Facebook response to Ofcom's Questionnaire Regarding Video-sharing platforms.

Intro

Facebook's mission is to give people the power to build communities and bring the world closer together. Building on this mission, Facebook's platforms and products are designed to help people connect with family, friends and communities.

The main objective of Facebook's platforms and products is allowing interaction amongst users within their communities. Users are free to achieve such interaction using the type of content which is most meaningful and appropriate for them, including text, photos, stickers and video. Video-sharing on our platform is therefore one aspect of a broad and inclusive multi-media platform.

As part of our mission, we want people to feel safe during their interactions on our platform. The conversations that happen on Facebook reflect the diversity of our community and, while the protection of the freedom of expression is part of Facebook's core values, we remain fully committed to making people feel safe when using Facebook. That is why we devote significant resources to handle abusive content on our platform. Facebook remains fully devoted to keeping European citizens, and minors in particular, subject to the highest possible level of protection when using our services.

Facebook has its European headquarters in Ireland via Facebook Ireland Limited, and as such we do anticipate that it will come under jurisdiction of Ireland for the purposes of the AVMSD regulation of VSPs. We do anticipate that some of the Facebook suite of services will be considered to meet the definition of a VSP by the Irish regulator. Which elements of the Facebook platform will be considered to fall under the VSP definition will be determined by the Irish transposition of the AVMSD and the interpretation of the VSP definition by the Irish regulator, taking into consideration the European Commission guidance on the interpretation of the ‘‘essential functionality' criterion. Given that the new Irish Media Commission has not been established yet, it is premature to make assumptions in relation to which aspects of Facebook services will be considered a VSP. We have therefore responded to this questionnaire providing general information about the existing policies and tools for the Facebook app.

Facebook seeks open dialogue with all regulators, and so we appreciate the opportunity to contribute to Ofcom's call for evidence on video-sharing platform regulation.
Facebook strongly supports a harmonized implementation of the AVMSD VSP rules, alongside the 'country-of-origin' principle enshrined in the Directive. A harmonized interpretation of the AVMSD principles would contribute to European citizens benefitting from the same level of protection across the EU.

Consistency is also crucial across different regulated areas at a national level and across Member States. Overlap with initiatives to regulate content online in different jurisdictions (outside AVMSD) introduces regulatory uncertainty and risk of the same service having to comply with competing/inconsistent requirements. As an example, Facebook's video-sharing platform services (as determined under the Irish implementation of AVMSD) may also be regulated services under new German laws (e.g. NetzDG and its revision) that overlap with or are very closely related to AVMSD (or another category of regulated service in another Member State). The amendments to the NetzDG include this obligation for services that are both social networks and video-sharing platform services. It is already unclear what the situation will be if Facebook's video-sharing platform services (as determined under the Irish implementation of AVMSD) are also deemed to be social networks under German law (or another category of regulated service in another Member State) as the requirements for each legal framework are not harmonised. This is a difficult question which demonstrates how important a coherent legal framework is. We welcome that Ofcom is considering the application of VSP regulations within the context of the upcoming Online Harms Bill regulations to ensure consistency and continuity for industry.

The need to comply with potentially diverging requirements would increase complexity and reduce the effectiveness of the overall system, undermining the objectives that they are intended to achieve.

Digital services are highly innovative and fast-changing. Equally, the challenges that emerge in the digital space can be equally as fast (for example new behaviours by bad actors). Requirements which are too prescriptive or detailed would reduce the necessary flexibility to adapt to the emerging challenges. There is a role in this sense for coregulatory models, as encouraged by the AVMSD, which creates dialogue channels for industry and regulators. Co-regulatory regimes in areas which are still developing, would provide the benefit of being able to be responsive and adaptable (for example in relation to mechanisms to protect minors).

Community Standards

Facebook is committed to making its service a place that is as safe and respectful as possible for all users to share and connect with others. Our users' safety is one of Facebook's core principles and one which it takes very seriously. It is necessary and important to take active steps to tackle abusive, harmful and illegal behaviour and content, and we recognise that we have a responsibility to do this.

This commitment begins with the Terms of Service, which all users must accept and which prohibits users from doing or sharing anything that is unlawful, misleading, or fraudulent or that infringes or breaches someone else's rights.

In addition to our Terms, we also have our Community Standards, the global set of policies that outlines what is and is not allowed on Facebook. Our Community Standards are publicly

available on our website and apply to everyone, all around the world, and to all types of content.

Facebook employs a team of more than 35,000 people who are devoted to safety and security globally, and employ a combination of technological and human review to remove abusive content violating our content policies. Facebook provides comprehensive reporting mechanisms, allowing users to flag alleged violating content, including photos, videos and ads.

Facebook's Community Standards is the basic policy governing what is allowed or not on our platform. They are developed for a global user base - they apply to everyone, all around the world, and to all types of content. Community Standards prohibit a significant amount of content that may affect minors, hate speech and illegal content.

The Community Standards are constantly revised based on feedback from our community and the advice of external experts in fields such as technology, public safety and human rights. This consultative and dynamic model aims at ensuring external voices are valued and that our policies evolve to respond to emerging challenges happening on our platform.

Commercial communications

Facebook has specific policies in relation to advertising on the platform. Our Advertising Policies provide guidance on what types of ad content are allowed on the platform. These policies complement our Community Standards, and set out additional rules governing paid ads on our platform.

According to our Ads Policies, ads must not violate our Community Standards and, in addition to that, advertising of some types of specific content is prohibited or restricted on the platform. 'Prohibited' content includes 30 categories such as advertising of tobacco products or of illegal products or services. As per 'restricted' content, ads that for example promote or reference alcohol must comply with all applicable local laws, required or established industry codes, guidelines, licenses and approvals.

Some of the policies include age-limitations, which affects content categories such as alcohol, adult products or services (which are age-gated to users over 18 years old). Other types of content which might be inappropriate for minors are just prohibited, such adult content (e.g. nudity). In addition, ads targeted to minors must not promote products, services, or content that are inappropriate, illegal, or unsafe, or that exploit, mislead, or exert undue pressure on the age groups targeted.

When advertisers place an order, each ad is reviewed against these policies. The ad review system relies primarily on automated tools, including machine learning classifiers that are trained to identify signals or patterns, to check for certain types of violations of these policies. This review happens automatically before ads begin running. The system is complemented by the reporting system (as explained in response to Question 2), which allows users to report ads that they think violate our content policies.

Branded content (i.e. commercial communications not monetized by Facebook) are managed through the Branded Content Tool. Branded Content is produced by a *publisher or creator*[1] for payment by a *business partner*[2], where the partner influences the content or is featured in it.

The Branded Content Policies govern the types of branded content which is allowed or not on the platform. Pages Policies require compliance with the Branded Content Policies.

As with our Ads Policies, the Branded Content Policies prohibit the promotion of certain types of content, which includes, amongst others, tobacco products, drugs or drug related products, illegal products or services or adult products. Some sensitive content can only be promoted with restrictions, which includes restricting who can see the post based on age. Age-gating applies to e.g. alcohol, subscription services, financial and insurance services or cosmetic procedures and weight loss.

All monetized advertisements uploaded on the platform are reviewed by Facebook and are automatically declared as Ads when published.

For Branded Content, creators and publishers should use the Branded Content tool to tag their business partner when the content is posted for payment. A paid partnership label is applied to each ad post managed through the Branded Content tool and highlighted. Facebook recommends using this tool or labelling a post appropriately when its systems detect that a post may be of commercial nature.

Pages policies require *compliance* with the Branded Content Policies and stats that "*Your Page must include all necessary disclosures to* people *using Facebook, such as any disclosures needed to indicate the commercial nature of content posted by you*."

Appeals processes

Any system that operates at scale and for a global user base will make errors. For this reason, Facebook has set up an appeal process for content removal. If a piece of content has been removed because it's found to violate our Community Standards, the user who uploaded it will be notified and given the option to request additional review.

Any system that operates at scale and for a global user base will make errors. For this reason, in April 2018, we launched appeals, globally, for content that was removed for violating our Community Standards for nudity or sexual activity, hate speech and violence. We've extended this option so that re-review is now available for additional content areas, including dangerous organisations and individuals (which includes our policies on terrorist propaganda), bullying and harassment, regulated goods, and spam.

We are continuing to roll out re-review for additional types of Community Standards violations,

[1] Creators include celebrities, influencers or public figures that post branded content. Publishers include media companies and entities that post branded content.
[2] Business partners include brands, advertisers, marketers or sponsors that sponsor branded content.

but there are some violation types – for example, severe safety policy violations – for which we don't offer re-review. We are also beginning to provide appeals not just for content that we took action on, but also for content that was reported but not acted on.

Here's how the process works for many of our appeals:

- If your photo, video or post has been removed for violating our Community Standards, you will be given the option to “Request Review” on both mobile and desktop.
- Appeals are reviewed by our Community Operations team typically within 24 hours.
- If we've made a mistake, the content will be restored and we will notify the person who requested the appeal.
- For certain high severity content violations, Facebook may not allow users to request another review.

Independent Oversight Board

We have created an independent review body called the “Oversight Board”. This board will help Facebook answer some of the most difficult questions around freedom of expression online: what to take down, what to leave up and why. The board will use its independent judgment to support people's right to free expression and ensure that those rights are being adequately respected. The board's decisions to uphold or reverse Facebook's content decisions will be binding, meaning that Facebook will have to implement them, unless doing so could violate the law.

When fully staffed, the board will consist of 40 members from around the world that represent a diverse set of disciplines and backgrounds. These members will be empowered to select content cases for review and to uphold or reverse Facebook's content decisions. The board is not designed to be a simple extension of Facebook's existing content review process. Rather, it will review a selected number of highly emblematic cases and determine if decisions were made in accordance with Facebook's stated values and policies. These decisions will be publicly available for everyone to see.

In addition to rendering binding decisions, the board will be able to recommend changes to Facebook’s content policies through official “policy advisory statements.” In accordance with the Oversight Board’s bylaws, Facebook must consider the board’s recommendation and publicly disclose whether we took action in accordance with the recommendation.

Reporting Mechanisms

Facebook is committed to making its service a place that is as safe and respectful as possible for all users to share and connect with others. Our users’ safety is one of Facebook’s core principles and one which it takes very seriously. It is necessary and important to take active steps to tackle abusive, harmful and illegal behaviour and content, and we recognise that we have a responsibility to do this.

This commitment begins with the Terms of Service, which all users must accept and which prohibits users from doing or sharing anything that is unlawful, misleading, or fraudulent or that infringes or breaches someone else's rights.

In addition to our Terms, we also have our Community Standards, the global set of policies that outlines what is and is not allowed on Facebook. Our Community Standards are publicly available on our website and apply to everyone, all around the world, and to all types of content.

Facebook employs a team of more than 35, 000 people who are devoted to safety and security globally, and employ a combination of technological and human review work together to remove abusive content violating our content policies. Facebook provides comprehensive reporting mechanisms, allowing users to flag alleged violating content, including photos, videos and ads.

Users can report video content if they think it violates our Community Standards.

Facebook provides comprehensive reporting mechanisms, as listed in e.g. the Reporting Abuse or How to report Things on Facebook sections in our Help Center. The user can report any type of content: photos, video, pages, profiles, ads, etc. Reporting works through a click on the three dots next to each piece of content.

We also provide users with the ability to submit a request to remove content they believe is unlawful according to local law through our Legal Removal Request form.

The reporting tools, which entail enforcement of the Community Standards, have implied the commitment of significant time and resources to develop from the ground up — and we have continued to devote significant resources to scaling and improving these important tools. At the moment, more than 35,000 people work on security and safety at Facebook, 15,000 of whom review content. This includes content reviewers who speak almost every language widely used in the world and who come from diverse backgrounds, to reflect the diversity of our community and to bring a wide array of professional experiences to the role. They consist of native language speakers who also demonstrate an in-depth understanding of local context and issues. Collectively, they review around 2 million pieces of content every day.

High level information regarding what effect has been given to content reports is provided to users on the 'Report Something' page (see What happens when I report something to Facebook? Does the person I report get notified? and Can I check the status of something I've reported to Facebook or cancel a report? sections).

Users have the option to *check* the status of their report from the Support Inbox. Only the user can see their Support Inbox. From this page, users can learn more about our policies, cancel a report, and see when we take action and the decision we made.

Protection of Minors

Keeping young people as safe as possible online is and has been a top priority for Facebook. We have in place a very diverse set of policies and measures aimed specifically to protect

minors from harm on our platform. Given the multi-media nature of our platform, the measures to protect minors cover all types of content, not just video.

Facebook's Community Standards are developed to ensure all users have as safe and enjoyable experience as possible on our platform. They prohibit an extensive list of behaviours/types of content considered to be inappropriate, illegal or harmful, including categories especially sensitive for minors, such as bullying or graphic violence.

.

***

Although our Community Standards are aimed at protecting all of our users, some policies are especially important to minors, such as those related to nudity, bullying or graphic violence.

As a general rule, content which may impair the physical, mental or moral development of minors should not appear or be accessible to minors on our platform. There are a variety of measures in place to ensure that young people who use our services have age-appropriate experiences. We set up an age limitation of at least 13 years old for people to be able to create a Facebook account. In addition, minors between 13 and 17 years old generally have a more limited experience on Facebook when it comes to the features they have access to, who they share and connect with, and the content they see (including ads).

For example, some types of content which might be inappropriate for minors are prohibited, such as tobacco, adult content (e.g. nudity) or illegal drugs. Other content is age-gated, meaning it will not be accessible to 13-18 year olds. For instance, although we allow some types of graphic content (with some limitations) to help people raise awareness about issues, this content may still be inappropriate for minors even if it does not violate our policies. If this content is reported to us, we age-gate it and we put a warning screen on it, letting those who can see it (18+) know that they are about to watch graphic material.

In addition, some of our product experiences are limited to those above 18, such as marketplace and blood donations, and certain services and products (such as alcohol, nicotine products or financial services) cannot be advertised to those under the age of 18.

** Age verification

Collection of age at sign-up is a fundamental step to ensure implementation of our policies based on age restrictions across the platform. These include: (i) minimum age requirements to register to Facebook services, (ii) age-gating of content limited to adults, and (iii) implementation of GDPR requirement on parental consent.

Facebook has set up an age limit, requiring everyone to be at least 13 years old before they can create an account (in some jurisdictions, this age limit may be higher). It violates our terms of service to provide a false age when creating an account.

We do not allow minors under 13 on our platform. Facebook undertakes a series of steps to prevent users under 13 from signing up for Facebook services and detect underage users.

- All users are required to enter their age as part of our multi-step registration experience.
- On Facebook we've added barriers to the registration flow that prevent minors from editing the initial birth date that they enter. Users are also prevented from attempting to register more than twice in a single session and enforce a waiting period before they can try again.
- We are working on changes to make it more challenging for people to give us an inaccurate age at sign up. For instance, we will not have a default age at or above the minimum required age to access our services.
- Users are encouraged to report potential underage accounts. Facebook has a standardized process for reviewing and disabling accounts if we think they are underage, either because they are reported or if we review them for other reasons.

Facebook has taken a variety of steps designed to ensure users are the appropriate age, and are committed to exploring additional ways we can further minimise the number of young people who see content that is inappropriate or access our platforms inappropriately.

Tools to check age are still a developing area. Age verification is a complex and industry-wide challenge requiring thoughtful solutions that protect children's safety and privacy without unduly restricting everyone's ability to access information, express themselves, and build community online or unduly invading their privacy.

In light of the ecosystem-wide nature of this problem, we believe that providers and users alike would benefit from a scalable, centralized solution that all apps and services could rely on to enforce minimum age restrictions. Age verification will be challenging to effectuate on a per-app basis, regardless of the precise mechanism for verifying age. A centralized approach would avoid each individual provider having to come up with their own unique scheme to verify age, an undertaking that will prove highly challenging for companies regardless of size, resources, and technical expertise. Users would also face significant challenges with such an ad hoc approach, as they would be forced to navigate cumbersome mechanisms for scores of apps and websites with potentially inconsistent requirements and varying levels of privacy and security measures in place. In contrast, centralizing age verification at a platform or device level would ensure much-needed consistency across apps and websites, and enable a more predictable and secure experience for users as well as minimizing the amount of data that most apps had to collect.

The issue of age verification should not be considered in isolation or viewed as a cure-all to the question of protecting children online, but viewed as one of many responses that comprise a holistic approach to the protection of minors online.

We welcome the opportunity to work with the government, civil society and others in industry on alternative solutions.

** Parental controls

As described above, most vulnerable minors (in terms of age) are not allowed on our platform. We require everyone to be at least 13 years old to register for Facebook services.

In addition to this age restriction, Facebook has put in place a number of measures to ensure minors have an age-appropriate experience on our platform.

Under GDPR, in some EU countries on Facebook, 13 to 15 year olds see a less personalized version of Facebook with restricted sharing and less relevant ads until they get permission from a parent or guardian to use all aspects of Facebook.

More generally, minors between 13 and 17 years old have a more limited experience on Facebook when it comes to the features they have access to, who they share and connect with, and the content they see. For example:

- we prevent minors from receiving messages from unconnected adults and we protect sensitive information such as minors' contact info, school or birthday appearing to a public audience.
- we take steps to remind minors that they should only accept friend requests from people they know.
- location sharing is off for minors by default. When either an adult or minor turns on location sharing, we include a consistent indicator as a reminder that they're sharing their location.
- new minor users are automatically defaulted to post with 'friends' only.
- if a minor wants to share publicly, the first time they go to do so they must go to their settings to enable the option and we remind them about the meaning of posting publicly.
- we age-restrict some products (e.g. marketplace or blood donations) and content (e.g. graphic violence or tobacco ads)

In addition to the platform settings and restrictions, Facebook provides Tools for Parents and Educators, which include details on how to report underage children, how to remove a child's image, information that parents cannot monitor their child's activity on Facebook, reporting tools and information on privacy.

The Safety Centre includes a Parents portal and a Youth portal.

Considering the multimedia nature on Facebook, offering granular controls of video content for parents risk disproportionately increasing complexity without materially increasing the level of protection. Minors of most vulnerable age ranges (under 13) are not allowed on the platform and extensive measures are in place to protect minors of 13-17 years old from harmful content.

In line with Art 28b(3), which sets out that appropriate measures shall be determined "in light of (...) the rights and legitimate interests at stake, including those of (...) and the users having created or uploaded the content as well as the general public interest", any parental control system should avoid undue burden on the ability for children to exercise their rights online. Given that minors on our platform are aged 13 to 17 years old, undue access for parents of a child's account raises concerns as to the right of young users to access information, express themselves, and build community online.

Media Literacy

Facebook is committed to media literacy and has developed extensive resources to encourage users to understand our policies and make use of the safety tools available on our platform.

The Help Center contains comprehensive information on our content policies and the reporting system, including specific tips in relation to video content across the website.

The Safety Center is our site devoted to providing tools and specific resources to users on safety related topics. It provides:

- Parents Portal - with guidance for parents on starting important online safety conversations and setting parameters.
- Youth Portal, with resources made specifically for our young users.
- Digital Literacy Library, with lesson plans designed by experts to help young people develop skills needed to navigate the digital world, critically consume information and responsibly produce and share content.
- Online Well-being, with tools for suicide prevention
- Not Without My Consent, with information on how to respond to intimate images shared without permission
- Bullying Prevention Hub, a resource for teens, parents and educators seeking support and help for issues related to bullying and other conflicts.

In addition, we run ad-hoc campaigns and programs addressed to local communities, in many instances young users, which include digital skills initiatives in the areas of safety or well-being.

We continuously work to make our platform safe for our users, which includes the creation or resources and tools aimed at providing information to vulnerable users, such as minors or teens, or on issues of especial social impact, such as suicide prevention.

Community Standards Enforcement Report

Facebook publicly shares information on a regular basis through the Community Standards Enforcement Report (CSER), that includes information about content that is removed from the platform because it violates our Community Standards. This Community Standards Enforcement Report also includes information regarding content detected and removed with the help of AI tools. We also regularly share other transparency reports regarding take-down requests for illegal content in specific countries, as well as Intellectual Property violations.

We are looking at opening up our content moderation systems for external audit. We're reaching out to key stakeholders spanning government regulators, civil society, and the advertising industry to help us develop our approach.

External audit will begin with the harmful content metrics we provide in our Community Standards Enforcement Report (CSER). This detailed report shows how we are doing at removing content that violates our Community Standards. The audit, which will be done by an

independent auditor, will show that “we’re not grading our own homework”. We want to give people confidence that the numbers we are reporting around harmful content are accurate. This builds on the work of the Data Transparency Advisory Group (DTAG), who assessed Facebook’s methods of measuring and reporting on its Community Standards enforcement policies. In DTAG’s final report published last May, it was noted that Facebook’s approach and methodology were sound and reasonable.

Section B

As we stated at the start of this response, we anticipate that any Facebook services that are to be regulated in line with the AVMSD, will fall within the remit of the soon to be established Irish Media Commission. Given that the new Irish Media Commission has not been established yet, it is premature to make assumptions in relation to which Facebook aspects will be considered as a VSP. For that reason, we haven't answered in detail the questions in section B, rather we have indicated the aspects we consider most relevant for regulators to consider in implementation.

Mitigating risks for Freedom of Expression to free speech

Any regulation should recognise the need to balance the removal of harmful but legal content with the protection of freedom of expression and other fundamental rights. Holding intermediaries liable if they do not remove individual pieces of third-party content will necessarily lead to overblocking of content, as intermediaries will be wary of incurring penalties and/or fines. Regulation should also recognise that intermediaries face challenges when they do seek to remove harmful content pursuant to their policies. At an increasing rate Courts in a number of European Member States are regularly ordering Facebook to reinstate content that violates Facebook's Community Standards -- and would be considered by many to constitute harmful content. These "wrongful removal and restoration" decisions limit Facebook's ability to remove harmful content and keep its platforms safe, and regulations should acknowledge that intermediaries cannot be held liable if they in good faith remove content from their platform in an effort to combat harmful speech.

This is why we have created an independent review body called the "Oversight Board." It will help Facebook answer some of the most challenging questions around freedom of expression online: what to take down, what to leave up and why. The board will use its independent judgment to support people's right to free expression and ensure that those rights are being adequately respected. The board's decisions to uphold or reverse Facebook's content decisions will be binding, meaning that Facebook will have to implement them, unless doing so could violate the law.

When fully staffed, the board will consist of 40 members from around the world that represent a diverse set of disciplines and backgrounds. These members will be empowered to select content cases for review and to uphold or reverse Facebook's content decisions. The board is not designed to be a simple extension of Facebook's existing content review process. Rather, it will review a selected number of highly emblematic cases and determine if decisions were made in accordance with Facebook's stated values and policies. These decisions will be publicly available for everyone to see.

In addition to rendering binding decisions, the board will be able to recommend changes to Facebook's content policies through official "policy advisory statements." In accordance with the Oversight Board's bylaws, Facebook must consider the board's recommendation and publicly disclose whether we took action in accordance with the recommendation.

How to assess the effectiveness of measures

In order to give our community visibility into how we enforce policies, respond to data requests and protect intellectual property, while monitoring dynamics that limit access to Facebook products, we publish regular transparency reports and other information. This allows for users to have an understanding of how their systems are performing to give the community visibility and to monitor the dynamics of content, so that we can continually improve.

Facebook already shares regular transparency and enforcement reports, such as the Community Standards Enforcement Report (CSER)[3] detailing how much content we remove for violating certain of our policies, how much of that content was detected proactively by our automated tools, how much content was appealed when people believed we had made a mistake, and how many of those appeals were successful. Additionally, we regularly publish another report that includes metrics on the number and nature of legal requests we receive from governments and other entities around the world – including requests for data and requests to restrict access to content which they believe violates local law. In addition, Facebook also publishes an IP transparency report, which sets out the number of copyright, trademark and counterfeit reports we receive, the number of pieces of content removed based on those reports, and the overall action rate.

Measuring the right metrics that actually help answer the right questions is not a trivial endeavour and requires a massive understanding, detailed caveats and significant effort in terms of human resources, processing time and financial commitments. In a number of cases, measurement such as retrospective analysis may not always be feasible. **Hence, systematic transparency should always be preferred over ad-hoc requests for metrics.**

However, the type and nature of what information should be made available should not be so prescriptive or fixed as to predetermine the business model of the platform. Transparency should avoid being overly detailed with regard to automated systems or how enforcement measures operate as doing so could allow bad actors to circumvent the systems. Different types of services may require different levels of transparency, and such level needs to be proportionate according to the characteristics and nature of the provider.

Not all platforms are the same, and therefore will not be able to have the same type of systems in place nor information available. We believe it is essential to have a structure of protection regarding any information shared with authorities. Care needs to be taken regarding:

[3] If a piece of content is reported to us by local authorities because of a potential violation of the local law and it also violates our Community Standards and is subsequently removed, this metric will be included in the CSER report

- *information requests to ensure those serve specific and defined regulatory functions.* In particular with AI-related information there is a process of continual development and improvement and information in relation to this should be within that context and understanding.
- *designating who is or isn't a trusted researcher, and for what purpose they want information*. There are risks associated with information becoming available that would allow bad actors to 'game' the platform and exploit systems.
- *Ensuring respect for user privacy and data*, as well as for company confidential information.
- *the legal basis for the information sharing* – if the information includes personal data then the information sharing must be compatible with the GDPR; if the information contains proprietary or confidential material, appropriate measures must be put in place to safeguard those rights.
- *what is necessary and proportionate to the objective* – the information must service specific and defined objectives, underpinned by the need for good regulatory outcomes. Furthermore, the information shared must be proportionate to that objective.
- *what is technically feasible* – not all platforms are the same, and therefore will not be able to have the same type of systems in place nor information available. It is important to recognise that in addition the utility of information will change over time.
- *security* – careful consideration needs to be given to the security implications of sharing the data with third parties – be that the information security risks posed by providing access to a dataset by a third party; and/or the risk of bad actors gaming the platforms and exploit systems once information becomes more widely available.

Independent Audit

We are looking at opening up our content moderation systems for external audit. We're reaching out to key stakeholders spanning government regulators, civil society, and the advertising industry to help us develop our approach.

Currently, we are preparing an audit of the harmful content metrics we provide in our Community Standards Enforcement Report (CSER). This detailed report shows how we are doing at removing content that violates our Community Standards. The audit, which will be done by an independent auditor, will show that "we're not grading our own homework". We want to give people confidence that the numbers we are reporting around harmful content are accurate. We have currently put out a request for a provider. This builds on the work of the Data Transparency Advisory Group (DTAG), who assessed Facebook's methods of measuring and reporting on its Community Standards enforcement policies. In DTAG's final report published last May, it was noted that Facebook's approach and methodology were sound and reasonable.

How to determine what is practical and proportionate in relation to a platform

Regulations need to be proportionate, according to the characteristics and nature of the service, and level of risk the service poses. Regulatory compliance can be difficult or burdensome when the rules are fragmented, however we would suggest that any exemptions designed to help smaller businesses to flourish would need to be determined very carefully. For example, full exemptions of regulatory requirements to small to medium sized businesses could result in niche services that would have no prospect of significant market impact, becoming the repository of content that would otherwise be removed and subject to regulation on larger services. Small and medium-sized companies can have higher risks of exposure to illegal and/or harmful activities conducted by their users. The regulation therefore needs to be proportionate to the characteristics and nature of the service, and as well as the risk of the company providing a service.

While Facebook supports the implementation of AVMSD, we believe a homogenous one-size-fits-all approach is not a viable solution. Obligations should be proportionate in relation to the nature and characteristics of the service, and include appropriate safeguards to protect the privacy of users in the course of legitimate and lawful activities, and any requirements should be tailored to the variety of business models involved and developed in collaboration with stakeholders and platforms.

The implementation needs to be mindful of the practical difficulties that platforms face when implementing measures to comply with different applicable laws and especially if these would affect the user experience. For example, being overly prescriptive about the design of how compliance should look, disrupt business models and the ability to innovate products for users. This implementation should recognise this and allow flexibility for different types of services to address AVMSD requirements in a way that is best suited for those services' offerings and technical capabilities rather than requiring uniform measures across platforms.