Snap Group Limited

# Snap UK OSA Report

## Illegal Content Risk Assessment Summary



November 2026

**Introduction**

Snapchat is a regulated service under the UK Online Safety Act 2023 ("**OSA**"), provided to UK users by Snap Group Limited ("**Snap**"). Snap must therefore carry out an illegal content risk assessment ("**ICRA**") to assess the level of risk of users encountering illegal content on Snapchat, of the service being used for the commission or facilitation of a priority offence, and of harm to individuals by each of those. Ofcom has assessed that Snapchat is a Category 1 service provider which means that Snap must also publish a summary of the findings of its most recent ICRA. Snap's summary of its ICRA is below.

Snap takes its duties under the OSA extremely seriously. Each risk level reflects the assessed level of potential risk for OSA purposes on Snapchat at the time of our ICRA. Snap remains committed to continuing to invest in its mitigation measures and further decrease the risk of illegal content on Snapchat.

**Disclaimer**

Our UK OSA Risk Ratings have been prepared solely for UK OSA purposes in accordance with our risk assessment methodology, which is dictated by Ofcom's Risk Assessment Guidance and the specific requirements of the UK OSA only. Ofcom's Risk Assessment Guidance requires providers of user-to-user services to assess themselves as high risk if the severity of a priority offence is high even if the evidence indicates relatively low incidents of that harm considering the numbers of users and the volume of activity on the service. Services are also required to err on the side of caution where there is more limited evidence of the effectiveness of a service's mitigation measures (as is the case for all private messaging services where there are legal and societal limits on providers' ability to carry out proactive scanning for illegal content or activity). It is important to note that our UK OSA Risk Ratings do not consider the risk to the average UK user of Snapchat. Separately, and not for UK OSA purposes, we have considered the risk to the average UK user and assessed the risk to be low.

| Harm Type | Terrorism |
|---|---|
| Nature and severity of potential harm | Terrorism content may encourage or promote terrorism. Potential harms from this type of content include radicalisation, recruitment, grooming and exploitation by extremists.<br><br>External sources indicate young people may be more susceptible to radicalisation exploitation by extremists. Snapchat has a teenage user base in the UK. |
| Likelihood assessment | **Assessment: Low**<br>● *Ofcom Specific Risk Factors:* 11 out of 17 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated:* Account names may be written to attract like-minded users; teenage user base; targeted advertising systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Do not indicate that Snapchat is a primary platform for this content. |
| Impact assessment | **Assessment: Moderate**<br>● Potential for severe impact if encountered by individuals (albeit impact could vary considerably).<br>● Some scope for this content to impact many individuals, but data indicates that measures to reduce risk are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● Automated detection and human moderation.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Recommender systems are designed not to promote this content.<br>● Snap works with industry groups to advance counterterrorism efforts.<br>● Crisis response protocol for major incidents. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | CSEA |
|---|---|
| Nature and severity of potential harm | Child sexual exploitation and abuse (CSEA) content includes forcing or enticing a child or young person to take part in sexual activities, online or offline. This includes grooming, child sexual abuse material (CSAM), and sharing links to CSAM (CSAM URLS). The consequences of this type of content can be severe, including lasting psychological and emotional harm.<br><br>External sources indicate that children and young people face significant risks of grooming and sexual exploitation online. Ofcom has also assessed that women and girls are disproportionately affected by grooming and CSAM, although boys can also be targeted. |
| Likelihood assessment | **Assessment: High**<br>● *Ofcom Specific Risk Factors:*<br>○ Image-based CSAM: 7 out of 13 apply.<br>○ CSAM URLs: 7 out of 10 apply.<br>○ Grooming: 12 out of 16 apply.<br>● *Additional service characteristics that could increase the risk of harm if not effectively mitigated include*: Friend recommendations; group chat with non-friends; teenage user base; account names may be written to attract like-minded users.<br>● *Internal data:*<br>○ All types of CSEA content: Existence of this type of harm but very low violative view rates on public content and low violative view rates on private content, and rapid response time for enforcement.<br>● *External sources:*<br>○ Image-based CSAM: Presents a mixed picture regarding Snapchat specifically, but some reports allege social media (including Snapchat) are involved in CSAM offences.<br>○ CSAM URLs: None identified concerning CSAM URLs being shared on Snapchat.<br>○ Grooming: Various sources contend that grooming (and sextortion) on social media platforms (including Snapchat) is a significant risk for teenage users. Snap expresses caution when considering this data, as it is limited and may be subject to misrepresentation. |
| Impact assessment | **Assessment: High**<br>● Image-based CSAM and CSAM URLs: Although Snap's data indicates that image-based CSAM and CSAM URLs are relatively uncommon on Snapchat, coming into contact with such content can have serious potential for immediate and long term impacts for those directly involved in and accessing it.<br>● Grooming: Although Snap's data indicates that grooming is relatively uncommon on Snapchat, grooming is a uniquely severe harm with serious potential for immediate and long term impacts on victims, even in isolated incidents.<br>Given the potential severity of harm, Snap has assessed the potential impact |

| | |
|---|---|
| | as High. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Automated detection (including proactive hash matching for image-based CSAM and CSAM URL detection/blocking).<br>● Limits on public visibility of teen profiles for under 16s.<br>● Use of a 'trusted flagger' program (for example, the NCA) to enable reporting of such content through a dedicated high-priority channel. |
| Overall risk level | Snap assessed the risk associated with this content as **High**. |

| Harm Type | Hate |
|---|---|
| Nature and severity of potential harm | Hate offences content includes stirring up racial, religious, or sexual-orientation related hatred, as well as hateful content directed at other protected characteristics, if it amounts to harassment or a public order offence. Potential harms include cumulative emotional harm and a corrosive effect on democratic participation and community trust.<br><br>External sources indicate that younger users and minority groups can be particularly vulnerable to online hate. |
| Likelihood assessment | **Assessment: Low**<br>● *Ofcom Specific Risk Factors:* 7 out of 11 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Teenage user base*;* account names may be written to attract like-minded users.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of it being encountered.<br>● *External sources*: Do not indicate that Snapchat is used disproportionately as a platform for hate-related activity. |
| Impact assessment | **Assessment: Low**<br>● Potential for impactful cumulative emotional harm and its effect on democratic participation and community trust, however severity of this harm is likely to be low given the relatively low volume of this content on Snapchat. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit hateful content.<br>● Automated detection and human moderation.<br>● Coverage of hate crimes are not eligible for recommendation on Spotlight.<br>● User controls such as block, mute, report, leave group chats, disable comments on certain public content. |
| Overall risk level | Snap assessed the risk associated with this content as **Low**. |

| **Harm Type** | **Harassment, Stalking, Threats & Abuse** |
|---|---|
| Nature and severity of potential harm | Harassment, stalking, threats and abuse content includes threatening or abusive behaviour, threats of violence, and conduct that causes harassment, fear, alarm or distress. This includes threatening communications and epilepsy trolling content. These can cause psychological harm, distress, fear of violence and escalation to offline abuse. This harm may be serious.<br><br>External sources indicate that teenagers may be particularly vulnerable to online harassment and that technology (such as location tracking) can be used to facilitate stalking and other forms of abuse, which can disproportionately affect women and girls. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 10 out of 13 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Teenage user base; friend recommendations; group chat with non-friends.<br>● *Internal Data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Various sources contend that individuals are likely to face these types of harms on online platforms, with Snapchat cited in some media coverage. |
| Impact assessment | **Assessment: Moderate**<br>● Impact can vary, although there is potential for psychological harm and escalation to offline abuse. Users who remove themselves from online platforms due to this harm may feel isolated from their communities.<br>● Limits on the extent to share this content on Snapchat plus internal data indicates that measures to reduce impact are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content. Threatening and abusive behaviours are treated as a severe harm with strict consequences for violators.<br>● Automated detection and human moderation.<br>● Location sharing off by default, and users can only share location on Snap Map with friends.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● User controls such as block, mute, report, leave group chats, disable comments on certain public content.<br>● Use of 'trusted flagger' program enabling reports of such content through a dedicated high-priority channel (such as the Southwest Grid for Learning in the UK). |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**.<br><br>This rating also covers the related non-priority offences of threatening |

| | |
|---|---|
| | communications and epilepsy trolling which Snap considers to be closely linked to this type of harm. |

| Harm Type | Controlling & Coercive Behavior |
|---|---|
| Nature and severity of potential harm | Controlling or coercive behaviour content involves repeated or continuous abuse between people who are personally connected, including in intimate or family relationships. Potential harms include fear of violence, serious alarm or distress, isolation, and loss of independence, which can substantially affect day-to-day activities.<br><br>External sources contend that technology may be used to monitor, isolate and stalk victim-survivors, and highlight the disproportionate impact of this abuse on women and girls. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 7 out of 8 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated:* None identified.<br>● *Internal data:* Existence of this type of content on Snapchat, but rapid median turnaround time from detection to enforcement and low prevalence and viewership metrics for violent and dangerous behaviours (including controlling or coercive behaviours).<br>● *External sources:* Indicate that a considerable number of individuals, particularly women, have experienced abuse (including controlling and coercive behaviour) through social media platforms. However, external sources do not indicate that Snapchat is commonly used in abuse cases. |
| Impact assessment | **Assessment: Moderate**<br>● Impact for affected individuals can be significant, including serious psychological harm and potentially life-threatening abuse.<br>● External sources have not identified Snapchat as a primary platform through which such behaviour is commonly perpetrated. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Automated detection and human moderation.<br>● Location sharing off by default and users can only share location on Snap Map with friends.<br>● User controls such as block, mute, report, leave group chats, disable comments on certain public content. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Intimate Image Abuse |
|---|---|
| Nature and severity of potential harm | Intimate image abuse content includes non-consensually sharing or threatening to share intimate images. Potential harms caused by intimate image abuse include humiliation, alarm, psychological harm, trauma or distress.<br><br>External sources indicate that women and younger users are particularly vulnerable to this harm, while Ofcom notes that men are increasingly targeted by financially motivated sexual extortion (sextortion). |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 5 out of 9 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Teenage user base.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered. (Note: Data relating to private messaging and story communications is more limited).<br>● *External sources:* Various sources contend that intimate image abuse is prevalent online, although Snapchat is not expressly identified as a common platform for the dissemination of such images. |
| Impact assessment | **Assessment: Moderate**<br>● The impact of intimate image abuse can include significant psychological harm, trauma and distress.<br>● Rapid removal of detected content limits the ability for intimate images to circulate widely and the resulting impact. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● Automated detection (including hash-matching from the StopNCII.org data base) and human moderation.<br>● Use of 'trusted flagger' program to enable reporting of such content through dedicated high-priority channel (such as the Revenge Porn Helpline in the UK).<br>● User controls such as block, mute, report and leave group chats.<br>● Teen account protective measures, including limits on public visibility of teen profiles for under-16s.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Safety warnings. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Extreme Pornography |
|---|---|
| Nature and severity of potential harm | Extreme pornography content includes that produced principally for sexual arousal, and which depicts extreme or obscene behaviours.<br><br>Exposure to extreme pornography can cause harm including affecting perceptions of sexuality and gender roles, which can be particularly harmful if negatively skewed in teens. |
| Likelihood assessment | **Assessment: Low**<br>● *Ofcom Specific Risk Factors:* 3 out of 4 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Account names may be used to attract like-minded users; teenage user base; recommender systems on Spotlight and Discover; targeted advertising systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* No indication that Snapchat is used disproportionately as a platform for this type of content. |
| Impact assessment | **Assessment: Moderate**<br>● Some scope for this content to impact many individuals although evidence suggests that systems in place to reduce the risk of this type of content are effective.<br>● The impact of encountering this type of content can be significant (particularly for younger users). |
| Existing controls / mitigations | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content.<br>● Automated detection and human moderation.<br>● No livestreaming or open newsfeed.<br>● Teen profile protective measures, including limits on public visibility of teen profiles for under 16s, prevention of sexually suggestive content being recommended.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Safety warnings. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| **Harm Type** | **Sexual Exploitation of Adults** |
|---|---|
| Nature and severity of potential harm | Sexual exploitation content includes the inducement of a commercial sex act generally by means of force, fraud or coercion, including adults who have been forced into sex work and consenting adult sex workers who are being exploited.<br><br>This type of content can cause severe physical and psychological harm. Women and younger adults are particularly vulnerable, with women and girls accounting for the majority of victims of sexual exploitation and domestic servitude. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 5 out of 6 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Recommender systems; teenage user base; targeted advertising systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of occurrence.<br>● *External sources:* Various sources contend that millions of people are forced into commercial sex work and that there is a large amount of adult sexual content availale online. Snapchat was not identified in the sources considered. |
| Impact assessment | **Assessment: Moderate**<br>● Impact can vary depending on the nature of the content and the individual, but there is a risk of physical harm for adult sex workers and victims forced into sex work and potential psychological harm for victims of sexual exploitation and viewers who encounter this type of content.<br>● Limits on the ability to share this content on Snapchat plus Snap's internal data indicates that measures to reduce impact are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content and treat it as a severe harm with swift consequences for violators.<br>● Automated detection and human moderation.<br>● User controls such as block, mute, report and leave group chats.<br>● No livestreaming or open newsfeed.<br>● Teen profile protective measures, including limits on public visibility of teen profiles for under-16s, prevention of sexually suggestive content being recommended.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Safety warnings. |
| Overall risk level | Snap assessed the risk associated with sexual exploitation of adults as **Medium**. |

| Harm Type | Human Trafficking |
|---|---|
| Nature and severity of potential harm | Human trafficking content includes arranging or facilitating the travel of another person, or taking a relevant action, with a view to them being exploited. It can involve the exploitation of both children and adults. The impact can be severe.<br><br>External sources indicate that social media platforms may be used to find and recruit potential victims, including through job advertisements and subsequent private communications. Children and young adults may be particularly vulnerable to exploitation, while women and girls may face elevated risks from certain forms of human trafficking. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 8 out of 10 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Friend recommendations; group chat with non-friends; teenage user base; account names may be written to attract like-minded users; targeted advertising systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources*: Various sources contend that social media platforms (in general) may be used to promote and facilitate human trafficking. |
| Impact assessment | **Assessment: Moderate**<br>● Scope for severe impact due to links to offline violence, youth vulnerability and organised criminal activity.<br>● However, evidence suggests that systems and processes in place to reduce this risk are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● Automated detection and human moderation.<br>● Location sharing off by default, and users can share location on Snap Map with friends.<br>● Working with industry groups (e.g. the EU Internet Forum) which work to tackle issues including human trafficking.<br>● Teen account protective measures, including limits on public visibility of teen profiles for under-16s.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Safety warnings. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Unlawful Immigration |
|---|---|
| Nature and severity of potential harm | Unlawful immigration content includes promoting, facilitating or enabling unlawful immigration.<br><br>External sources indicate that social media platforms may be used to promote unlawful immigration services and connect individuals with organised crime groups. Potential harms include serious risk to life and safety, particularly in connection with dangerous small boat crossings. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 7 out of 7 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Friend recommendations; group chat with non-friends; account names may be written to attract like-minded users; recommender systems; targeted advertising systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources*: Various sources contend that social media platforms (in general) are increasingly used to promote and facilitate unlawful immigration, particularly through advertisements and subsequent private communications. |
| Impact assessment | **Assessment: Moderate**<br>● Scope for severe impacts including promotion and facilitation of unlawful immigration. However, evidence suggests that systems and processes in place to reduce this risk are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content.<br>● Updated enforcement framework to better identify this content.<br>● Automated detection and human moderation.<br>● No marketplace for the sale of goods. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Fraud and Financial Services Offences |
|---|---|
| Nature and severity of potential harm | Fraud and financial services content may facilitate fraudulent schemes, misleading financial promotions, investment scams, phishing, impersonation and the promotion of fraudulent goods or services. Potential harms include fraudsters obtaining money or personal information from victims which may result in financial loss as well as negatively affect the mental and/or physical health of victims.<br><br>People of any age can be victims of fraud, although investment scams mainly target older age groups. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 12 out of 14 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Group chat with non-friends; targeted advertising systems; recommender systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid medium turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Various sources contend that social media and direct messaging platforms (in general) are used to carry out fraudulent activity, but no indication that Snapchat is a primary platform for this type of activity. |
| Impact assessment | **Assessment: Moderate**<br>● Impact can vary depending on the nature of the content, the individual affected and how the activity is carried out.<br>● The harms can be broad and can affect individuals' mental and physical health.<br>● Limits on the ability to share this content on Snapchat reduces the resulting impact. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content.<br>● Automated detection and human moderation.<br>● No marketplace for the sale of goods.<br>● Snap's Advertising Policies prohibit this content in ads.<br>● Use of 'trusted flagger' program to enable such content to be reported through dedicated high-priority channel (such as the City of London Police and DWP).<br>● Snap is a signatory to the Online Fraud Charter. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Proceeds of Crime |
|---|---|
| Nature and severity of potential harm | Proceeds of crime content promotes or enables criminals to unlawfully obtain money or assets gained by their criminal activity and money laundering. Potential harms include the recruitment of people as money mules, arranging money laundering, and offering stolen personal information for sale to facilitate fraud.<br><br>External sources indicate that social media platforms may be used to recruit individuals into these activities. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 6 out of 6 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Large user base; friend recommendations; group chats with non-friends; targeted advertising systems; recommender systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Various sources contend that increasing use of social media platforms (including Snapchat) promotes and facilitates proceeds of crime-related activities. |
| Impact assessment | **Assessment: Moderate**<br>● Scope for significant impact due to links with organised criminal activity. However, evidence suggests that systems and processes in place to reduce the risk are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content.<br>● Automated detection and human moderation.<br>● No marketplace for the sale of goods.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Snap engages with working groups to maintain awareness of proceeds of crime offences online. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Drugs and Psychoactive Substances |
|---|---|
| Nature and severity of potential harm | Drugs and psychoactive substances content includes the unlawful supply or offer to supply controlled drugs and psychoactive substances.<br><br>External sources indicate that teenagers and young adults may be particularly vulnerable to these harms. Studies indicate that men can be at an elevated risk of certain crimes, particularly those involving drugs. |
| Likelihood assessment | **Assessment: High**<br>● *Ofcom Specific Risk Factors:* 11 out of 14 apply.<br>● *Additional service characteristics that could increase the risk of harm if not effectively mitigated include*: Friend recommendations; group chat with non-friends; account names may be written to attract like-minded users.<br>● *Internal data*: Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered. Slightly higher volume of enforcement activity than other harms.<br>● *External sources*: Various sources contend that social media platforms (including Snapchat) are allegedly used in the sale of illegal drugs. |
| Impact assessment | **Assessment: High**<br>● Although Snap's data indicates that this harm has a low likelihood of being encountered on Snapchat, Snap recognises the potential for severe impact, particularly the illegal sale of lethal drugs. Snap has therefore erred on the side of caution, and considers there is a potential for high impact. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● Automated detection and human moderation.<br>● No marketplace for the sale of goods.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● In-app educational tools.<br>● Snap is a member of working groups tackling illegal drugs sales. |
| Overall risk level | Snap assessed the risk associated with this content as **High**. |

| Harm Type | Firearms, Knives, and Other Weapons |
|---|---|
| Nature and severity of potential harm | Firearms, knives and other weapons content includes the illegal sale, promotion, or distribution of firearms, knives and other weapons. Potential harms include youth offline violence and organised criminal activity. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 4 out of 8 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Teenage user base; group chat with non-friends; account names may be written to attract like-minded users; recommender systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Various sources contend that online content that promotes or advertises the sale of weapons is widespread, including on social media platforms (including Snapchat). |
| Impact assessment | **Assessment: Moderate**<br>● Potential for significant harm due to potential links to offline youth violence and organised criminal activity.<br>● Some scope for this content to impact many individuals but comprehensive systems and processes in place that evidence indicates are effective in reducing the risk. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines expressly prohibit this content and treat it as a severe harm with strict consequences for violators.<br>● Automated detection and human moderation.<br>● No marketplace for the sale of goods.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● Teen account protective measures, including limits on public visibility of teen profiles for under-16s, friend recommendation safeguards, and safety warnings. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Encouraging or Assisting Suicide and Serious Self-Harm |
|---|---|
| Nature and severity of potential harm | Suicide or serious self-harm content encourages or assists a person to end their life or self-harm. Exposure to illegal or harmful self-harm and suicide content can present a risk of serious harm, particularly for teenagers and young adults. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 9 out of 13 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Search function.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Various sources contend that this content is on social media platforms (including Snapchat) but do not indicate that Snapchat is used disproportionately to disseminate suicide or self-harm activity compared to other platforms. |
| Impact assessment | **Assessment: Moderate**<br>● Scope for serious harm, particularly for teenagers and younger users.<br>● Some scope for this content to impact many individuals, however, evidence suggests that systems and processes in place to reduce this risk are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content.<br>● Automated detection and human moderation.<br>● User controls such as block, mute, report, leave group chats and disable comments on certain public content.<br>● This content is not eligible for recommendation on Spotlight. If Snap detects self-harm or suicide that is not prohibited (such as news or public commentary) it is marked as sensitive to prevent its distribution.<br>● Signposting to support resources.<br>● Snap is a member of a signal-sharing initiative focused on suicide, self-harm, and eating disorder content. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Foreign Interference |
|---|---|
| Nature and severity of potential harm | Foreign interference content can include misrepresentation of a person's identity or purpose, or in the presentation of the information, for example, through state-backed disinformation campaigns. This type of illegal content also includes false communications content.<br><br>Harm can arise if this content manipulates the UK information environment, misleads users or undermines democratic institutions. |
| Likelihood assessment | **Assessment: Low**<br>● *Ofcom Specific Risk Factors:* 10 out of 13 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Targeted advertising systems; misuse of account creation tools to set up inauthentic or impersonating profiles.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources*: Various sources contend that foreign influence operations remain a threat globally, but do not identify Snapchat as a primary platform for this activity. Platforms with higher political engagement or more open content-sharing systems tend to be used instead. |
| Impact assessment | **Assessment: Low**<br>● Potential to mislead individuals and undermine democratic integrity. However, evidence suggests that systems and processes in place to reduce the risk of this type of content are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content.<br>● Pre-approval requirements before political content can be recommended.<br>● No livestreaming or open newsfeed.<br>● Policies for generative AI tools tackling impersonation. |
| Overall risk level | Snap assessed the risk associated with this content as **Low**.<br><br>This rating also covers the related non-priority offence of false communications. |

| **Harm Type** | **Animal Cruelty** |
|---|---|
| Nature and severity of potential harm | Animal cruelty content includes the unnecessary suffering of animals. This includes obscene content showing torture of humans and animals.<br><br>External sources indicate that exposure to this content can cause emotional distress or lead to harmful imitation. |
| Likelihood assessment | **Assessment: Low**<br>● *Ofcom Specific Risk Factors:* 5 out of 6 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Recommender systems.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered.<br>● *External sources:* Various sources contend that animal abuse content has been reported online more generally, but do not identify Snapchat as a platform of concern for this content. |
| Impact assessment | **Assessment: Low**<br>● Impact may vary depending on the nature of the content and the individual, but could include emotional distress or lead to harmful imitation. Severity of impact likely to be low given the relatively low volume of this content on Snapchat as well as the comprehensive systems and processes in place to reduce the risk of this type of illegal harm. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content.<br>● Automated detection and human moderation.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● No livestreaming or open newsfeed.<br>● User controls such as block, mute, report and leave group chats. |
| Overall risk level | Snap assessed the risk associated with this content as **Low**.<br><br>This rating also covers the related non-priority offence of obscene content showing torture of humans and animals which Snap considers to be closely linked to this type of harm. |

| Harm Type | Cyberflashing |
|---|---|
| Nature and severity of potential harm | Cyberflashing content includes where a person intentionally sends a photograph or film of genitals for the purposes of causing alarm, distress or humiliation or for the purpose of obtaining sexual gratification. Potential harms include psychological harm, alarm, trauma, distress or humiliation.<br><br>External sources indicate that younger adults are more likely to experience sexual harms and that cyberflashing is a gendered offence whereby women are significantly more likely than men to have received an unsolicited sexual image. |
| Likelihood assessment | **Assessment: Moderate**<br>● *Ofcom Specific Risk Factors:* 6 out of 7 apply.<br>● *Additional service characteristics that could increase the risk of this harm if not effectively mitigated include:* Teenage user base.<br>● *Internal data:* Existence of this type of illegal harm, but rapid median turnaround time from detection to enforcement and low likelihood of being encountered. (Note: Data relating to private messaging is more limited)<br>● *External sources*: Various sources contend that there are high numbers of intimate images being shared online, although a smaller proportion of this is criminal activity. Do not identify Snapchat specifically but Snap is aware of the potential for this type of harm to arise on private and limited broadcast content services. |
| Impact assessment | **Assessment: Moderate**<br>● Impact can include psychological harm, trauma and distress. However, evidence suggests that systems and processes in place to reduce this harm are effective. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines prohibit this content.<br>● Automated detection and human moderation.<br>● An accepted friend connection or contact book connection is required for direct messaging.<br>● User controls such as block, mute, report other users and leave group chats.<br>● Teen account protective measures, including limits on public visibility of teen profiles for under-16s, direct messaging only to accepted Snap friends or people already in contact book and safety warnings. |
| Overall risk level | Snap assessed the risk associated with this content as **Medium**. |

| Harm Type | Other Illegal Activities |
|---|---|
| Nature and severity of potential harm | Other illegal activities include illegal content or activity that is not already covered by the specific OSA Illegal Content Categories assessed above.<br><br>As Snapchat allows users to publish content, it is possible that information relating to other illegal activities may be shared on the service. |
| Likelihood assessment | **Assessment: Low**<br>● *Ofcom Specific Risk Factors:* N/A<br>● *Data and external sources:* Snap did not identify data or external sources suggesting that a significant new category of illegal content or activity is occurring on Snapchat. |
| Impact assessment | **Assessment: Low**<br>● Snap's assessment of a low likelihood of a significant new category of illegal content or activity on Snapchat, Snap considers the impact to be limited. |
| Existing controls / mitigation measures | Include (but are not limited to):<br>● Snap's Community Guidelines generally prohibit content that is illegal or harmful to users.<br>● Users can report illegal content.<br>● Monitor indicators of other illegal content.<br>● Engagement with UK law enforcement and regulatory partners to ensure ongoing compliance and rapid response. |
| Overall risk level | Snap assessed the risk associated with other illegal activities as **Low**. |