube:
    - Our Community Guidelines outline the types of content that are not allowed on YT, including cyberbullying, suicide and self harm, and content that endangers the emotional and physical well-being of minors. They also include policies on age-restriction, which requires users who want to view certain types of content to be signed-in and their account age must be 18 or older. In addition, we have content policies for YouTube Kids and supervised accounts that help guide which content is eligible to be included in the different content settings available for families. Finally, we've worked with third-party experts to develop kids quality principles, which we use to guide how we surface content in YouTube's recommendations, include content in the YouTube Kids app, and shape our monetization policies.
    - For users under 18, we do not serve targeted ads and we restrict sensitive ads categories; for our youngest users on YT Kids, Made for Kids content and in supervised experiences. We prohibit ads in additional categories such as foods and beverages, religion, or politics, as well as ads with inappropriate content such as scary imagery, crude humor, or sexual innuendo.
    - For users under 18, we set the default upload, livestream, and livechat settings to the most private setting available, and use transparency notices to remind them who can see their content with each setting option.
- Search:
    - Users under 18 have Safesearch turned on by default, helping to filter explicit content, including sexually explicit content like pornography, violence, and gore.
    - In the coming months, Search will roll out additional protections by blurring explicit imagery even if SafeSearch filtering isn't turned on. This setting will be the new default for people who don't already have the SafeSearch filter turned on.
    - Our ranking systems help you steer clear from unexpected shocking content and are designed to reduce unwanted sexualization by limiting porn or racy results for queries that aren't seeking them out.

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.

        - Thanks to advances in language understanding through AI, in 2022 we announced a 30% reduction in unexpected sexual results.
    - We apply the same approach to queries with potentially violent associations.

- Ads:
    - Google blocks all ads personalization and targeting based on age, gender, or interests for Google account users under the age of 18 around the world.
    - Expanded safeguards to prevent age-sensitive ad categories from being shown to teens, such as tobacco and alcohol, dangerous activities, weight loss, sweepstakes, etc.

Google takes a comprehensive approach to online child safety, involving large global teams and experts with a deep understanding of the key issues. In our experience, this comprehensive approach is the best way to avoid over-relying on any single tool and ensures we can address the different aspects that contribute to a safer online experience. Our approach has three elements: strong policies and guidelines, technological innovation and enforcement, and working in partnership with others:

- **Policies and guidelines**: Google has clear and publicly available product-specific "policies" and "community guidelines" developed in partnership with experts. Our response to Question 1 of Ofcom's first phase call for evidence provides more detail on our overall approach to policies and guidelines, and our response to Questions 11-15 in this document provides more detail on our approach to policies and guidelines in relation to children, specifically.

- **Innovation and enforcement**: Safety is incorporated into the design of each of our services and we constantly iterate and improve. We have built a range of products, tools and approaches across our different services that ensure children can have a safe and age-appropriate experience. For example, YouTube Kids provides a separate YouTube experience designed especially for children that parents can customise. Our answers to the questions throughout this consultation provides more detail on the systems and tools we have implemented specifically to keep children safe.

- **Partnership**: Whilst fully accepting our own responsibility to keep children safe on our own services, we have also established long standing partnerships with experts in areas such as media literacy for children and young people. Our work here includes the "Be Internet Legends" online safety learning programme that we deliver in partnership with ParentZone, which helps children to be confident and safe explorers of the online world. The programme has now reached 80% of UK primary schools; independent assessment of the programme in 2021 by IPSOS Mori found that, following the training, children are twice as likely to show an improved understanding of internet safety than those who haven't received the training. We'll be repeating our evaluation again in 2023 and will be

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.

glad to share the outcomes. Our response to Question 20 of Ofcom's first phase call for evidence provides more detail on our efforts to equip children and young people with the media literacy skills to experience the internet in a safe and positive way.

Google supports thoughtful regulation and will look to play its part

We believe thoughtful regulation is good for society, business and the internet. We recognise and fully support the Bill's particular emphasis on keeping children safe. We believe that policy interventions must support and balance all the rights of children, both to protect them from harm and to allow them access to the digital world and support their development. The Bill's goal to strengthen online child safety whilst minimising inadvertent impacts on their privacy and ability to access online services requires a careful balance to be struck.

This is an opportunity for Ofcom to establish a genuinely world-leading approach that provides for safe and age-appropriate access for children to the benefits provided by online services. We recognise that Ofcom will be entrusted with difficult decisions, particularly in balancing children's safety, privacy, and access to online services. We will support Ofcom by providing evidence that draws on our own extensive experience of keeping children safe online.

Children's access to services

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

Is this a confidential response? No

Google's services and child users

Google offers a wide range of services. With some limited exceptions, such as business-to-business services, most of our services are intended to provide value to our child users, whether that be in

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

helping them to access information and advance their education, express themselves and their identity online, or communicate with their peers.

We take two broad approaches to providing services for children:

1. Designing services specifically for children

We have led the way in designing products specifically for kids and families (e.g., Family Link, YouTube Kids, Supervised Experiences on YouTube, Teacher Approved Apps, and Google Kids Space).

As outlined above, our products have a minimum age requirement of 13 in the UK. For younger users wishing to use our services, we have designed a product with the needs of children specifically in mind. These services are adapted to children's developmental needs and have specific protections in place. YouTube Kids is a separate YouTube app for the youngest users, a safer and simple place where young kids can learn and explore their interests. YouTube Kids prioritises enriching and inspiring content - a result of content policies developed in collaboration with external experts; does not include any personalised advertising; and includes more tools for parents and caregivers to control and customise the experience for their families. For example, parents can block a video or channel, can handpick which videos to make available for their children, and also have an option to only allow content selected by trusted partners such as UNICEF.

2. Designing services for both adults and children and tailoring the service to each

For our flagship products that are built for everyone, we have a number of protections in place for younger users. A few highlights:

- **Search:** SafeSearch on by default, image search removal u18s
- **YouTube Supervised Experiences:** For older children whose parents decide their child is ready to explore some of YouTube's vast universe of content. Like YT Kids, it does not include any personalized ads

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

- **Wellbeing:** expanded wellbeing tools (take a break, bedtime reminders)

Most of our business-to-consumer services have mixed audiences of adults and children. These services require a thoughtful approach to ensuring users have an age appropriate experience. In particular, we believe it is important to empower parents and to give them the tools and knowledge to protect their families from the potential risk. These tools put parents in control of the content and experience that their children can access, giving them the flexibility to choose what is right for their children and their families, taking into account different maturity levels and developmental abilities.

Family Link supports parents and carers to help their children with safe access across Google's different products. All users under the age of 13 have to have an account managed with Family Link. We also offer Family Link parental controls by default on all Android devices for parents that want to set up content restrictions for their child. This functionality helps parents stay in the loop and guide their children as they explore and enjoy the internet. It allows parents to set and tailor digital ground rules that work for their family (including the ability to set screen limits, manage their apps, and lock their device). We also offer tips for families to help parents guide their children to make smart choices when using their own devices.

Additionally, YouTube also launched a new experience for parents to allow their children supervised access to YouTube. This supervised experience will come with content settings and limited features and our answer under the *"Providing an age appropriate experience across Google's services"* section will explain this function in greater detail.

We have taken extensive measures to encourage use of these tools, including marketing campaigns. We also provide user-friendly guides that cover both how to use our parental control tools and broader important principles for having a safe and positive online experience. We recognise the current focus of the Bill is on the steps taken by online services to protect users. It is important to also recognise the role of user agency, the actions of individual users, and parental support in contributing to a safer online environment. Where parents allow children to use their own devices, for example, this circumvents the protection offered by Family Link and increases the

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

risk of children accessing age-inappropriate content. Alongside fulfilling our own responsibilities, we are willing to discuss further with Ofcom how to strengthen awareness of parents' role in keeping children safe.

Principles in seeking to understanding users' age

Google is committed to ensuring that children and teens have appropriate experiences when using our products. Getting a better understanding of the age of our users is a part of that and we're committed to working with regulators, industry, civil society, and our users to get it right, in a way that balances the need to protect children with users' rights to access information and services, their privacy.

In the UK, in addition to age declaration, we use a model to help us infer if a user is over or under the age of 18 based on a variety of behavioral signals. Users have the option to verify their age using a government ID or a credit card. This method allows us to meet the regulatory expectations in those jurisdictions in a proportionate way. Not all jurisdictions place the same emphasis on age assurances, with the implications for data collection, as a tool for child protection. We work to deliver the most appropriate solutions in different jurisdictions within the existing legal framework, social norms and expectations.

Google's approach to age assurance respects privacy principles. In order to implement data minimisation principles for age assurance, we assess the types and amount of data to be collected, and carry out a balancing test that considers the risks posed by the processing and the safeguards implemented to ensure a proportionate approach.

We also consider other important principles, such as ensuring that users have notice, control over their data and choice. We have implemented a multi-layered approach to age assurance and give the user the opportunity to object to or seek correction of a determination about their age. We have also implemented steps to mitigate inclusion risks.

Google's processes to establish users' age

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

Across Google's consumer-facing products, **age declaration** is the first method for establishing the age of our users. We ask users for their date of birth during the Google account creation process. We do this in a neutral way, not nudging them to tell us a specific age and not allowing them to go back to the age screen once the date has been introduced, to avoid users gaming the system. If a user tells us they are under the age of consent, we direct them to our Family Link account creation flow to create a supervised account. If a user tells us they are above the age of consent but below 18, we offer them our set of default protections for users under 18.

As an added measure of protection, in the UK, we also utilise **age inference technology** across some of our products including Search and YouTube. Specifically, we deploy machine learning models that help to provide an additional level of assurance with regard to the veracity of a user's declared age, or to provide an indication as to whether or not a user is an adult where the user has not declared their age. This in turn helps us to determine the appropriate levels of protection to apply for our users. These models use a variety of signals, such as the types of sites a user is searching for or the categories of videos that they have watched on YouTube, as well as indicators like the longevity of an account, to make a determination on age. For example, searches for mortgage lending sites or tax assistance would be signals that are indicative of the likelihood that a user is an adult. We are constantly iterating the model to improve its accuracy. We do not collect any additional data for this purpose, but rather utilise only the data that is available in accordance with the user's privacy settings.

The model allows us to make an assessment of the likelihood of the user being above or below 18, but is not able to get to a fine-grained determination of whether the user is above or below the age of consent. Only hard-identifiers or other kinds of technology that would require the collection of additional data would be able to provide that information and we don't think the use of those for all types of processing across our services would be proportionate. For example, would it be proportionate to ask for an ID just to search for the nearest library? How would users react considering that some of them are already reluctant to provide their ID even for situations where it may be justified, for example to allow access to mature content? What about people who do not

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

have a passport or credit card – would they be excluded from accessing online services? We strongly agree with the ICO's recommendation in the Age Appropriate Design Code to avoid giving users no choice but to provide hard identifiers unless the risks really warrant such an approach.

We also require users to provide **additional verification** of their age in limited circumstances. For example, we might require age verification if a user is trying to access age-restricted content or services and we cannot otherwise establish with sufficient certainty that they are an adult, or if our age assurance process has classified the user as under 18 and they wish to access age-restricted content.

We use the following tools to obtain additional verification of age:

- **Government ID:** A user can submit an electronic copy of a valid, government-issued ID that shows the user's date of birth. That submission is then reviewed and approved (normally within 24 hours). To protect users, we then delete the copy after we have validated the user's date of birth. The user can cover the national identification number on their ID to keep it confidential.

- **Credit card verification:** A user can also verify their age by providing a valid credit card. The user is not billed as part of this process but may see a small authorisation on their account from the request, which is subsequently removed.

Beyond our current approach to assessing users' age, as a company we also engage in proactive outreach to identify and assess emerging solutions. Vendors in the market also sometimes approach us about potential solutions. We assess the viability of these solutions by considering important factors including the assurance level, coverage, use of data and implications for users' privacy, security.

Providing an age appropriate experience across Google's services

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

The information we obtain about the age of users is used to ensure that children and young people have a safe and age appropriate experience across our different services:

- On **Search:**
    - **SafeSearch:** By default, on Google Search, we turn on SafeSearch for declared users under the age of 18. Our SafeSearch feature helps filter explicit results from Google Search results, even when they might be relevant for the query. While these algorithms will never be completely accurate, turning on SafeSearch helps to filter explicit content, like pornography, from Google search results. Parents and schools have the option to lock in SafeSearch for supervised minors.
    - More generally, we also more prominently surface digital wellbeing features, and provide safeguards and education about commercial content. Our answer to Question 16-19 provides examples of how we help improve the visibility of authoritative information on mental health-related queries. We have recently announced an additional safeguard to protect users from inadvertently encountering explicit imagery on Search. As an added protection we will soon be launching a new safety service that will blur explicit imagery if it appears in Search results when SafeSearch filtering isn't turned on. This setting will be the new default for users who don't already have the SafeSearch filter turned on, with the option to adjust settings at any time.
    - **Removal:** While we already provide a range of removal options for people using Google Search, children are at particular risk when it comes to controlling their imagery on the internet. We therefore have a policy in place that enables anyone under the age of 18, or their parent or guardian, to request the removal of their images from Google Image results.[4]

[4] Note: while we can prevent an image from appearing in our Google search results, we cannot remove it from websites that host it. We therefore also give guidance to users about how to contact the site's webmaster and ask them to remove the content entirely.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

- On **YouTube:**
    - We age-restrict content on YouTube that does not violate our Community Guidelines but that may still not be appropriate for viewers under 18. For example, videos that contain vulgar language, violent or gory content, depictions of adults engaged in dangerous behaviour that might be emulated by minors, and adults consuming products that are not legally available to children (for example, alcohol or legal drugs). We continue to build on our approach of using machine learning to detect content for review, by developing and adapting our technology to help us automatically apply age-restrictions. Viewers attempting to access age-restricted videos on most third-party websites will be redirected to YouTube where they must sign in as an 18+ user to view it.
    - For users under 18, we do not serve targeted ads and we restrict sensitive ads categories; for our youngest users on YT Kids, Made for Kids content and in supervised experiences, we prohibit ads in additional categories such as foods and beverages, religion, or politics, as well as ads with inappropriate content such as scary imagery, crude humor, or sexual innuendo.
    - For those users under 18 who can create content, we set the default upload and livestream settings to the most private setting available, and use transparency notices to remind them who can see their content with each setting option.
    - We believe that it is vital that Section 14 in the latest version of the Bill on duties to protect news publisher content is implemented in a way that does not jeopardise our ability to protect children. The duties require Category 1 services to notify news publishers before "taking action" in relation to content. We would welcome further discussions and clarity from Ofcom on how to ensure that this will not prevent us from removing content that violates our Guidelines and is

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

harmful to children, or from applying warning labels, interstitials, or age restrictions to this content.

- **YouTube Kids:** YouTube Kids offers a set of parental controls to customise their child's experience. Parents can decide what content to make available for their child to watch (including content settings by age, approved channels only, or approved collections of channels only), set a timer to control screen time, block videos or channels, turn off search, and more.

- **YouTube supervised experiences:** We offer parents the ability to create supervised accounts using Family Link. Supervised experiences allow children under the age of consent to access YouTube with parents choosing the right content setting for their children: "Explore", "Explore More", or "Most of YouTube" (which generally align with content ratings for viewers 9+, 13+, and almost all content except Mature/18+). The YouTube supervised experience looks much like YouTube's flagship platform but with additional safety features.

Implications for Ofcom's approach to age assurance and age verification

Getting a better understanding of the age of users inevitably requires us to collect more data - the more certainty and/or granularity required, the more data that will need to be collected.

In our view, when assessing effectiveness and desirability of age assurance mechanisms, it is important to consider a number of aspects, including:

- The balance that needs to be struck between age assurance and privacy. Often, the higher the accuracy of an age assurance mechanism, the more intrusive that mechanism needs to be. That intrusiveness may not be proportionate in all scenarios or appropriate to the risks associated with the service in all cases.
- Even the most state-of-the-art technology has limitations, so no mechanism can be 100% effective.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

- The implementation of proportionate age assurance mechanisms sometimes requires a more gradual approach given that the level of assurance about the age of the user may get more accurate the more they use the service. This gradual approach can, of course, be complemented with other safeguards while the organisation reaches the relevant level of confidence about the user's age.
- Users' views should be considered**.** Users may disagree with the level of intrusiveness established. For example, based on our research and the feedback we have received, not all users agree to provide their ID card to prove their age, as they consider this to be sensitive information. Some users are also reluctant to provide their credit card details if it is not associated with a payment obligation. This makes it hard to impose such mechanisms for all use cases.
- The implications for digital exclusion. The 2021 Census indicates that eight million adults do not have a passport and the Electoral Commission 2022 tracker indicates that around 1.9m adults do not have any pre-existing forms of photo ID. In addition to that, there are as many as 80,000 cared-for children in England alone; these are substantial figures that illustrate the extent to which hard age verification can lead to digital exclusion.

Overall, we support a nuanced approach to age assurance that respects privacy principles and accounts for the implications for users' access to online services, including the risk of entrenching the digital divide. This includes considering the implications of specific age assurance approaches for specific groups - for example, relying on parental verification for children in care, or on ID verification for adults who lack photo identification, or who simply do not want to share their documentation.

In addition, as the remainder of this consultation response demonstrates, we believe that a holistic and comprehensive approach is the best way to keep children safe while preserving their access to online services. This includes combining effective policies and guidelines, systems and tools, and broader partnerships that protect children from harmful and age-inappropriate content and enable them to have a positive experience online. We believe that this holistic, outcomes-based approach is ultimately more effective than over-relying on any single tool, such as age verification.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Question 3: What information do services have about the age of users on different platforms (including children)?

Question 4: How can services ensure that children cannot access a service, or a part of it?

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

Content that is harmful to children

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

Is this a confidential response? No

Google's approach to age inappropriate content

Google is committed to minimising children and young people's access to age inappropriate content. The approach we take to this is tailored by product.

- **YouTube**
  - YouTube has its own Terms of Service and Community Guidelines.
  - As a baseline, we implement measures to help keep all users safe on the platform Our YouTube Community Guidelines prohibit a range of content regardless of the age of the user, including:
    - Content that encourages dangerous or illegal activities that risk serious physical harm, including content that praises, glorifies or encourages viewers to imitate anorexia or other eating disorders. We also prohibit

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

content that poses a serious risk of egregious harm by spreading medical misinformation about approved vaccines and content about COVID-19 that poses a serious risk of egregious harm.
    - Content that promotes suicide or self-harm, that is intended to shock or disgust or poses a considerable risk to viewers.
    - Content that is violent or gory intended to shock or disgust viewers, or content encouraging others to commit violent acts.
    - Content that endangers the emotional and physical well-being of minors.
    - Content that is meant to be sexually gratifying.
    - Content that threatens individuals or which targets an individual with prolonged or malicious insults based on intrinsic attributes.
- We also offer users an optional setting (Restricted Mode) that helps to screen out potentially mature content that some users may prefer not to view. This feature is designed to filter out graphic content that is permitted under Community Guidelines, but which is more appropriate for mature users.
- Users attempting to access mature/18+ content are subject to the age assurance pathway outlined in our response to Questions 2 - 5. Examples of content that may be subject to age restriction include:
    - Adult themes in family content: Content meant for adult audiences but that could easily be confused with family content. This includes cartoons that contain adult themes such as violence, sex or death.
    - For violent or graphic content, we may apply an age-restriction rather than remove content if that content provides enough context to understand it. For example, content showing victims' injuries in a road accident may be removed, but we may age-restrict that same content if presented with news coverage that explains the situation and context.
- We are constantly evolving these policies, and we have robust mechanisms to monitor compliance with, and enforce them.
    - In September 2020, we announced that we would be expanding the application of our age-restriction policies. While creators and our Trust and Safety content moderator teams had always been able to age-restrict content, from September 2020 onwards we developed and adapted our machine-learning technology to help us automatically apply age-restrictions to content.
- We measure our global enforcement of our Community Guidelines for YouTube and publish this information in quarterly Community Guidelines Enforcement Reports. Between July and September 2022, over 5.6 million videos were removed

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

from YouTube globally and more than 65,000 videos were removed in the UK. The breakdown of reasons for these removals is as follows:

- 2,017,231, removals in "child safety" category (26,528 removals in the UK)
- 1,032,392 removals due to "violent or graphic" content (4,701 removals in the UK)
- 769,692 removals due to "nudity or sexual" content (10,183 removals in the UK)
- 638,672 removals due to "harmful or dangerous" content (6,426 removals in the UK)
- 561,185 removals in "harassment and cyberbullying" category (5,381 removals in the UK)
- 218,885 removals in "spam, misleading and scams" category (2,875 removals in the UK)
- 156,676 removals in "hateful or abusive" category (4,757 removals in the UK)
- 121,423 removals in "misinformation" category (2,537 removals in the UK)
- 67,516 removals in "promotion of violence and violent extremism" category (818 removals in the UK)

- In addition to the data that we publish which shows the volume of content that we have removed from our services, we also publish "Violative View Rates" (VVRs) in respect of YouTube content. These VVRs show how many times content has been viewed before it is removed for breaching our policies. Our analysis shows that, between July to September 2022, of the 5.6 million videos removed globally from YouTube for violations of Community Guidelines, 36.4% were never viewed and a further 32.1% were viewed under ten times. In Q3 2022, the VVR was 0.1-0.11%, meaning that out of every 10,000 views on YouTube, only 10-11 came from violative content.

- **Search**
  - Google Search works at the scale of the web. Given the volume of content on the web, the speed at which new content is created, and the need for human review to identify certain types of harmful content (in particular, where legal nuance, competing rights, and context are relevant), it is not currently possible for Google to quantify the amount of content harmful to children on the open web at any given time.
  - However, Search can take action to either "delist" or "demote" content:
    - The term "delisting" refers to a process by which we "remove" links to certain web pages from the lists of displayed search results. This stops returning to users links to certain web pages (at times for all search

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

queries, and at times following only certain search queries) and thereby prevents those web pages from being accessed through Search.
        - The term "demoting" (referred to in the Bill as giving content "a lower priority in search results") refers to the process of ranking links to certain web pages lower in response to certain search queries and thereby makes it less likely that those specific web pages are accessed through the Search service.
    - Our content policies for Google Search specify that we:
        - Delist search results that lead to child sexual abuse imagery or material that appears to victimise, endanger, or otherwise exploit children (we also report CSAM, as explained in our response to Question 10 of Ofcom's first phase call for evidence).
        - Delist non-consensual explicit imagery (NCEI).
    - We also turn on SafeSearch for users under the age of 18. Our SafeSearch feature helps filter explicit results from Google Search results, even when they might be relevant for the query. Our response to Questions 2 - 5 explains how Safe Search works in more detail.
    - More broadly, our most effective protection is to design systems that rank high-quality, reliable information at the top of our results. We largely understand the quality of content through "signals": clues about the characteristics of a page that align with what humans might interpret as high quality or reliable. For example, the number of quality pages that link to a particular page is a signal that a page may be a trusted source of information on a topic.
    - For topics where quality information is particularly important—like health, finance, civic information, and crisis situations—we place an even greater emphasis on factors related to expertise and trustworthiness. We've learned that sites that demonstrate authoritativeness and expertise on a topic are less likely to publish false or misleading information, so if we can build our systems to identify signals of those characteristics, we can continue to provide reliable information.

Risk assessment and management

Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

Is this a confidential response? No

Please refer to material shared within Google's September 16th, 2022 submission to Ofcom and during Google's January 23rd engagement with Ofcom for the most relevant perspective on our general approach to risk assessments and governance structures.

As we continue to ingest new regulations focusing on these topics, Google will further refine its approach to risk assessments to be effective and in line with regulatory expectations. We look forward to providing additional updates on Google's risk assessment processes and practices later this year.

Terms of service, policy statements, reporting and complaints

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

Is this a confidential response? No

Terms of service

When we produce or update our terms of service and content policies, we take great care to ensure that they are clear to all users. This includes considering the needs of different categories of users, including children, parents and people with accessibility needs. Except for YouTube Kids, our users need to be at least 13 years old to use our services. Accordingly, we design our terms of service to ensure they are clear for children of that age and above. Our response to Questions 6 - 7 provides more detail on our terms of service on Google Search and YouTube, including how they relate to the indicative categories of primary priority and priority harmful content.

We track traffic to the webpage policies.google.com which contains our universal Terms of Service. In 2021, there were 425 million visits to this page over a 90-day period.

When we drafted the last major revision to our universal Terms of Service in 2020, and when we updated those terms further in 2022, we used the Flesch-Kincaid Readability Test to assess the readability and comprehensibility of the wording in those terms. The results of this test showed that users with a 9th grade level of education (i.e. 14-15 year olds) are able to read and understand our universal Terms of Service.

[5] See A1.2 to A1.3 of the call for evidence for more information on the indicative list of harms to children.

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

We continually review our internal standards on the language of our policies and how we can make our policies clear and intelligible to all users, including children, in areas such as how we write, format and present our policies externally. Internal guidelines are reviewed and made available to teams within Google responsible for policies

Reporting mechanisms

Our reporting mechanisms for policy violations are designed to allow users to immediately flag content of concern and ensure that users provide the information that we need to quickly assess the content for policy violations and to take action, where necessary. The specific mechanisms for flagging content to Google vary from service to service. We have adopted the following principles to ensure that user reporting is as easy and accessible for all users as possible:

- In many of our products and services, we provide clear "buttons" for reporting content.

- We provide detailed and user-friendly information for users on how to make complaints. This means that child users who may not notice or understand the buttons for flagging or complaining about content are still able to guide themselves through the process.
    - For example, Google's Help Center (available here) provides information for users on reporting, as well as information on how users can report inappropriate behaviour towards children, including grooming and other forms of child sexual exploitation.

- We include accessibility features (such as providing explanations on how to make

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

complaints in a read-aloud format in different languages).

- We provide information for users to understand the process once they have flagged content.
    - For example, on YouTube, we have produced a video on The Life of a Flag to help users understand what happens to content they have flagged.

For YouTube Kids, specifically, we operate a simpler reporting process than on the main YouTube service. In YouTube Kids:

1) The user first clicks the hamburger button instead of the "settings" button in order to find the "report" button.
2) The "report" button is 1 of only 2 options available under the hamburger button.
3) There are only three reporting options: inappropriate video, inappropriate audio, and thumbs down).

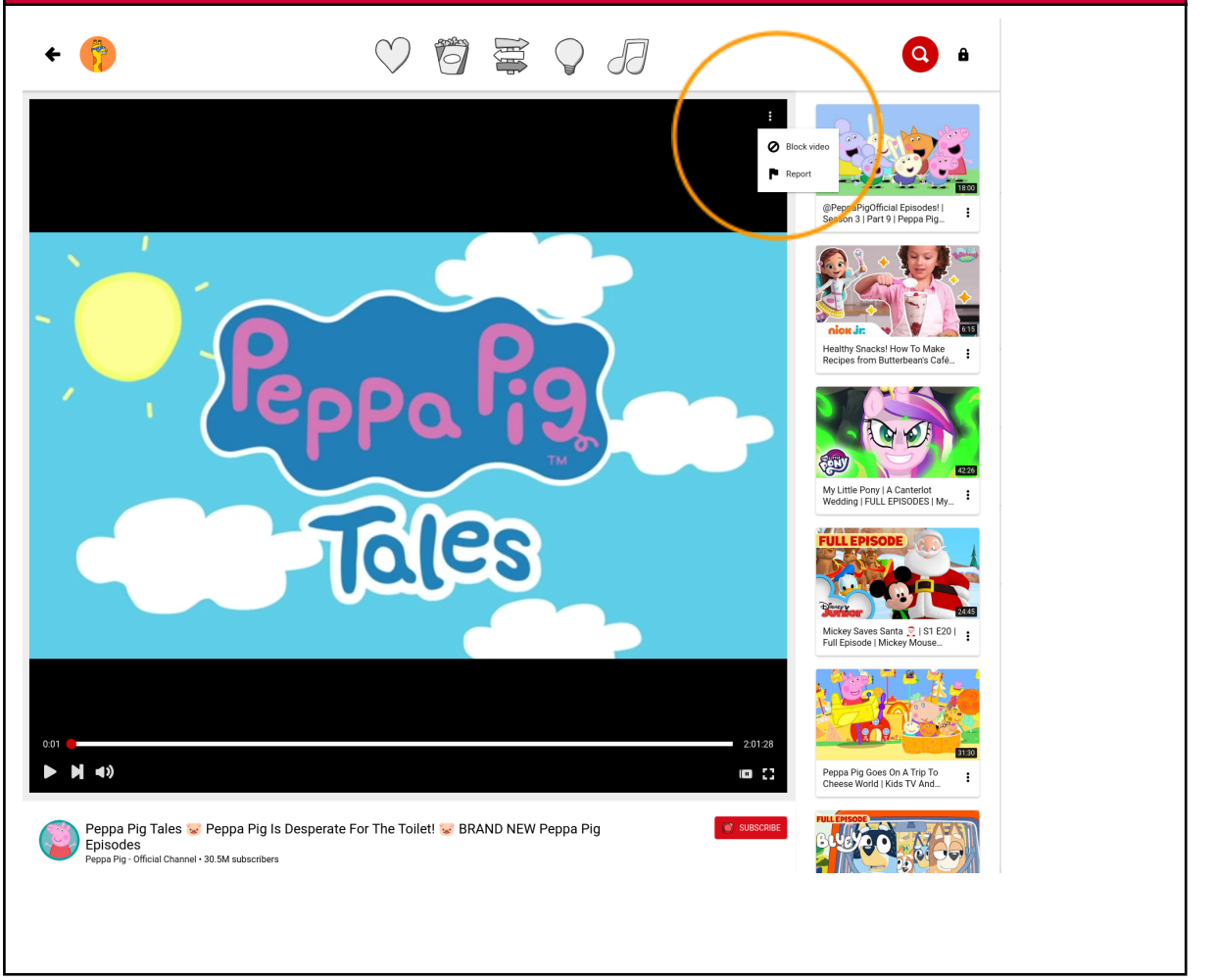
Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?



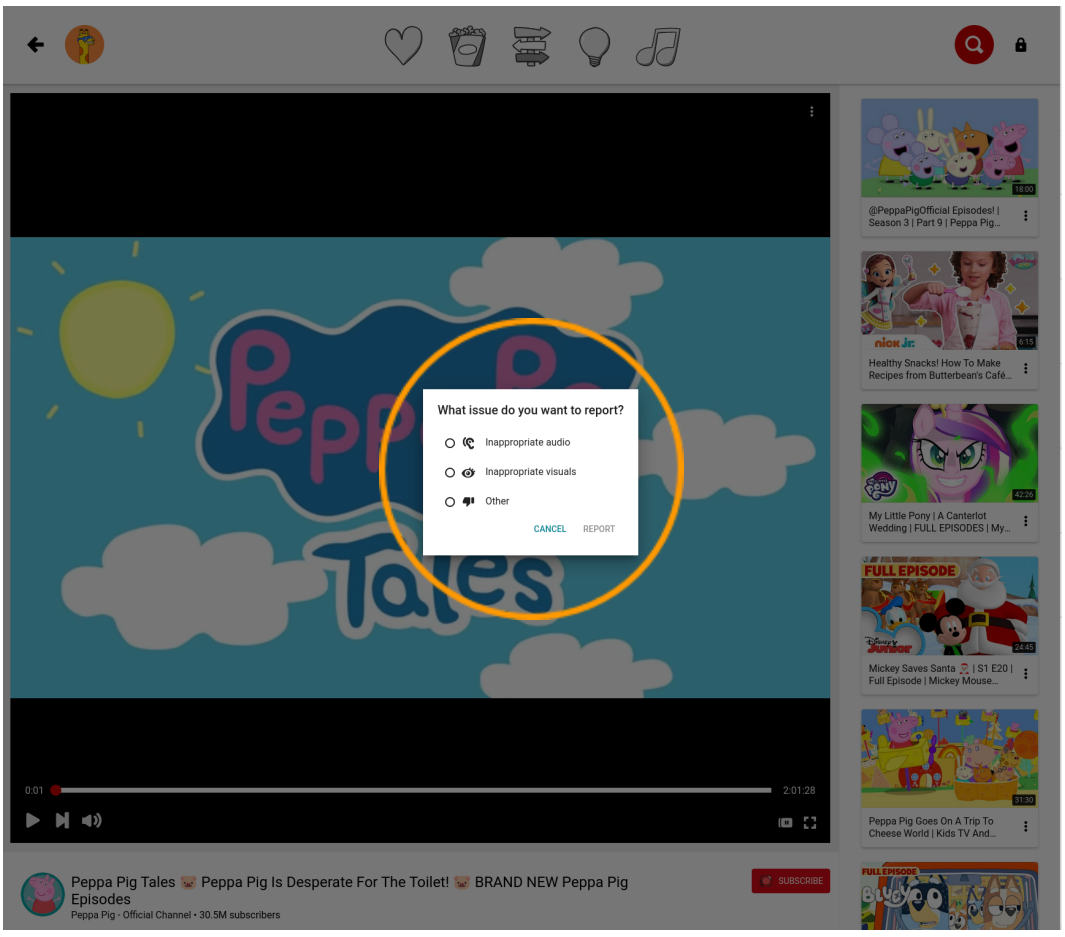
Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?



Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

Our response to Question 8 of Ofcom's first phase call for evidence provides more detail on our reporting processes across YouTube, Search, and other Google services.

Complaints mechanisms

We are focused on providing easy and accessible complaints mechanisms for all users, including for children. Our response to Question 9 of Ofcom's first phase call for evidence provides more detail on our mechanisms for responding to complaints or "appeals" to decisions we make to restrict or remove content that has been generated, shared or uploaded on YouTube by users, and the limited cases in which we allow appeals against the delisting of content on Search.

Response to user reports

Reviewing user reports requires careful and nuanced judgments by human reviewers. The context of content can be very important. The same piece of content with a different context can indicate very different intent from the user. For example, there is a clear difference between content intended to shock or disturb a viewer, potentially causing them harm, and content which documents real-world events where there is a public interest in keeping that content available. Our response to Question 10 of Ofcom's first phase call for evidence provides more detail on our processes of responding to user reports on YouTube and Search. Our response to Question 8 of Ofcom's first phase call for evidence outlines how user flags on YouTube can have very low accuracy rates, especially flags made in bad faith or based on a dislike of a particular video.

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)? **Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[5]** **Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?** **Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?** **Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?**

Design and operation of the service, including functionalities and algorithms

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

Is this a confidential response? No

Designing our services to provide a safe and positive online experience for children

Google believes there is no one-size-fits-all approach to designing services in a way that minimises the risk of harm. Different functionalities work well on different services, for different types of content. Different approaches and design features need to be used across our services, depending on the nature of the content on that service and its functionalities.

We employ a safety by design approach, incorporating it into the core aspects of each of our services' functionalities, including in the design of algorithms, such as recommendation algorithms. We design all of our services to mitigate the risk of harm to all users, and in particular to children. Parental controls are an important part of our approach. We include more detail on our approach to parental controls in our response to Questions 2 - 5.

As an example, in developing our policies, tools and features around self-harm or suicide queries or content, we consult with both internal and external experts in psychology, mental health, and related areas. These include not only academics and clinicians, but also practitioners who provide direct services to vulnerable populations. We know how important it is to increase awareness around help-seeking behaviours, while decreasing risk-taking and reducing stigma. This issue is complex and requires highly specialised expertise, which is why we have a dedicated Health team, led by Google's Chief Health Officer, with whom we work closely to inform our product design.

Beyond this overarching approach, we also use specific tools on each of our products to protect all users, including children.

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

- **Search:**
  - Within Google's population of users, there are some for whom it is particularly critical that we get our results right: users in crisis. Some of our users turn to Google as their first or last resource after going through a traumatic event. For these reasons, we've done work to refine our systems to help improve the visibility of authoritative information, such as national hotlines and text services, in search results for queries that indicate a high intent of self-harm or suicide. When users in the UK express urgent intent around suicide, a feature will appear at the top of their Search results page. This feature surfaces phone numbers of the Samaritans that support users in "SOS" situations, free of charge. This suicide hotline feature is Google's approach to connecting vulnerable users facing imminent harm with helpful and free resources immediately. We also ensure that support charities', NHS, or other authoritative websites appear at the top of the list of results.
  - Independent academics have studied our results for suicide-related queries, and published research that found that Google performs better than other search engines when it comes to handling suicide-seeking queries. They found that other major search engines returned more harmful URLs when compared to Google, and they were also less likely to display help messaging at the top of the search results page.
  - We deploy extensive policies for search features where Google curates content, which include knowledge panels, "top stories" carousels, enhancements to web listings, predictive and refinement features (such as auto-complete), and results and features spoken aloud. These policies don't apply to organic web results. For example, we do not allow the following categories of content in our search features:
    - Dangerous content
    - Deceptive practices

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

        - Harassing content
        - Hateful content
        - Manipulated media
        - Medical content
        - Regulated goods
        - Sexually explicit content
        - Terrorist content
- **YouTube**:
    - YouTube regularly updates its family product experiences and policies in consultation with experts in children's media, child development, digital learning and citizenship from a range of academic, non-profit and clinical backgrounds. This advisory committee is a collection of independent experts that weigh in on products, policies and services that we offer to young people and families. Together, YouTube and our advisory committee are dedicated to fostering a safe, high-quality, helpful platform that enriches the lives of youth and families around the world.
    - Our answer to Questions 2 - 5 provides more information on the controls we offer to parents to customise their child's experience as part of our separate YouTube Kids service, tailored to provide an age appropriate experience for children aged under 13 years old.
    - Our answer to Questions 2 - 5 provides more information on the ability we offer to parents to supervise and customise the experience of tweens and teens on the main YouTube service as part of YouTube supervised experiences.
    - As part of our approach to mitigating the risk of harm to children, users in the UK aged under 13 years old cannot upload any content to YouTube, nor can they livestream or comment on videos. This includes users in YouTube Kids and in YouTube's Supervised Experiences.

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

- With respect to teenagers, our approach protects their right to express themselves on the platform, while ensuring that they can access the service with appropriate safeguards. For users aged 13-17 uploading content, the default is set to the most private sharing setting, where videos can only be seen by the uploader and users they choose to make the content available to. We display a teen-friendly privacy prompt when the minor changes this setting, making a video public, stating "Remember that anyone can see public videos". The first time that a user aged 13-17 makes a public comment on the platform, we also display a similar prompt.
- For livestreaming, the livestream visibility default is set to the most private option available, which is "unlisted" on mobile and "private" on web. Unlisted means that only viewers that have the specific link to the livestream would be able to view it.
- We have introduced a number of features to promote Digital Wellbeing and are working hard to ensure we understand how we can best protect our users by carrying out research and updating our products. This includes turning reminders to take a break and bedtime reminders to on by default for these users.

- **Off-platform**
  - We have established long-standing partnerships in the UK aimed at strengthening media literacy for children and young people. Our response to Question 20 of Ofcom's first phase call for evidence provides more detail on how these initiatives aim to support the safety and wellbeing of children and young people.

Algorithmic design

Algorithms are an integral part of how our services function and meet the needs of our users. We ensure that they are designed to prioritise access to the most helpful information on our services. We also use algorithms to help detect content that may be in violation of our policies. We acknowledge that, without appropriate risk assessment and mitigation, the deployment of

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

algorithms by online services can present the risk of exposing children to harm. We therefore operate extensive and sophisticated processes of assessing the risks presented by algorithms and implementing safeguards to mitigate those risks.

Google has developed cross-product artificial intelligence principles, which set out our commitment to developing technology responsibly. These include the principle that our algorithms are built and tested for safety. We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm.

- **Search:**
    - We use automated systems to deliver the most relevant and reliable information. These systems consider many factors, including the words in a query, the content of pages, the expertise of sources, and the user's language. We provide more information on how Search works in our publicly available guide. We have a rigorous testing process to ensure that our automated systems return high quality results and mitigate the risk of harm to children using Search. Every change to Search goes through a launch process, including testing and review, before it's approved. Our response to Questions 4 and 21 of Ofcom's first phase call for evidence provides more detail on the rigorous process we follow to evaluate changes to our search algorithms.
- **YouTube**:
    - Recommendations on YouTube are designed to minimise the chances that all users, including children, will see problematic content. We prominently surface high-quality content in recommendations, using classifiers to identify whether a video is "authoritative". These classifications rely on human evaluators who assess the quality of information in each channel or video. Our response to Question 21

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

of Ofcom's first phase call for evidence provides more detail on the effectiveness of the safeguards we have implemented for YouTube's recommender system.

- For YouTube Kids, specifically, we have published a set of principles around low quality content that clearly demonstrate themes that should be avoided in content creation. This content doesn't violate our Community Guidelines, but includes things like heavily commercial content, content that encourages negative behaviours and sensational content. This content is less likely to be recommended by our recommendation algorithms on YouTube Kids.
- For users that declare to be under 18 when they create their Google Account, autoplay on YouTube has been turned off by default on all devices. Our autoplay feature is set to default off in YouTube Kids and the YouTube Supervised Experience, and we offer additional parental controls for this feature in the YouTube Kids app, including the ability for a parent to choose a "locked" default autoplay setting so their children cannot change this control. We have now also changed our settings so that for all new users, we first ask the user to decide if they would like autoplay on or off, instead of having it as default on.
- We also give users control over their recommendations through YouTube settings in several ways:
  - Users can view, pause, edit, or clear their watch history at any time through the YouTube history settings.
  - Users can also clear their search history, remove individual search entries from search suggestions, or pause search history using the YouTube History settings.
  - In-product controls enable users to remove recommended content - including videos and channels - from their Home pages and Watch Next.
  - Users can disable autoplay in their setting or on any video watch page.

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

Signed-in users can also choose to have their YouTube search and watch history deleted automatically after a certain period of time through their Google "My Account" settings.

Moderation

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Question 23: What training and support is or should be provided to moderators?

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 a confidential response? No

We agree that there can always be improvements made in the development and enforcement of content moderation.

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Question 23: What training and support is or should be provided to moderators?

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

Across Search and YouTube, we are constantly working to improve and evolve, introducing new policy changes, hiring new people dedicated to safety policy, and continuing to invest in technology to help us deliver greater protection for children.

We are also mindful of the need to avoid unduly restricting user activity. With that in mind, we welcome the Bill's protections for freedom of expression, including for children and young people, as well as Ofcom's interest in the subject and in ensuring the Bill's protections for freedom of expression are given practical consideration. Our response to Question 11 of Ofcom's first phase call for evidence provides broader reflections on how the regulatory framework can work best to drive improvements to user safety and to the protection of users' rights.

Automated content moderation

We have long invested in the most effective automated systems for protecting all users from harmful content and have developed effective automated detection tools, which we have opened up to the wider industry.

Our response to Question 12 of Ofcom's first phase call for evidence provides more information on the automated systems we have in place for CSAM. This is the area in which our solutions are the most advanced because the legal framework is clear (CSAM is nearly universally illegal).

Machines also can help to flag content that is harmful to children, such as content about self-harm and eating disorders, but these categories are highly dependent on context and require human review to make nuanced and accurate decisions. Mandating an approach that relied excessively on automated tools to identify and remove this material would be likely to lead to the removal of legitimate self-help material, such as expert information about seeking help for disordered eating and body-positive content which is proven to have a positive impact on self-image and social comparison. It is important that children suffering from issues such as eating disorders and self-harm – including the 10,000 UK children and young people started NHS disorder treatment

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Question 23: What training and support is or should be provided to moderators?

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

between April and December 2021[6] – are not prevented from accessing legitimate material that can help them to explore and address these issues. We would be happy to discuss with Ofcom further the nuances in our policies on this issue and the limitations of an approach that overly relies on automation.

As an example of what can happen when an over-reliance on automated removal is prioritised over careful human review, in Q2 2020, as COVID-19 lockdowns meant that fewer human content moderators were able to work, YouTube depended more heavily on automated technology to remove content violating our policies. The number of appeals by users of content removal decisions doubled as compared to Q1: 50% of appeals resulted in reinstatement in Q2, compared with less than 25% in Q1.

- **YouTube**
  - The majority of content takedowns are made under our Community Guidelines. As our response to Questions 6 - 7 explains, we are awaiting further detail on the definitions of the different categories of content so far identified as indicative 'primary priority' and 'priority' content and how they may align with our Guidelines.
  - We have developed a number of classifiers aimed at proactively detecting these types of content. According to our last reported figures from Q3 2022, over 94% of video removals were initially flagged by our automated systems, with almost one third removed before they had a single viewing and over two thirds removed before they had more than 10 viewings. However, careful human review is necessary to assess much of the content that is first flagged by automated systems, to avoid the large-scale removal of legitimate content and ensure the video's context is taken into account.
  - In September 2020, we announced that we would be expanding the application of our age-restriction policies (see our response to Questions 6 - 7 for more detail on

[6] https://www.england.nhs.uk/2022/03/nhs-treating-record-number-of-young-people-for-eating-disorders/.

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Question 23: What training and support is or should be provided to moderators?

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

these policies). While creators and our Trust and Safety content moderator teams had always been able to age-restrict content, from September 2020 onwards we developed and adapted our machine-learning technology to help us automatically apply age-restrictions to content.

- **Search**
    - Automation is Google Search's most powerful tool in dealing with policy-violating content. Google Search's ranking systems are designed to elevate the most relevant and reliable information possible and reduce the spread of low quality content, e.g. protect users from being met with unexpected shocking content.
    - These systems are continuously improved through rigorous testing. Google Search works with external Search Quality Raters to measure the quality of Search results on an ongoing basis. Raters assess how well content fulfils a search request, and evaluate the quality of results based on a set of guidelines. Under these guidelines, raters are instructed to assign the lowest rating to pages that potentially are harmful to users, harmful to groups, misleading, untrustworthy, and spammy. Through these and other efforts, ranking allows us to elevate the relevant information that our algorithms determine is the most authoritative and trustworthy above information that may be less reliable or problematic. To ensure a consistent and transparent approach, Google makes the Search quality rater guidelines publicly available. SafeSearch ranking filters explicit content, so that these explicit sites don't surprise young users - with options for parents and schools to lock SafeSearch on for supervised minors.
    - At the heart of Google's automated systems is language understanding, helping Google steer clear of unwanted results that may be shocking out of context. The design of these systems is the greatest defence against low-quality content, and is work that Google has been investing in for many years.

Human content moderators

On YouTube, our human moderators may decide whether to:

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Question 23: What training and support is or should be provided to moderators?

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

- remove content - where it violates our Community Guidelines;
- restrict access to the content (for example, based on age where the content is not appropriate for all audiences); or
- leave the content live when their judgement is that it doesn't violate our guidelines.

In general, where content has been flagged by automated systems, we refer to human moderators to make these judgments where a more nuanced decision is required. In other cases, content identified by machines will be automatically removed - for example, if it is content we have already reviewed previously.

Our moderators receive regular training, including to identify content that is harmful to children, in line with our policies on harmful or dangerous content, harassment and cyberbullying, and other relevant policies outlined in our response to Questions 6-7 of this document. This training is updated when new policies are introduced or new abuses of our services are identified.

We have a rigorous quality assurance mechanism for moderators where we assess moderation decisions for accuracy (whether the Community Guidelines are correctly interpreted) and consistency. We draw samples regularly from the work done by our moderators, which are then re-reviewed by our specialised teams of quality analysts. Our response to Question 13 of Ofcom's first phase call for evidence provides more information on our processes for training, support, and safeguarding of content moderators, in addition to our industry-leading research and technological innovation in the field of content moderation.

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what? **Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?** **Question 22: How are human moderators used to identify and assess content that is harmful to children?** **Question 23: What training and support is or should be provided to moderators?** **Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?**

Other mitigations to protect children

Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access? **Question 26: What other mitigations do services currently have to protect children from harmful content?** **Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?** **Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?**
Is this a confidential response? No

Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access?

Question 26: What other mitigations do services currently have to protect children from harmful content?

Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?

Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?

Removing content that is harmful to children

Our response to Questions 6-7 of this document outlines our content policies for YouTube and Google Search, and the extent to which they map on to the indicative categories of primary priority and priority content.

Content that is harmful to children and which contravenes our content policies is removed from our services where:

- we are able to find it, for example, using one of the tools that we deploy;
- we are able to verify that it violates one of our policies (or we are required to remove it pursuant to local law, if it is a legal removal); and
- where we have the justification to remove (or, more accurately in the case of Search, delist or deprioritise) it.

Our response to Question 15 of Ofcom's first phase call for evidence provides more information on the circumstances in which content is removed on YouTube and Search. Our response to Questions 20-24 in this document provides more information on the automated and human content moderation systems we have in place to identify and assess harmful content that may be in contravention of our policies.

Other mitigations and measures

Our response to Questions 16 - 19 sets out our approach to implementing safeguards that mitigate the risk to children and young people on our services. Measures and technology to mitigate risk to children are continually evolving. We welcome the opportunity to brief Ofcom in the future as we introduce new tools and measures to mitigate risk and harm to children.