

Snap DSA Report

Risk Assessment Results and Mitigations



Public Copy - 30 November 2024



Publication Statement

Background

Providers of Very Large Online Platforms and Very Large Search Engines (as designated by the Commission) are required to complete a report every year setting out the results of their risk assessment and details of its mitigations pursuant to Article 42.4.(a), (b) and (e) of Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (the “Digital Services Act” or “DSA”).

Providers must: (i) share their reports with the Commission without undue delay following their completion; and (ii) publish the report on their website within 3 months of Snap having received its final auditors report (as this includes an audit of Snap’s compliance with its risk and mitigation assessment and reporting obligations) pursuant to Article 42(4) of the DSA i.e. approximately 1 year and 3 months after each report is completed.

2023 Report

Snapchat’s online platforms were designated by the Commission on 25 April 2023. Snap completed its first risk and mitigation report for Snapchat’s online platforms in August 2023 (the “2023 Report”), which it sent to the Commission without undue delay. The 2023 Report contains an extensive assessment of risks and mitigations with respect to Snapchat’s online platforms, as well as a full explanation of our methodology.

On 30 November 2024, Snap also published a public version of its 2023 Report on the [European Union transparency page](#). This public version can be found in the **Annex** to this document. It has been redacted to remove confidential and other information permitted pursuant to Article 42(5) of the DSA and, where possible, replaced with a summary of the redacted information.

Updates since the 2023 Report

Snap has produced a second risk and mitigation report on Snapchat’s online platforms (the ‘2024 Report’). The 2024 Report is an update to the 2023 Report to reflect new internal and external changes that may impact our assessment of risk and mitigations. As the 2024 Report is not due for publication until November 2025, we have voluntarily shared highlights from that 2024 Report to explain what has and has not changed with respect to Snapchat’s online platforms since we completed the 2023 Report in August 2023:



Section 1 - Introduction

- [Snapchat 101](#) - In all material respects, Snapchat's online platforms remain the same since our 2023 Report. We have not deployed functionalities that were likely to have a critical impact on the risks identified pursuant to Article 34 of the DSA. We have flagged to the Commission two significant product changes that we are currently evaluating, Simple Snapchat and a Public Profiles experience that is suitable for 16-17 year olds (profiles are private by default on Snapchat). [Simple Snapchat](#) is primarily a cosmetic change that simplifies the Snapchat experience, and is not likely to have a critical impact on our risk assessment. Public Profiles 16-17 will not be rolled out to the EU until we have finalized our review and completed an update to our risk and mitigation assessment as needed.
- [Snapchat Community](#) - We continue to observe positive growth in our user base globally, and in the European Union we grew to [92.4 million average monthly active recipients](#) of our Snapchat app (as at 1 August 2024). Our community demographics have not seen any significant changes since our 2023 Report.

Section 2 - DSA Risk Assessment Scope

- [Scope Assessment](#) - Since the 2023 Report: (i) our Snapchat designation has not changed; (ii) the Commission has not issued any new guidance relating to scope and (iii) the functionality of Snapchat has not significantly changed. We have therefore confirmed that Snap still considers the Spotlight, Discover, Public Profiles, Snap Map, Lenses, and Advertising online platforms of Snapchat to fall within the scope of our risk assessment and mitigation obligations in Articles 34 and 35 of the DSA.

Section 3 - DSA Risk Assessment Methodology

- [DSA Risk Assessment Methodology](#) - We have not made any material changes to our risk assessment methodology since our 2023 Report.

Section 4 - DSA Risk Assessment Results

- There have been no changes to the conclusions we reached in our 2023 Report i.e. we continue to assess that we have reasonable, proportionate and effective mitigation measures for every systemic risk identified in Article 34 of the DSA.
- We have seen further substantial reductions in prevalence rates across all of the illegal and other violating content categories.
- Although the likelihood of all of the risks tracked by Snapchat are very low, this further reduction in prevalence has resulted in four risks moving to even lower likelihood categories according to our methodology set out in section 3: Adult Sexual Content, Fraud and Spam, Regulated Goods Content and Harassment and Bullying Content.
- We have also reflected on recent events and studies:



- We described our recent positive experiences in the run up to and during the European elections and to address the Commission’s Guideline for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes.¹ This has helped to further demonstrate that Snapchat’s online platforms do not present risk of negative effects on democratic and electoral processes and we have reasonable, proportionate and effective mitigation measures.
- We have identified new quantitative studies, in particular from the Netherlands, that compare online platforms on well-being, self-esteem and friendship closeness.² This found that Snapchat was the only online platform studied that positively impacts well-being and Snapchat also has a strong positive effect on friendships and no net negative effect on self-esteem. This has helped to further demonstrate that Snapchat does not present a risk of negative consequences on physical and mental wellbeing and we have reasonable, proportionate and effective mitigation measures.

Section 5 - Specific Mitigations

There have been some updates to the mitigations relating to Snapchat’s online platforms since our 2023 Report:

- Transparency - We made updates to our [Privacy Policy](#), added additional safety resources to our [Safety Center](#) and launched parents.snapchat.com - a dedicated microsite for parents, educators and other caregivers.
- Protection of Minors - We have provided updates on the mitigation measures we put in place to protect minors, including:
 - Further information on our use of inferred age techniques to identify teen users of adult accounts and confirmation that we disabled targeted advertising to minors;
 - Further information on the readability of our terms and conditions, administration and oversight, and how we apply the age appropriate design code; and
 - New functionality added to [Family Center](#) and the launch of our new parents site which provides additional guidance for parents and caregivers on risks and support.³

Since completing our 2024 Report, we have published summaries of our safeguards for ‘Snapchat Teen Accounts’ on our website:

- [Teens | Snapchat Privacy](#)
- [Snapchat Safeguards for Teens](#)
- Content Authenticity - We have updated this section to confirm the steps that Snap has taken to mitigate risks that (i) its generative AI tools may be used to create illegal or

¹ Guidelines for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes, April 2024, [url](#).

² Social Media Use Leads to Negative Mental Health Outcomes for Most Adolescents, Amber van der Wal, Ine Beyens, Loes H. C. Janssen, and Patti M. Valkenburg, 2024, [url](#) (preprint)

³ <https://parents.snapchat.com>.



otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on Snap’s or third party services may be disseminated to the public via Snapchat’s online platforms.

- Codes and Crisis Protocols - Snap has engaged with the European Commission and other stakeholders on updates to the EU Hate Speech Code. We have occasionally actioned our internal crisis protocols in response to international events.

There have not been material updates with respect to our other migrations: (i) Snapchat Design and Function; (ii) Terms; (iii) Moderation; (iv) Enforcement; (v) Algorithmic Systems; (vi) Targeted Advertising; (vii) Trusted flaggers; and (viii) Dispute Settlement Bodies.

Section 6 - Ongoing Risk Detection & Management

Since our 2023 Report, there have also been some updates to the measures we have in place to detect and manage risks on an ongoing basis for Snapchat’s online platforms, in particular:

- Prevalence Testing - We conduct a random sampling of Public Stories to identify how many samples contain Snaps that are illegal or otherwise violate our terms (“violating Snaps”). We then work out how many times the sampled violating Snaps have been viewed (“violating views”). We used the number of violating Snaps and violating views to estimate:
 - (i) the percentage of Snaps in all Public Stories that will be violating Snaps (“percentage of distinct violating Snaps”); and
 - (ii) the percentage of views of Snaps in all Public Stories that are violating views (known as the ‘Policy Violating Prevalence’ (PVP) rate)

This is a continuous daily sampling exercise. We have used this to uncover blindspots and prioritized efforts to close those gaps such as improvements to our proactive detection mechanisms, infrastructure improvements and agent training. We continue to measure the prevalence of illegal and otherwise violating content on Snapchat’s video sharing platforms to assess the effectiveness of content moderation, enforcement and other mitigation efforts.

As explained in Section 3 of our 2023 Report, we classify the likelihood of our risks using the PVP rate (amongst other available data):

Low Likelihood	This means this risk has the highest chance of occurring on Snapchat vs other risks. Where Prevalence Testing data is available, this risk has a percent of policy-violating prevalence (PVP) of 0.5% - 1.5% or greater.
----------------	--



Very Low Likelihood	This means this risk has an average chance of occurring on Snapchat vs other risks. Where Prevalence Testing data is available, this risk has a percent of policy-violating prevalence (PVP) of between 0.05% and 0.49%.
Extremely Low Likelihood	This means this risk has the lowest chance of occurring on Snapchat vs other risks. Where Prevalence Testing data is available, this risk has a percent of policy-violating prevalence (PVP) of 0.049% or less.

We continue to be pleased with progress from our prevalence testing and have observed further significant reductions in PVP rates since the 2023 Report. In particular:

- All illegal or other violating content categories now fall within our Very Low or Extremely Low likelihood categories (**with the vast majority falling within the Extremely Low Likelihood category**).
- We have observed a further **substantial decrease** in the overall percentage of views of illegal or otherwise violating content (i.e. the overall PVP rate)
- We have also observed a further **substantial decrease** in the percentage of distinct violating Snaps.

	2023 Report (as at 9 August 2023)	2024 Report (as at 30 July 2024)	Change
Overall PVP	Low	Very Low	86% decrease
Percentage of distinct violating Snaps	Low	Very Low	89% decrease

- **External Request Monitoring and Review** - We also continue to rely on data relating to content reports, enforcements and appeals, as well as advertising review rejections, advertising reporting and enforcements, ‘privacy, data protection and DSA’ requests and general community support requests, to assess risk and the effectiveness of our mitigations. Since the DSA came into force, we have also been monitoring DSA queries raised via our dedicated email address and community support page.
- **Snap Advisory Groups** - Our work with the Snap Safety Advisory Board continues to progress since our 2023 Report, including in particular the establishment of a new [Snap Council for Digital Well-Being](#) to hear from teens and their parents on online safety and well-being topics, and gain their perspectives on how these topics should be approached.
- **Digital Well-Being Index (DWBI) Initiative** - We have [published](#) the results from the Year Three industry wide DWBI research study.
- **Audit** - We have noted the completion of our external DSA audit of Snap’s compliance with its obligations under Chapter 3 of the Digital Services Act for the audit period



between August 25th 2023 and June 30th 2024 pursuant to Article 37. A public version of our audit and audit implementation reports have been published with this 2023 Report on our risks and mitigations, pursuant to Article 42.4(c) and (d).

There have been no significant changes to our other migrations to detect and manage risk on an ongoing basis: (i) Platform Principles Framework; (ii) DSA Compliance Team and Cross-Functional Working Groups; and (iii) Privacy and Safety by Design.

Conclusion

- [Conclusion](#) - Since our 2023 Report, we continue to drive down the likelihood of all DSA risks occurring on Snapchat's online platforms, with all risks now falling within the very low or extremely low likelihood categories. We continue to assess that we have in place reasonable, proportionate and effective mitigation measures for Snapchat's online platforms, tailored to the specific systemic risks identified.



Annex - 2023 Report

Contents

Foreword	16
1. Introduction	17
Snapchat 1.01.....	17
Snapchat Community.....	19
2. DSA Risk Assessment Scope	20
2.1 Spotlight.....	21
2.2 For You (previously Discover).....	22
2.3 Public Profiles.....	22
2.4 Snap Map.....	23
2.5 Lenses.....	27
2.6 Advertising.....	27
3. DSA Risk Assessment Methodology	29
3.1 Identification of Risks.....	29
3.2 Likelihood Analysis.....	30
3.3 Severity Analysis.....	30
3.4 Overall Potential Risk Prioritization Assessment.....	31
3.5 Snap's Mitigations and Conclusions.....	32
4. DSA Risk Assessment Results	34
4.1 Category 1 - Dissemination of content that is illegal or violates our terms and conditions..	34
4.1.1 Dissemination of child sexual abuse material.....	36
Likelihood.....	36
Severity.....	37
Overall potential risk prioritization.....	37
Snap's Mitigations.....	37
Conclusion.....	40
4.1.2 Dissemination of illegal hate speech.....	40
Likelihood.....	40
Severity.....	41
Overall potential risk prioritization.....	41
Snap's Mitigations.....	41
Conclusion.....	44
4.1.3 Dissemination of information related to the sale of prohibited products or services (such as dangerous products, counterfeit products or illegally-traded animals).....	44



Likelihood..... 44

Severity..... 45

Overall potential risk prioritization..... 45

Snap's Mitigations..... 45

Conclusion..... 49

4.1.4 Dissemination of terrorist content..... 49

 Likelihood..... 49

 Severity..... 50

 Overall potential risk prioritization..... 50

 Snap's Mitigations..... 50

 Conclusion..... 53

4.1.5 Dissemination of content that infringes on intellectual property rights..... 53

 Likelihood..... 54

 Severity..... 54

 Overall potential risk prioritization..... 55

 Snap's Mitigations..... 55

 Conclusion..... 57

4.1.6 Dissemination of adult sexual content..... 57

 Likelihood..... 58

 Severity..... 58

 Overall potential risk prioritization..... 58

 Snap's Mitigations..... 59

 Conclusion..... 61

4.1.7 Dissemination of content regarding harassment & bullying..... 61

 Likelihood..... 62

 Severity..... 63

 Overall potential risk prioritization..... 63

 Snap's Mitigations..... 63

 Conclusion..... 66

4.1.8 Dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders..... 66

 Likelihood..... 66

 Severity..... 67

 Overall potential risk prioritization..... 67

 Snap's Mitigations..... 68

 Conclusion..... 70

4.1.9 Dissemination of content encouraging or engaging in violent or dangerous behavior.. 71

 Likelihood..... 71



Severity.....	72
Overall potential risk prioritization.....	72
Snap's Mitigations.....	72
Conclusion.....	75
4.1.10 Dissemination of harmful false information.....	75
Likelihood.....	75
Severity.....	76
Overall potential risk prioritization.....	76
Snap's Mitigations.....	76
Conclusion.....	80
4.1.11 Dissemination of fraud and spam.....	81
Likelihood.....	81
Severity.....	81
Overall potential risk prioritization.....	81
Snap's Mitigations.....	81
Conclusion.....	84
4.1.12 Dissemination of information related to other illegal activities.....	85
Likelihood.....	85
Severity.....	85
Overall potential risk prioritization.....	85
Snap's Mitigations.....	86
Conclusion.....	88
4.2 Category 2: Negative Effects on Fundamental EU Rights.....	89
4.2.1 Right to human dignity.....	90
Likelihood.....	90
Severity.....	91
Overall potential risk prioritization.....	91
Snap's Mitigations.....	91
Conclusion.....	94
4.2.2 Right to freedom of expression and assembly.....	94
Likelihood.....	95
Severity.....	95
Overall potential risk prioritization.....	95
Snap's Mitigations.....	95
Conclusion.....	98
4.2.3 Right to private life.....	99
Likelihood.....	99
Severity.....	99
Overall potential risk prioritization.....	99



Snap's Mitigations.....	100
Conclusion.....	103
4.2.4 Right to data protection.....	103
Likelihood.....	103
Severity.....	104
Overall potential risk prioritization.....	104
Snap's Mitigations.....	104
Conclusion.....	108
4.2.5 Right to non-discrimination and freedom of religion.....	108
Likelihood.....	109
Severity.....	109
Overall potential risk prioritization.....	109
Snap's Mitigations.....	109
Conclusion.....	114
4.2.6 Children's Rights.....	114
Likelihood.....	114
Severity.....	115
Overall potential risk.....	115
Snap's Mitigations.....	116
Conclusion.....	119
4.2.7 Right to consumer protection.....	119
Likelihood.....	119
Severity.....	120
Overall potential risk prioritization.....	120
Snap's Mitigations.....	120
Conclusion.....	123
4.2.8 Right to Property.....	123
4.3 Category 3: Negative effect on democratic and electoral processes, civic discourse and public security.....	124
4.3.1 Negative Effect on Democratic and Electoral Processes.....	124
Likelihood.....	125
Severity.....	125
Overall potential risk prioritization.....	126
Snap's Mitigations.....	126
Conclusion.....	130
4.3.2 Negative Effect on Civil Discourse.....	130
Likelihood.....	131
Severity.....	131
Overall potential risk.....	131



- Snap's Mitigations.....131
- Conclusion.....135
- 4.3.3.Negative Effect on Public Security..... 136
 - Likelihood..... 136
 - Severity.....137
 - Overall potential risk prioritization..... 137
 - Snap's Mitigations.....137
 - Conclusion.....141
- 4.4 Category 4: Negative Effects on Public Health..... 142
 - 4.4.1 Negative Effects on Public Health.....143
 - Likelihood..... 143
 - Severity.....143
 - Overall potential risk prioritization..... 143
 - Snap's Mitigations.....143
 - Conclusion.....147
 - 4.4.2 Negative Effects on gender-based violence.....147
 - Likelihood..... 147
 - Severity.....147
 - Overall potential risk prioritization..... 148
 - Snap's Mitigations.....148
 - Conclusion.....151
 - 4.4.3 Negative Effects on Children..... 151
 - Likelihood..... 151
 - Severity..... 151
 - Overall potential risk prioritization..... 152
 - Snap's Mitigations..... 152
 - Conclusion..... 155
 - 4.4.4 Serious Negative Consequences on physical and mental well-being..... 155
 - Likelihood..... 156
 - Severity..... 156
 - Overall potential risk prioritization..... 156
 - Snap's Mitigations.....157
 - Conclusion.....161
- 5. Specific Mitigations..... 162**
 - 5.1 Snapchat Design and Function.....162
 - Privacy and Safety by Design from day one..... 162
 - Adaptations and Mitigations.....163
 - Integrations with other mitigations..... 168
 - Conclusion.....169



5.2 Terms.....	170
Terms of Service.....	170
Community Guidelines.....	170
Platform Specific Terms.....	174
Languages.....	175
Conclusion.....	175
5.3 Transparency.....	175
1. Information we provide on our website.....	175
Privacy and Safety Hub.....	176
Privacy Center.....	176
Safety Center.....	176
Transparency Center.....	178
News Page.....	178
2. Information provided in app stores.....	179
3. Information we provide in our application.....	179
Languages.....	183
Conclusion.....	184
5.4 Moderation.....	184
Overview.....	184
Conclusion.....	185
5.5 Enforcement.....	185
Introduction.....	185
Conclusion.....	186
5.6 Algorithmic Systems.....	186
Introduction.....	186
How do our Recommender Systems work?.....	187
Benefits.....	187
Adaption and Testing.....	187
Conclusion.....	191
5.7 Advertising Systems.....	192
Introduction.....	192
How do our Advertising Systems Work?.....	192
Benefits.....	193
Adaptation and Testing.....	195
Conclusion.....	201
5.8 Protection of Minors.....	202
How We Think About the Protection of Minors.....	202
Specific Safeguards for Teens.....	202
Family Center / Parent Tools.....	202



Age Gating and Age Appropriate Content.....	205
Conclusion.....	206
5.9 Content Authenticity.....	207
Conclusion.....	209
5.10 Trusted Flaggers.....	209
Trusted Flagger Program.....	209
Conclusion.....	210
5.11 Dispute Settlement Bodies.....	210
Conclusion.....	210
5.12 Codes and Crisis Protocols.....	210
Cooperation.....	210
Codes of Practice.....	211
EU hate speech Code.....	211
FSM Code of Conduct.....	211
EU disinformation code.....	212
EU AAD Code.....	212
Crisis Protocols.....	212
Conclusion.....	212
6. Ongoing Risk Detection and Management.....	213
Platform Risk Framework.....	213
DSA Compliance Team and Cross-Functional Working Groups.....	213
Compliance Function.....	214
DSA Cross-functional Working Groups.....	214
Integrating DSA into our Privacy and Safety by Design Process.....	215
Ongoing Risk Management.....	215
Existing Privacy and Safety by Design review process.....	215
Integrating DSA Risk Assessments into our existing process.....	216
Prevalence Testing.....	216
External Request Monitoring and Review.....	216
Digital Well-Being Index (DWBI) Initiative.....	217
Safety Advisory Board (SAB).....	218
Audit.....	219
7. Conclusion.....	220
8. Final Words.....	223
Annex - Community Guidelines: Explainer Series.....	224
Sexual Content.....	224
Harassment & Bullying.....	224
Threats, Violence & Harm.....	224
Harmful False or Deceptive Information.....	224



Illegal or Regulated Activities..... 224
Hateful Content, Terrorism, and Violent Extremism..... 224
Severe Harm..... 224
Snapchat Content Moderation, Enforcement, and Appeals..... 224



Foreword

This Risk Assessment Results and Mitigations Report (**Report**) has been prepared to comply with Snap’s obligations under Article 42.4.(a), (b) and (e) of Regulation (**EU**) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (the “Digital Services Act” or “DSA”).

This Report is divided into eight sections: (1) Introduction; (2) DSA Risk Assessment Scope; (3) DSA Risk Assessment Methodology; (4) DSA Risk Assessment Results; (5) Specific Mitigations; (6) Ongoing Risk Detection and Management; (7) Conclusion; and (8) Final Words.



1. Introduction

At Snap, our mission is to contribute to human progress by empowering people to express themselves, live in the moment, learn about the world, and have fun together.

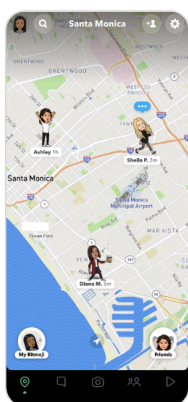
Even as Snap grows and faces new opportunities and challenges, we remain grounded in kindness. Our engineers, designers, product managers, and other team members build our products and services to serve people. The well-being of the community informs our decision making, which in turn creates more value for our business over the long term.⁴

Snapchat 1.01

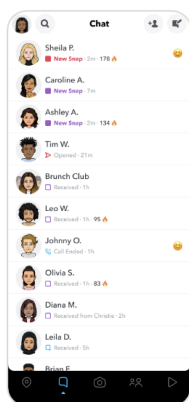
Snapchat is a communications app designed for people ages 13 and up, who primarily use it to talk with their close friends, similar to the ways they interact in real life. It's similar to how older generations use text messaging or their phone to stay in touch with friends and family. Our data shows that the vast majority of our users are primarily using the messaging aspects of our platform. While the products detailed in this report and within scope of the DSA primarily revolve around our public content surfaces, our core use is a messaging app, which sets us apart from many other VLOPs, and we believe is an important lens through which to view Snap and our platform.

We purposely designed Snapchat differently from traditional social media. It doesn't open to a public news feed powered by an algorithm with likes and comments. Instead, Snapchat opens to a camera and has five tabs: Camera, Chat, Map, Stories, and Spotlight.⁵

Platform Architecture



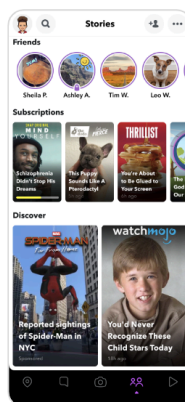
Map



Chat



Camera



Stories



Spotlight

⁴ See our [Citizen Snap Report](#) for more details.

⁵ View our [Snapchat 101 video](#) for more details.



Camera

Snapchat opens into a camera, making it an easy and visual way for people to share what's on their mind with the people that matter most to them. Snapchatters can Snap a quick video or photo with our augmented reality Lenses to put fun and educational layers on the world, and get creative by overlaying text, stickers, and more.

Chat

To the left of the Camera is Chat, where Snapchatters can talk with their friends and family using text and pictures. Chats will show when both friends are there at the same time. They'll also indicate when a friend has opened and viewed a Snap.

Snaps and Chats delete-by-default to mirror real life conversations, where what one says or does isn't recorded forever and shared with a bunch of strangers. This helps people feel more comfortable expressing themselves, the same way they would if they were just hanging out with friends in person. While Chats and Snaps delete by default, Snapchatters do have the option to save Chats – simply by tapping on the ones they want to save.

In Chat, you can also make voice and video calls and join group conversations and chat with My AI, our chatbot powered by OpenAI's ChatGPT technology.

Map

Swipe to the left of Chat for the Map. Our Map is an interactive way for Snapchatters to share their favorite spots, discover new places, and see what their friends are up to – but only if they choose to share their location with their friends.

Profile

My Profile features a user's Snapchat info, like their [Bitmoji](#) (which is an avatar representation of the user), location on the Map, friend info, and more. My Profile is also where Snapchatters can manage their friendships, and report, block, or remove a friend.

Public Profile

Public Profiles enable Snapchatters to be discovered in the app. If Snapchatters want a Public Profile, they will need to create one first. Once they have created a Public Profile, they can showcase their favorite public Snaps and share Lenses and other information.

Stories / For You

Swipe to the right of the Camera for Stories. Snapchatters can add Snaps to their Stories to share more of their day with friends and family, and scroll down to discover new Stories and content about the world — produced by trusted media publishers and popular creators.



Spotlight

Right next to Stories is our entertainment platform, Spotlight. This is where Snapchatters can submit and watch short, fun, and creative videos for our community.

In [Section 2](#) of this report we provide more details on the products and services that are in scope of the DSA Risk Assessment.

Snapchat Community

We reach over 750 million⁶ monthly active users around the world, and see a path for Snapchat to reach over 1 billion people in the next 2-3 years at our current growth rate. Additionally, we have over 397 million⁷ daily active users (up 14% year-over-year), and have been growing at strong rates for many consecutive quarters. We provide information on the average monthly active recipients of our Snapchat app, across the EU and per EU Member State, in our [European Union transparency page](#) on our website.

Our European Snapchatter community consists of a diverse range of ages and genders. While Snapchat does have a young demographic, by far the largest age category is 18-24, the second largest age category is 25-34, 35+ makes up the third place, and 13-17 is the smallest age category. In terms of gender, our analysis indicates that our community is fairly balanced but with a slightly higher percentage of the community identifying as female. A more detailed analysis of gender shows a slightly higher percentage of our 18-24 age group identifying as male, with a higher percentage of our 35+ age group identifying as female.

⁶ Snap Inc. internal data Q4 2022.

⁷ Snap Inc. internal data Q2 2023 and as compared to Q2 2022 for year-over-year percentage.



2. DSA Risk Assessment Scope

Articles 34 and 35 apply to Very Large Online Platforms designated by the European Commission. Snapchat was designated as a Very Large Online Platform by the Commission on 25 April 2023 because the Average Monthly Active Recipients of Snapchat exceeds 45 million.

The Commission Decision to designate Snapchat as a Very Large Online Platform states that it only applies to services provided as part of Snapchat that meet the definition of online platform laid down in Article 3, point (i), of Regulation (EU) 2022/2065. The designation does not apply to services that are provided together with Snapchat, such as a private messaging service, and that, based on their technical functionalities, do not in themselves meet the definition of online platform laid down in Article 3, point (i), of Regulation (EU) 2022/2065.

Article 3.(i) of the DSA defines ‘online platform’ as:

“a hosting service that, at the request of a recipient of the service, stores and disseminates information to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation”.

Recital 14 explains that:

“The concept of ‘dissemination to the public’, as used in this Regulation, should entail the making available of information to a potentially unlimited number of persons, meaning making the information easily accessible to recipients of the service in general without further action by the recipient of the service providing the information being required, irrespective of whether those persons actually access the information in question.

Accordingly, where access to information requires registration or admittance to a group of recipients of the service, that information should be considered to be disseminated to the public only where recipients of the service seeking to access the information are automatically registered or admitted without a human decision or selection of whom to grant access. Interpersonal communication services, as defined in Directive (EU) 2018/1972 of the European Parliament and of the Council ⁽²⁴⁾, such as emails or private messaging services, fall outside the scope of the definition of online platforms as they are used for interpersonal communication between a finite number of persons determined by the sender of the communication.

However, the obligations set out in this Regulation for providers of online platforms may apply to services that allow the making available of information to a potentially unlimited number of recipients, not determined by the sender of the communication, such as through public groups or open channels. Information should be considered disseminated to the public within the meaning



of this Regulation only where that dissemination occurs upon the direct request by the recipient of the service that provided the information.”

Taking account of this definition and guidance, and that fact that Snapchatters are automatically registered without a human decision or selection of whom to grant access, Snap considers the Spotlight, For You (previously known as Discover), Lenses, Public Profiles, Snap Map, and Advertising services of Snapchat within the scope of risk assessment and mitigation obligations in Articles 34 and 35. These services entail making information published by recipients of those services easily accessible to other recipients of Snapchat in general without further action by the recipients publishing the information in question. When we refer to “Snapchat” or “Snapchat’s in-scope services” in this Report, therefore, we are referring to those six services in Snapchat unless the context is clear that it is referring to Snapchat as a whole.

The six in scope services of Snapchat are described in more detail in the following sections.

2.1 Spotlight

What is Spotlight?

Spotlight is the Snapchat community’s destination for entertaining short-form video content. Launched in November 2020, Spotlight provides users a simple way to view short-form videos created and submitted by the Snapchat community via a personalized feed. All users can post videos to Spotlight either via the Snapchat app or on the website, and videos on Spotlight are public and visible to users on the Snapchat app, on the web, and a link to the Spotlight video can be shared to other platforms. Users can also add Comments to Spotlight videos, which go through moderation before being shown to the creator to either accept or reject. If accepted, the Comment is publicly visible on the Spotlight video. Spotlight Comments may be deleted or reported, and viewers can also indicate fondness by clicking on a heart icon.

In addition to compliance with Spotlight Terms, users must also comply with the [Community Guidelines](#) and the [Spotlight Guidelines](#). Certain higher profile Snapchatters have the opportunity to receive revenue from their content if they, and their Spotlight Snaps, meet certain [eligibility criteria during the Eligibility Period](#).

How does Spotlight work?

Spotlight provides a content experience that is intended to entertain and delight users in the same app they use to communicate with their friends and family. It offers creators at all stages of their career a variety of opportunities and tools to help them grow their audiences, build sustainable businesses and make content creation a full-time career. Spotlight is an easy entry point to start a creator journey and is a source of relevant cultural trends and credible partner to the industry (media, music, sports, fashion, etc.) that offers meaningful reach, relevance and revenue.



The content shown in Spotlight is personalized to provide the user with a more relevant experience. Spotlight’s ranking algorithm is described [here](#). Users may opt out of personalization as described [here](#). Spotlight content is moderated using a combination of auto-moderation and human moderation, and all Spotlight content is human moderated before being widely distributed. Spotlight also uses various engagement and metadata to determine eligibility to receive revenue from their content.

2.2 For You (previously Discover)

What is For You?

For You or Discover is part of the 4th tab in the Snapchat app, below your friends’ Stories. We are currently in the process of renaming this product from Discover to For You. Hence, any reference to Discover in this Report or in supporting documentation should be interpreted as a reference to For You.

For You is dedicated to Creator Stories, which includes Media Partner content, Snap Originals, and some user generated content (“UGC”) created from Snaps by popular users (“Creator Content”). The UGC that appears on For You includes the Public Stories from Snap Stars and other users who meet a follower count threshold. The videos in For You are accessible to all users including those between 13-17 years old.

How does For You work?

For You displays personalized content to users. For You achieves this using its ranking algorithm, which is described [here](#). The intended purpose of this processing is to personalize For You and make it easy for users to discover new content that is relevant to their interests. The intended effect/impact on users is that they enjoy what they are watching and remain engaged users of Snapchat. Users may opt out of personalization as described [here](#).

For You also generates information about how Snapchatters interact with the content in For You. It achieves this by generating ‘event’ metadata each time a user does something noteworthy, like viewing or skipping a video. The intended purpose of this processing is to select content the user is likely to be interested in, in order to further personalize content on For You and elsewhere in Snapchat (such as other content areas like Spotlight and also Advertising - the revenue from which is used to pay for Snapchat). The intended effect/impact on users is that they enjoy their experience and remain engaged users of Snapchat.

2.3 Public Profiles

What are Public Profiles?

Public Profiles enable Snapchatters to be discovered and followed in the app and showcase their favorite public Snaps, Lenses and other information. Snapchatters (including businesses) can create and access Public Profiles and grow an audience with their public identity. Public Profiles



enables Snapchatters to showcase Stories, Spotlights and Lenses and lets users make their name visible on Spotlight posts. For more information, see [here](#).

How do Public Profiles work?

Snapchatter accounts aged 18 and over can opt into having a Public Profile if Snapchatters want to share a bit more about themselves with a wider audience (beyond their immediate friends).

Creating a public profile is straightforward. An eligible Snapchatter is required to: (i) tap their Bitmoji or Story icon at the top to go to My Profile; (ii) Scroll down to the 'Public Profile' section and Tap 'Create Public Profile' and (iii) and then follow the simple instructions to create their Public Profile.

With a public profile, a Snapchatter can:

- Add a Photo, Bio, Description, Location, Stories and Lenses to your Public Profile
- Be Subscribed to by other Snapchatters
- Show their Subscriber Count
- View Public Story, Lens, and Audience Insights
- Add Snaps to their Public Story.

Snapchatters with a public profile that are particularly active can have their accounts upgraded to a Creator Account. These have advanced features that are designed to enable professional Creators to connect and grow with their audience. Creator Accounts are eligible to have their content shown in the For You section of Snapchat.

Note: Since completion of this 2023 Risk Assessment, we have announced that we are reviewing a [16-17 Public Profile experience](#) with specific mitigations for this age group. This product is not yet available in the EU. It will not be rolled out to the EU until we have finalized our review and completed an update to our risk and mitigation assessments as needed.

2.4 Snap Map

What is Snap Map?

Snap Map is designed to open up a world of possibilities for our community, enabling friends to experience something new in the world every day. Through an interactive map interface, Snap Map shows users what's happening nearby and around the world, anchored by the context of friends' Bitmojis. It's a personal map that starts with the user at the center and reflects the people, places, and activities they care about, and helps users meet up with friends, express themselves, find things to do, and explore places elsewhere. The Snap Map was developed with the privacy and safety of our community of Snapchatters in mind.

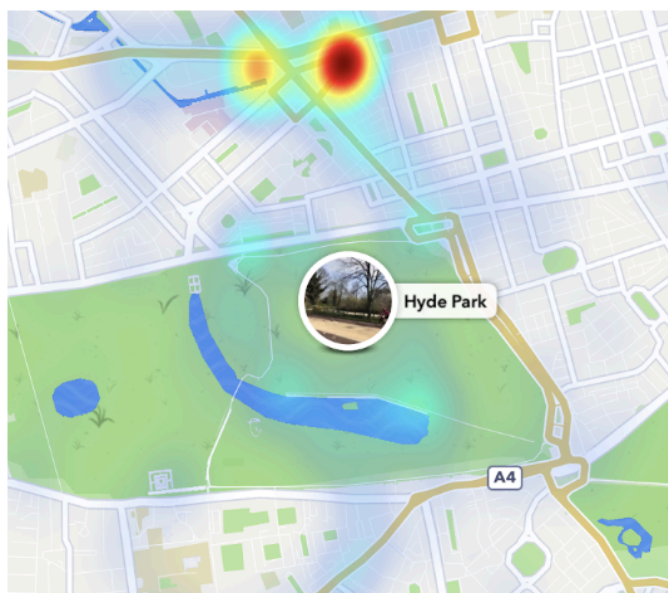


How does Snap Map work?

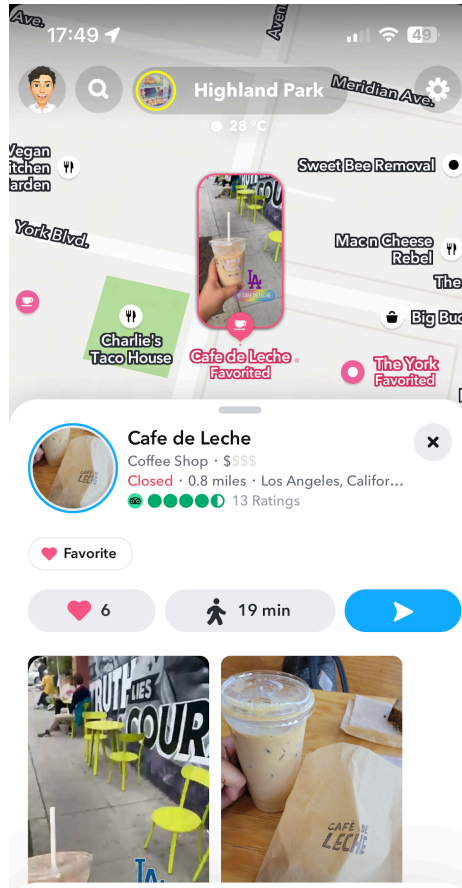
Snapchatters can share their Snaps to the Map by selecting “Snap Map” on the “Send To” page. If the Snapchatter has a Public Profile or is sharing their My Story with everyone, they may also have their Snap shared on Snap Map when it's tagged to a place or venue. Snapchatters can also choose to share their location on the Map with friends while the Snapchat app is actively being used, or share their live location with them even when the app is backgrounded.

Snap Map features five types of user-generated content that can be served:

Map Stories include thumbnails on the map that highlight interesting events and popular places on the Map



Place Stories appear on Place profiles. They contain Public Stories snaps explicitly tagged with the place, using either venue filters or place stickers. Snaps up to 90 days old can appear in a Place Story.



City Stories appear in the header of the Map and display the best snaps in that locality from the last 7 days. They can appear for cities and neighborhoods.

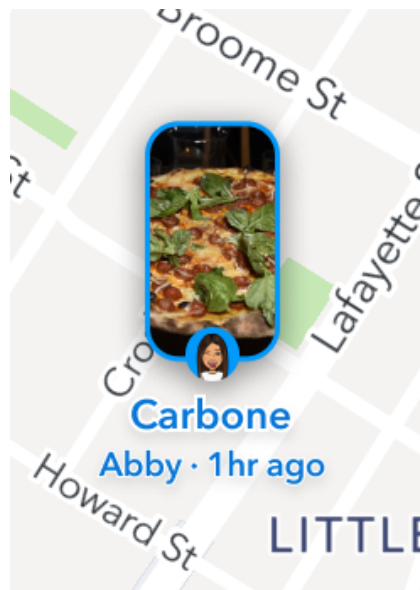




The **Heat Map** is used to visualize the volume and recency of content that's submitted to Public Stories. Content up to seven days old can be accessed by the heatmap. Heat spots represent areas where there is recent, high volume content.



Friend Stories tagged with Places presents a view of snaps that have been tagged with Places by a user's friend, along with the Bitmoji of the friend, that would appear with the Place on the basemap. This helps to personalize places on the user's Map, highlighting the places friends have recently visited.





2.5 Lenses

What are Lenses?

Snapchat Lenses are [augmented reality \(“AR”\)](#) experiences designed to transform the way users look and the world around them. Snapchatters frequently use Lenses for entertainment purposes, for example by creating Snaps with added 3D effects, digital objects, characters, and transformations to their image and voice. For example, Lenses can be used to add a layer of make-up to the user’s face, to distort the user’s face, to add a different background or certain elements to the surroundings. The most popular Lenses at the moment can be found [here](#). Snapchatters can interact with Lenses in the Carousel, via Search, and via Lens Explorer. In addition, we offer advertisers the possibility of creating [Sponsored Lenses](#).

How do Lenses work?

Lenses (in popular language often dubbed as ‘filters’) are created by a relatively limited number of community developers, and Snap’s internal Lens Team. The transformational effects of Lenses are often accomplished through object detection, which is an algorithm designed to help a computer generally understand what objects are in an image. For example, it lets us know that a nose is a nose or an eye is an eye. There are numerous AR development tools Snap has made publicly available through Lens Studio, Snaps’ Lens development platform and there are also internal tools that only the Lens Team can use to develop Lenses. Snap’s AR development tools are reviewed by privacy engineering and legal before being used in Lenses. Some examples of AR development tools are object detection, text to speech, location landmarks and ML models and algorithms to support AR effects like tools for depth and context understanding, all designed to help a computer generally understand what objects are in an image.

We provide provided further information about how Lenses works in product specific support pages:

- [How to use lenses](#)
- [Create Your Own Filters & Lenses • Snapchat](#)

Snapchatters can create or develop Lenses in the desktop application ‘Lens Studio’. There is a Public version and an internal Snap version of Lens Studio. Lens developers may publish Lenses through ‘My Lenses’, a web based portal. Lenses built by Snap’s Lens Team are organic Lenses.

2.6 Advertising

What is Snap’s Advertising product?

Snap relies on online advertising to support its business. Snap Advertising is a digital ad product created for advertisers who would like to easily create and manage ads that target relevant audiences on Snapchat. We process user information about Snapchatters to serve them with ads within Snapchat that we think they might be interested in.



An overview of Snap's ads services can be found [here](#) and [here](#). Some of Snap's advertising tools allow advertisers to provide Snap with data about their customers to improve their advertising campaigns. These tools are explained here:

- [Custom List Audiences](#)
- [Snap Pixel](#)
- [Conversion API](#)
- [Advanced](#) and [Estimated](#) Conversion

In addition, we offer advertisers the possibility of creating [Sponsored Lenses](#).

How does Advertising Work?

Our ad ranking algorithm determines which ads are displayed to a Snapchatter who is in the selected audience for those ads. The ad ranking algorithm uses various signals, including prior ad interactions and social signals, to determine which ads that user is more likely to interact with and then combines this with the results of advertiser ad action for that Snapchatter, to select an ad to display. Snap analyzes prior ad interactions to target advertisements. For example, we may determine that a user is likely to swipe up on certain types of ads or download certain types of games when they see an ad on Snapchat. We may then use this information to show that user similar ads.

Snapchatter interactions with the ad (i.e. impression data) is then logged to (a) attribute impressions to conversion events (such as a purchase on an advertiser website or download of an advertiser app) to demonstrate the performance of the ad and (b) to further train the ad ranking algorithm.



3. DSA Risk Assessment Methodology

In order to meet its obligations under Articles 34 and 35 of the DSA, Snap has applied a standard risk methodology adapted from that commonly used to assess risks in other contexts, including the EU general risk assessment methodology for product safety⁸ and the [ICO's DPIA template](#). This methodology has several steps:

3.1 Identification of Risks

As a first step, Snap identified potential systemic risks for each of the four categories outlined in the DSA:

- a. **Category 1 (Article 34.1.(a) / DSA Recital 80):** Dissemination of illegal or violating content, particularly rapidly and widely or as a result of intentional / automated manipulation, including:
 - i. child sexual abuse material
 - ii. illegal hate speech
 - iii. criminal offenses and the conduct of illegal activities, such as the sale of prohibited products or services, dangerous or counterfeit products, or illegally-traded animals.
- b. **Category 2 (Article 34.1.(b) / DSA Recital 81):** Impact on fundamental EU rights, including in particular rights for:
 - i. Human dignity
 - ii. Freedom of expression and of information, including media freedom and pluralism
 - iii. Private life
 - iv. Data protection
 - v. Non-discrimination
 - vi. Children
 - vii. Consumer protection
- c. **Category 3 (Article 34.1.(a) / DSA Recital 82):** Negatively effects on:
 - i. Democratic and electoral processes
 - ii. Civic discourse
 - iii. Public security
- d. **Category 4 (Article 34.1.(a) / DSA Recital 83):** Negative effects, in particular from design and use/misuse such as a coordinated disinformation campaign, on:
 - i. Public health
 - ii. Gender-based violence
 - iii. Children
 - iv. Physical and mental well-being (including addictions)

⁸ EU general risk assessment methodology (Action 5 of Multi-Annual Action Plan for the surveillance of products in the EU (COM(2013)76)), [url](#).



3.2 Likelihood Analysis

As a second step, Snap analyzed the extent to which the identified risk(s) are likely to occur on Snapchat. **In practice the prevalence of almost all of Snapchat's risks are considered to be very low likelihood or extremely low likelihood, in part because of robust mitigations and the inherent design of relevant Snapchat functionality**, so Snap used a measure of relative likelihood between each risk on Snapchat so we can continue to prioritize and improve (as explained in the following table). Note: this is not measuring likelihood relative to other platforms; it is measuring likelihood relative to risks assessed by Snap.

With this in mind, Snap used three levels of relative likelihood:

<i>Relative likelihood of risk occurring on Snapchat</i>	<i>Description</i>
Low Likelihood	This means this risk has the highest chance of occurring on Snapchat vs other risks. Where Prevalence Testing data is available, this risk has a percent of policy-violating prevalence (PVP) of 0.5% - 1.5%.
Very Low Likelihood	This means this risk has an average chance of occurring on Snapchat vs other risks. Where Prevalence Testing data is available, this risk has a percent of policy-violating prevalence (PVP) of between 0.05% and 0.49%.
Extremely Low Likelihood	This means this risk has the lowest chance of occurring on Snapchat vs other risks. Where Prevalence Testing data is available, this risk has a percent of policy-violating prevalence (PVP) of 0.049% or less.

3.3 Severity Analysis

As a third step, Snap analyzed the severity of the identified risk(s) by considering evidence of the potential harm they have caused individuals or society in general. In practice the severity of all the identified risks could cause at least significant harm (which is why they have been identified). So we used a measure of **relative severity** between each risk so it can continue to prioritize and improve.



With this in mind, Snap used three levels of severity:

<i>Harm classification industry wide</i>	<i>Description</i>
Severe harm industry wide	This means this risk has the highest severity vs other risks. We consider severe harm to include both (1) harms that risk significant damage to the physical or emotional well-being of Snapchatters and society at large e.g. external parties influenced by (other people’s use of) Snapchat, and (2) the imminent, credible risk of severe harm, including threats to human life, safety, and well-being.
Serious harm industry wide	This risk has a medium level of severity vs other risks. We consider these risks not to be severe (as defined above) but still have the potential to cause serious harm.
Significant harm industry wide	This means this risk has the lowest severity vs other risks. While not the most severe or serious, these risks still have the potential to cause significant harm.

The safety of Snapchatters is our top priority. We take behavior that threatens the safety of our community very seriously. We collaborate with experts, safety groups, and law enforcement on these topics in order to better educate ourselves and our community, and to ensure we are sufficiently informed to analyze different levels of severity for each risk.

3.4 Overall Potential Risk Prioritization Assessment

As a fourth step, Snap confirmed an overall potential risk prioritization for each identified risk taking account of the likelihood and severity analysis outlined above. This prioritization helps us to assess whether the mitigations we have put in place (as described in Snap’s Mitigations) are proportionate, reasonable and effective as required by Article 35. As a guide we use the following matrix that is commonly used in risk assessment methodologies to determine the overall potential risk. However, this is only an approximation and we make a decision on the overall potential risk, and therefore the prioritization, of a particular issue on a case by case basis depending on the harm classification industry wide or the relative likelihood of risk occurring on Snapchat. As a result, there are instances where we deviated from the overall potential risk prioritization matrix below (and we have explained each of these deviations in the relevant sub-sections of Section 4 - DSA Risk Assessment Results).

Overall Potential Risk Prioritization Matrix

<i>Harm classification</i>	Severe harm industry wide	Level 3	Level 1	Level 1
----------------------------	---------------------------	---------	---------	---------



<i>industry wide</i>	Serious harm industry wide	Level 3	Level 2	Level 1
	Significant harm industry wide	Level 3	Level 3	Level 3
		Extremely Low	Very Low	Low
<i>Relative likelihood of risk occurring on Snapchat</i>				

3.5 Snap's Mitigations and Conclusions

As a fifth step, Snap assessed the mitigation measures that it has taken for each risk category, taking into account its overall potential risk prioritization, and confirmed whether they were reasonable, proportionate and effective (and therefore whether Snap considers the resulting residual risk for recipients of Snapchat's online platform to be low).

When considering these mitigations, Snap has taken into account in particular the list of possible mitigations set out in Article 35.1. For ease of reference, we have set out a table below that maps the Article 35.1 list of mitigations to the corresponding section of this report where Snap has explained how it is using that mitigation measure on Snapchat.

#	DSA Mitigation	Relevant Report Section
a	Adapting the design, features or functioning of their services, including their online interfaces.	Snapchat Design and Function
b	Adapting their terms and conditions and their enforcement.	Terms and Enforcement
c	Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Moderation
d	Testing and adapting their algorithmic systems, including their recommender systems.	Algorithmic Systems



e	Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.	Advertising Systems
f	Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.	Risk Detection and Management
g	Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.	Trusted Flaggers
h	Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.	Codes and Crisis Protocols
i	Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Transparency
j	Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate;	Protection of Minors
k	Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.	Content Authenticity



4. DSA Risk Assessment Results

In this Section of this Report, we explain the result of the risk assessment of Snapchat’s in-scope services that Snap has carried out pursuant to Article 34 of the DSA. This risk assessment was conducted in accordance with the scope and methodology explained in Section 1 of this Report. One general point to note is that these risks impact a wide range of individuals, including our Snapchatter community, victims of crime, the general public and the moderators that review the content on Snapchat. The results of this risk assessment apply to all such individuals, and where appropriate we have noted impacts that extend beyond Snapchat (including the wellness of our moderators).

It is Snap’s mission to reduce virtually all harmful content on our platform. To that end, we are continually improving our systems every single day, and are investing into (machine learning) technology, human moderation, and other measures to make our platform safer for our community. As described in the [Ongoing Risk Management](#) section below, Snap has reasonable, proportionate and effective measures to detect and manage risks on an ongoing basis.

4.1 Category 1 - Dissemination of content that is illegal or violates our terms and conditions

(Article 34.1.a / DSA Recital 80)

In this first part we report on our assessment of the risk of illegal content or content that is incompatible with our [Terms](#) being disseminated on Snapchat as required by Article 34.1.a (“Category 1”), including in particular the illegal content identified in Recital 80. In our assessment, we have taken account of the extent to which these risks are influenced by intentional manipulation, including by inauthentic use or exploitation of the service, as well as the extent to which Snapchat allows for amplification and potentially rapid and wide dissemination.

The table below provides a summary of the results of our assessment of likelihood, severity and overall potential risk prioritization, together with our conclusions given the mitigations that Snap has put in place for each Category 1 risk.

<i>Category 1 - Dissemination of content that is illegal or violates our terms and conditions (including our Community Guidelines)</i>					
<i>Category</i>	<i>Relative likelihood of risk occurring on Snapchat</i>	<i>Harm classification industry wide</i>	<i>Risk Prioritization</i>	<i>Approach</i>	<i>Conclusion</i>



4.11 Dissemination of child sexual abuse material	Extremely Low Likelihood	Severe harm industry wide	Level 1	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.12 Dissemination of illegal hate speech	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.13 Dissemination of information related to the sale of prohibited products or services (such as dangerous products, counterfeit products or illegally-traded animals)	Extremely Low Likelihood	Drugs - Severe harm industry wide	Level 1 (Drugs)	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
	Extremely Low Likelihood	Weapons - Serious harm industry wide	Level 2 (Weapons)	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
	Extremely Low Likelihood	Other goods - Significant harm industry wide	Level 3 (Other goods)	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.14 Dissemination of terrorist content	Extremely Low Likelihood	Serious harm industry wide	Level 2	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.15 Dissemination of content that infringes on intellectual property rights	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.16 Dissemination of adult sexual content	Extremely Low Likelihood (adult sexual crimes)	Adult sexual crimes - Serious harm industry wide	Level 2 (Adult Sexual Crimes)	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
	Low likelihood for other adult sexual content	Other adult sexual content - Significant harm industry wide	Level 3 (Other adult sexual content)	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations, which are being monitored to confirm prevalence continues to decline and further measures are not required.
4.17 Dissemination of content regarding harassment & bullying	Very Low Likelihood	Serious harm industry wide	Level 2	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations, with monitoring of the quality of reports resulting from new reporting actions.
4.18 Dissemination of content that glorifies self-harm.	Extremely Low	Serious harm industry wide	Level 2	Mitigated	Low Risk / Reasonable, proportionate and



including the promotion of self-injury, suicide or eating disorders	Likelihood				effective mitigations
4.19 Dissemination of content encouraging or engaging in violent or dangerous behavior	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.110 Dissemination of harmful false misinformation	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.111 Dissemination of fraud and spam	Low Likelihood (Advertising)	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations, which are being monitored to confirm prevalence continues to decline and further measures are not required.
	Very Low Likelihood (Content)				
4.112 Dissemination of information related to other illegal activities	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations, which we will keep under review (including with data from our new reporting option).

4.1.1 Dissemination of child sexual abuse material

Snap has been sensitive to the issue of dissemination of child sexual abuse material (**CSAM**) on internet platforms and services, including on Snapchat, for some time. Without mitigations, CSAM can conceivably appear in any of Snapchat's in-scope services displaying user generated content, from videos featured on Spotlight / For You to Lenses being used to add a Lens on top of CSAM content or upload an image containing CSAM (elements) in the Lens creation flow. Note, internally and externally, Snap uses the term Child Sexual Exploitation and Abuse Imagery (**CSEAI**) to refer to CSAM. Throughout this report, we will largely be using CSEAI to refer to CSAM.

Likelihood

Snap measures Policy Violating Prevalence (PVP) via random sampling of Public Stories to estimate the percent of policy-violating views. All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms). As of today, the CSEAI content represents an extremely low percentage of total views of Snaps in Public Stories. The steps Snap has taken to mitigate this risk have substantially diminished the likelihood that Snapchatters will encounter



CSEAI on Snapchat's in-scope services, such that it falls into the **Extremely Low Likelihood category**.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if CSEAI were to materialise on an online platform, the risk of harm would fall within our **'severe' category**.

Overall potential risk prioritization

Although the prevalence of CSEAI on Snapchat is considered to be at the lowest level of all our risks, due to the potential for the most severe harms to be caused by CSEAI, Snap considers CSEAI to be a Level 1 risk priority. As described in our [risk methodology section](#), we assess overall potential risk on a case by case basis and Snap reserves the option to deviate from the overall potential prioritization risk matrix we use as a guide. This is one of the cases where we have chosen to deviate.

Snap's Mitigations

Highlights

Snap prohibits any activity that involves sexual exploitation or abuse of a Teen, including sharing child sexual exploitation or abuse imagery, grooming, or sexual extortion (sextortion), or the sexualization of children. By using Snapchat, users agree under our [Terms](#) not to post, save, send, forward, distribute, or ask for nude or sexually explicit content involving anyone under the age of 18 (this includes sending or saving such images of themselves).

It is possible, despite Snap's policies and enforcement efforts, that malicious actors will find ways to circumvent Snap's enforcement mechanisms and practices in order to post CSEAI, which could then appear on Snap's public surfaces. Preventing and addressing potential CSEAI is a top priority for Snap, and is considered a "severe harm" under Snap's [Community Guidelines](#), and we apply our most swift and severe response against violators as explained in our [Severe Harms explainer](#).

As explained in the [Moderation](#) section (specifically, the section on [CSEAI](#)), we also proactively scan all Stories and Spotlight submissions using PhotoDNA and Google CSAI Match, and enforce against accounts found to be sending CSEAI. Snapchatters can also report CSEAI to us via in-app reporting options and anyone can submit a report through the Snapchat Support Site.



When Snap becomes aware that CSEAI is present on its platform, the content is immediately removed from the platform and reported to NCMEC, and we take enforcement action on the user account. This is detailed further in the [Moderation](#) and [Enforcement](#) sections of the Report. Upon notice of any of the following activity, Snap will immediately take enforcement action and report the user's account to NCMEC.

Snap works with the U.S. National Center for Missing and Exploited Children (NCMEC) and other safety experts to learn about these types of harms and how they may manifest themselves on our platform, and to report such harm to the proper authorities. Snap also has [trusted flaggers](#) to bring these and other types of harms to the attention of our trust and safety teams.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigations	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, fundamental design decisions mean, for example, that Teens are less likely to come into contact with strangers.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit CSEAI and they are strictly enforced with the most serious consequences given the risk of severe harm.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, specific proactive and reactive moderation procedures to prevent and remove CSEAI.
Algorithmic Systems	Yes, our algorithmic systems do not



Testing and adapting their algorithmic systems, including their recommender systems.	knowingly recommend CSEAI i.e. there is no 'CSEAI' interest category.
<p>Advertising Systems</p> <p>Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	Yes, other mitigations listed here also apply to our Advertising Systems.
<p>Risk Detection and Management</p> <p>Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	Yes, for example we have specific prevalence testing and a transparency report which we use to help detect and manage CSEAI.
<p>Trusted Flaggers</p> <p>Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	Yes, we cooperate with trusted flaggers in relation to CSEAI/child safety.
<p>Codes and Crisis Protocols</p> <p>Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	Yes, we cooperate with other providers through various groups e.g. EUIF, the Technology Coalition, WeProtect Global Alliance.
<p>Transparency</p> <p>Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), and how to get help in our Safety Center.
<p>Protection of Minors</p> <p>Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity</p> <p>Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	General content authenticity measures.



Conclusion

Given the severity of the harm industry-wide, Snap treats CSEAI as a Level 1 risk priority. In response, it has put in place a range of mitigation measures. This includes, in particular, our proactive content [moderation](#) which is designed to detect and prevent CSEAI from appearing on each of Snapchat's in-scope services – for example, our automated and human review on Spotlight. Our prevalence testing has allowed us to improve this proactive content moderation. As a result, we've reduced the prevalence of CSEAI on Snapchat to the lowest likelihood level. In the second half of 2022, the proactive moderation detected and actioned 94% of the total child sexual exploitation and abuse violations reported in our Transparency Reports.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for the dissemination of CSEAI.

4.1.2 Dissemination of illegal hate speech

Public spaces displaying user-generated content have the potential for the dissemination of illegal hate speech. We recognise that, [without mitigations](#), hate speech could conceivably appear in any of Snapchat's in-scope services displaying user-generated content, from videos featured on Spotlight / For You, to edits made to place labels on the Snap Map to changes to the names and content of our Lenses and ads using hate speech or demeaning representations of a particular culture, race or ethnicity.

Likelihood

Snap is sensitive to the issue of hate speech on internet platforms, as well as the damaging effects hate speech can have on a community. Thankfully, hate speech is rarely found on the public surfaces of Snapchat. Our prevalence testing showed that hate speech accounted for an extremely low percentage of total views of Snaps in Public Stories in August 2023 (see our [Prevalence](#) chapter). Recent assessments by European authorities have confirmed the low incidence of hate speech across Snapchat. A July 2023 report issued by ARCOM indicates that NGOs and other Trusted Flaggers submitted zero reports related to hateful content on Snapchat over the course of the year.⁹ Snap Lenses are also not a popular medium for hate speech, with an extremely low percentage of all reviewed Lenses falling within the Hate Speech category (**all of which were rejected at submission**). The steps Snap has taken to mitigate this harm have substantially diminished the likelihood that Snapchatters will encounter hate speech on Snapchat's in-scope services, such that Snap considers this to fall within the **extremely low likelihood** category.

⁹ ARCOM, 'Lutte contre la diffusion de contenus haineux en ligne', July 2023, [url](#).



Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if illegal hate speech were to materialise on an online platform, the risk of harm would fall within our ‘**significant harm**’ category.

Overall potential risk prioritization

Because Snap qualifies hate speech as ‘significant’ in terms of severity but ‘lowest’ in likelihood given the lowest relative prevalence on the platform, Snap would consider hate speech to fall within our **Level 3 risk prioritization** category.

Snap's Mitigations

Highlights

Hate speech is strictly prohibited on Snap, and we work with subject matter experts, such as the FSM - Freiwillige Selbstkontrolle Multimedia Diensteanbieter e.V. in Germany, among others, to help identify hate speech, remove it from the platform, and take appropriate action against users who post such content. Snap also works with law enforcement, where appropriate, to take action against users who post illegal content. We are constantly learning, and will calibrate wherever necessary to ensure that our products and policies function to keep Snapchatters safe.

As explained in the [Terms](#) section of this Report, “hate speech” as defined in Snap’s [Community Guidelines](#), includes both illegal and legal but harmful speech. As such, Snap’s definition of hate speech is more inclusive than most legal definitions of hate speech, because Snap wants to tackle harmful (but) legal speech as well.

It is possible, despite Snap’s terms and policies prohibiting such practices, as well as Snap’s enforcement mechanisms, that malicious actors will find ways to circumvent Snap’s enforcement mechanisms and practices in order to post illegal hate speech, which could then appear on Snap’s public surfaces.

As explained in our [Moderation](#) and [Enforcement](#) sections of our report, on our potentially high-reach surfaces, like Spotlight and For You, we take a proactive approach to moderating any content that may violate our rules on hate speech. Our in-app reporting tool also allows users to directly report hateful content or activities that support terrorism or violent extremism. When hateful content is reported, our teams will remove any violating content and users who engage in repeated or egregious violations will have their account access locked. Lenses identified with



hate speech were rejected when found during submission and disabled in For You upon review if subsequently identified. As an additional measure, we encourage Snapchatters to block any users who make them feel unsafe or uncomfortable. Snap removes hate speech as soon as it becomes aware of it, and will disable accounts dedicated to hate speech, hate symbols or groups, or the glorification of hate groups or members of a hate group. Our median turnaround time for hate speech reports in the second half of 2022 was **30 minutes**.

Our adaptation of Snapchat's in-scope services to include moderation and enforcement tools and processes also encompasses service-specific adaptations to address illegal or violating content such as illegal hate speech.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, our in-scope services have been adapted to include proactive moderation for illegal hate speech.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit illegal hate speech and they are strictly enforced. Our median turnaround time for hate speech reports in the second half of 2022 was 30 minutes.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, specific proactive and reactive moderation procedures to prevent and remove illegal hate speech.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend illegal hate speech i.e. there are no interest categories relating



	to hate speech.
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	Yes, other mitigations listed here also apply to our Advertising Systems.
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	Yes, for example we have specific prevalence testing and transparency reporting which we use to help detect and manage illegal hate speech.
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	Yes, we cooperate with trusted flaggers in relation to illegal hate speech, in particular Licra in France and the Department for Internet Services and Social Media.
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	Yes, we cooperate with other providers through various groups in relation to illegal hate speech, in particular Snap is currently a signatory of the EU Code of Conduct to counter illegal hate speech online.
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	Yes, we provide guidance on harms (see also Annex) and how to get help in our Safety Center.
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use</p>	General content authenticity measures.



functionality which enables recipients of the service to indicate such information.	
---	--

Conclusion

Snap considers illegal hate speech a Level 3 risk prioritization. In response it has put in place a range of mitigation measures. These include in particular our alignment to the EU Hate Speech code of practice and our proactive content moderation which is designed to detect and prevent illegal hate speech from reaching a broad audience on Snapchat's in-scope services. We monitor the prevalence of hate speech in general via our [Prevalence Testing](#) and external reporting which we publish in our [Transparency Reports](#). As a result of the mitigation measures Snap has taken, hate speech continues to be an extremely low prevalence risk.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for the dissemination of Illegal Hate Speech.

4.1.3 Dissemination of information related to the sale of prohibited products or services (such as dangerous products, counterfeit products or illegally-traded animals)

The dissemination of information related to the sale of prohibited products or services is a pervasive challenge for digital platforms. On Snapchat, without mitigations, information related to the sale of prohibited products or services could conceivably appear in any of Snapchat's in-scope services displaying user generated content, including information in videos featured on Spotlight / For You and information about places to facilitate sales on Snap Map. It may also include ads promoting the sale of illegal goods, e.g. drugs or malicious content/malware.

Likelihood

We are encouraged that the prevalence of information related to the sale of prohibited products or services on Snapchat is amongst the lowest of all risks. As a reminder, all of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms). According to our analysis, in August 2023, content related to the sale of illegal goods is measured at an extremely low rate of prevalence (PVP); content related to regulated (though not illegal) activities is measured at an extremely low prevalence rate (PVP), see our [Prevalence](#) chapter. These data provide evidence that the likelihood of encountering such violating content on Snapchat's in-scope services falls within the **extremely low likelihood category**.



Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if information related to the sale of prohibited products or services were to materialise on an online platform, the risk of harm would fall within our:

- i. **'Severe harm' category** where a credible threat to human life, safety, or well-being existed, in particular the depiction or use of, or attempts at buying, selling, exchanging, or facilitating sales of illegal lethal drugs.
- ii. **'Substantial harm' category** where there are attempts to buy or sell weapons and depicting or brandishing weapons in a threatening or violent context; and
- iii. **'Significant harm' category** with respect to the use of online platforms for selling other illegal goods or services

Overall potential risk prioritization

Thankfully, this is not a common issue on Snap. Our [prevalence testing](#) revealed that communication around Illegal goods and activities made up an extremely low fraction of Policy Violating Prevalence (PVP) . However, due to the severity of some potential products and services (such as communication around dangerous or illicit drugs), prevalence is not the determinative factor for Snap's prioritization of this issue. Snap prioritizes severe harm and legal compliance over prevalence on the platform, and for this category has decided to deviate from the standard risk framework.

Snap would consider the overall risk of this type of content to be in the Level 1 category (due to the level of severity) in cases where it concerns dangerous and illicit drugs, or any other prohibited products or services that pose a threat to human life, safety, or well-being. Snap considers this issue a Level 2 overall potential risk prioritization in relation to weapons and a Level 3 potential overall risk prioritization in relation to other prohibited products and services.

Snap's Mitigations

Highlights

Snap is sensitive to the issue of internet platforms being misused to advertise or sell prohibited products or services. The steps Snap has taken to mitigate this harm have substantially



diminished the likelihood that Snapchatters will find information related to prohibited products or services on our platform.

Snap's [Terms](#) prohibit users from posting content that's illegal in their jurisdiction or using Snap for any illegal activity. The Community Guidelines also prohibit promoting, facilitating, or participating in criminal activity, such as buying, selling, exchanging, or facilitating sales of illegal or regulated drugs, contraband (such as child sexual exploitation or abuse imagery), weapons, or counterfeit goods or documents. They also prohibit promoting or facilitating any form of exploitation, including sex trafficking, labor trafficking, or other human trafficking. Snap also prohibits the illegal promotion of regulated goods or industries, including unauthorized promotion of gambling, tobacco or vape products, and alcohol.

It is possible, despite Snap's terms and policies prohibiting such practices, as well as Snap's enforcement mechanisms, that malicious actors will find ways to circumvent Snap's enforcement mechanisms and practices in order to post information related to the sale of prohibited products or services, which could then appear on Snap's public surfaces.

As explained in the [Moderation](#) section of the Report, we have proactive and reactive moderation processes in place to detect and moderate content relating to the sale of illegal goods and services, and we have aggressively focused on enforcement of severe and serious harms. For example, during the second half of 2022, we enforced against more than 290,000 pieces of content and more than 200,000 accounts relating to drug content, based on both proactive detection of drug sales content and reports in-app and through the support site. Moreover, our enforcement of accounts for violating our Community Guidelines relating to weapons has doubled during the same period.

We also use [Prevalence Testing](#) to continuously improve our moderation. As one example, we have used violating Snaps relating to illegal Drug sales, which we consider a Level 1 potential overall risk prioritization for illegal goods and services, to help train our proactive machine learning detection models with the purpose of detecting novel drug content. This has increased proactive detection and enforcement volumes for Drugs. By examining violating labeled Snaps, we have also identified opportunities to close gaps, such as improving Optical Character Recognition (OCR) models and emoji detection relating to drug sales.

As explained in the [Enforcement](#) section of our Report, Snap complies with relevant legal requirements to remove content about the sale of illegal goods and services, and takes appropriate action against egregious or repeat violators. Snap works with law enforcement, safety organizations, and subject matter experts to continue to educate ourselves and our community, and to take appropriate action where these threats may arise on our platform.

When we identify violators engaging in the attempted buying, selling, exchanging, or facilitating sales of dangerous and illicit drugs, we disable their accounts and, in some instances, refer the



conduct to law enforcement. For less severe harms, a user will be warned and their content removed. Repeat violations will result in violators' accounts being disabled.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, design decisions, including adding proactive moderation to Snapchat's in-scope services, make it difficult for the sale of prohibited products or services to reach a large audience.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, terms prohibit the sale of prohibited products or services and they are strictly enforced with the most serious consequences.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, specific proactive and reactive moderation procedures to prevent the sale of prohibited products or services.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend content concerning the sale of prohibited products or services content i.e. there is no interest category for this content.
Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the	Yes, other mitigations listed here also apply to our Advertising Systems.



<p>presentation of advertisements in association with the service they provide.</p>	
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, for example we have specific prevalence testing and transparency reporting which we use to help detect and manage information related to the sales of prohibited products and services.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to the sale of prohibited products or services, in particular the Danish Safety Technology Authority.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups and share signals, especially in relation to drug dealers with the EU Internet Forum which has recently expanded its work to tackle drug sales online.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers, Family Center, reporting, and guidance.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>



Conclusion

Despite the very low prevalence, we consider the overall risk of the dissemination of the sale of dangerous or illicit drugs, or any other prohibited products or services that pose a threat to human life, safety, or well-being, to be in the Level 1 category due to the level of severity. Snap considers that the sale of weapons poses a Level 2 overall potential risk, and a Level 3 potential overall risk in relation to other prohibited products and services. Snap has taken steps to mitigate these harms, which has substantially diminished the likelihood that Snapchatters will find information related to prohibited products or services on Snapchat's in-scope services. Snap continues to invest significant resources to further combat these harms, and further reducing this risk remains a top priority for Snap.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of information related to the sale of prohibited products or services.

4.1.4 Dissemination of terrorist content

Sadly, online influences have been depicted as major drivers for the propagation and adoption of extremist ideologies, which often contain an element of collective grievance, and subsequent acts of violence. It is conceivable that, without mitigations, bad actors could disseminate terrorist content on Snapchat, as with any other online platform. This could include, in particular, terrorist content appearing in videos featured on Spotlight / For You and extremist content and individuals promoted via Public Profiles.

Likelihood

According to our testing, terrorist content is measured to have an extremely low prevalence on our platform—the lowest percentage of any of our risk categories (see our [Prevalence](#) chapter). In the second half of 2022, for example, we removed 23 accounts in the EU for violations of our policy prohibiting terrorist and violent extremist content, as recorded in our [Transparency Reports](#), and these low numbers have remained consistent since. We have also sought independent analysis via third party intelligence vendors that track extremist activity online who have verified that Snapchat does not fall into the top 100 communications platforms used by extremist groups to communicate.

These data indicate that the likelihood of encountering terrorist content on Snapchat is within the **lowest category**.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. They show an extremely low prevalence on Snapchat and our 2024 Report shows similar information.



Severity

Snap has assessed information published by governments and other third party sources and considers that, if information related to terrorist content were to materialise on an online platform, the risk of harm would fall within our ‘severe harm’ category due to the high threat to human life, safety, or well-being.

Overall potential risk prioritization

Due to the very low prevalence of extremist content on Snapchat, the overall risk would normally be assessed to be Level 3. However, due to the consequences of terrorism and potential for severe harm to human life, safety or wellbeing,, Snap has decided to deviate from the standard risk framework, and has marked terrorist content as Level 2 (and Snap will always consider the overall risk to be Level 1 risk prioritization where there is an immediate risk to human life, safety, or well-being.

Snap's Mitigations

Highlights

Snap is sensitive to the issue of dissemination of terrorist content on internet platforms and services. The steps Snap has taken to mitigate this risk have substantially diminished the likelihood that Snapchatters will encounter terrorist content. In addition, unlike many of our peers, Snap does not offer an open news feed where unvetted publishers or individuals have an opportunity to broadcast terrorist content, nor does Snapchat offer a ‘reshare’ functionality that would encourage virality, and does not allow user-generated content to gain wide viewership without going through human review.

Snap’s [Terms](#) and [Community Guidelines](#) expressly prohibit terrorist organizations, violent extremists, and hate groups from using our platform. We consult the expertise and work of civil rights organizations, human rights experts, law enforcement agencies, NGOs, and safety advocates to help enforce these Guidelines. Such expert knowledge comes from sources such as the Anti-Defamation League, the Southern Poverty Law Center, the Election Integrity Partnership, the Atlantic Council, the Stanford Cyber Policy Center, the members of Snap’s Safety Advisory Board, and individual domain experts (including a former Ambassador to the UN Human Rights Council, leading digital rights scholars and advocates, former regulators and policymakers, and geopolitical experts). We are constantly learning, and will calibrate wherever necessary to ensure that our products and policies function to keep Snapchatters safe.

Our prohibitions against Terrorism and Violent Extremism extend to all forms of content that promotes terrorism or other violent, criminal acts committed by individuals or groups to further ideological goals. These rules also prohibit any content that promotes or supports foreign terrorist organizations or extremist hate groups—as designated by credible, third-party experts—as well as recruitment for such organizations or violent extremist activities.



It is possible, despite Snap’s terms and policies prohibiting such practices, as well as Snap’s enforcement mechanisms, that malicious actors will find ways to circumvent Snap’s enforcement mechanisms and practices in order to post terrorist content, which could then appear on Snap’s public surfaces.

As explained in our [Moderation](#) section, on our high-reach surfaces, like Spotlight and For You, we take a proactive approach to moderating any content that may violate these rules.

Our in-app reporting tool allows users to directly report hateful content or activities that support terrorism or violent extremism. Snap removes such content, disables accounts, and cooperates with law enforcement as such issues may arise; see our [Enforcement](#) section for more information. Users engaged in terrorist activities or violent extremism will lose account privileges. Accounts we discover engaging in the following activity will immediately be disabled and where appropriate, reported to law enforcement.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat’s in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **[link in the left hand column to a full summary](#)** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, Snapchat is not an attractive platform for terrorist content because it is difficult to reach a large audience on Snapchat, and Snap proactively moderates Snapchat’s in-scope services that provide an opportunity to reach a larger audience. As a result, we experience very few instances of terrorist content on Snapchat.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit terrorist content and they are strictly enforced with the most serious consequences.
Moderation	Yes, specific proactive and reactive moderation procedures to prevent and



<p>Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>remove terrorist content and accounts.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend terrorism content i.e. there is no 'terrorism' interest category.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, for example we have specific prevalence testing and transparency reporting which we use to help detect and manage terrorist content. We have also sought independent analysis via third party intelligence vendors (SITE and MEMRI) that track extremist activity online who have verified that Snapchat does not fall into the top 100 communications platforms used by extremist groups to communicate.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>No, we don't have a specific trusted flagger group we currently work with on terrorism content in the European Union. This is due to the low prevalence of terrorist content on Snap.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups, including the EU Internet Forum (EUIF), that consider terrorist content. Note, due to the low prevalence of terrorist content on Snap, we do not participate in primary multi stakeholder organization: The Global Internet Forum to Counter Terrorism (GIFCT).</p>
<p>Transparency</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to</p>



<p>Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

Despite the very low prevalence, we consider the potential risk of the dissemination of terrorist content to be in the Level 2 risk prioritization category due to the level of severity. Snapchat's design and its proactive detection measures make Snap a very unpopular place for the dissemination of terrorist content. The dissemination of terrorist content has the lowest likelihood of all risks with only 23 accounts being deleted for terrorist content in the second half of 2022.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of terrorist content. Snap monitors this category to confirm whether further mitigating measures might be required.

4.1.5 Dissemination of content that infringes on intellectual property rights

All platforms that allow users to upload and share media have the potential for those users to choose to upload material that they do not have the right to share (for example, clips from films), or that users may create and share original material that infringes on another party's intellectual



property (for example, a Lens using a copyrighted character). Without mitigations, such material could conceivably appear on Snap's public surfaces including in particular videos on Spotlight and For You.

Likelihood

Snapchat's platform architecture does not favor the mass distribution of unauthorized copyrighted content. Snapchat does not have a live-streaming feature. A typical "Snap" is 10 seconds or less and expires in 24 hours. Content creation and consumption on Snapchat favors very short, original content and in-the-moment communication between friends; other platforms are more attractive to those seeking to flout intellectual property law.

Snap maintains a public [Transparency Report](#) which includes data on enforcement actions related to intellectual property infringement.

In H1 of 2022:

- We received 558 copyright notices; 78% of those requests led to the removal of some content.
- We received 96 trademark notices; 29% of those requests led to the removal of some content.

In H2 of 2022:

- We received 905 copyright notices; 73% of those requests led to the removal of some content.
- We received 172 trademark notices; 13% of those requests led to the removal of some content.

This data shows an overall increase in reports of intellectual property issues over time but a consistently low prevalence in absolute terms. As a result, Snap considers the likelihood of encountering content that infringes intellectual property on Snapchat is within the **extremely low likelihood category**.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. They show an extremely low prevalence on Snapchat, and our 2024 Report shows similar information.

Severity

Snap has assessed information published by governments and other third party sources and considers that if information infringes intellectual property rights were to materialise on an online platform, the risk of harm would fall within our 'significant harm' category.



Overall potential risk prioritization

We consider the dissemination of content that infringes on intellectual property rights is a **Level 3 overall potential risk** on Snapchat. We take reports seriously and the reported infringement of intellectual property often leads to content removal or, in some cases, the deletion of the user's account.

Snap's Mitigations

Highlights

Snapchat respects the intellectual property of others, and expects our users to do the same. As we explain in the [Terms](#) section of the Report, Snap's [Terms of Service](#) clearly prohibit the use of Snap's services to infringe on someone else's intellectual property rights.

The [Enforcement](#) section of the Report states that if someone believes that any content on Snapchat infringes their intellectual property (IP), they can let us know via our reporting menu or online forms for [Copyright Infringement](#) or [Trademark Infringement](#). Snap honors copyright laws, including the Digital Millennium Copyright Act and European Copyright Directive, and takes reasonable steps to expeditiously remove from our Services any infringing material that we become aware of. If Snap becomes aware that a user has repeatedly infringed copyrights, we will take reasonable steps within our power to suspend or terminate the violator's account.

Snap respects the doctrine of "fair use" (where applicable) i.e., that there are certain circumstances (such as news reporting, social commentary on issues of public interest, criticism, parody, or education) where copyrighted material could be distributed without permission from, or payment to, the copyright holder.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function	Yes, content on Snapchat is typically short in nature, the average Snap is 10 seconds,



<p>Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>and proactive moderation and reporting tools help with the detection of IP infringing material.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, terms prohibit IP infringements and they are strictly enforced.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to prevent and remove content that infringes intellectual property rights.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Our algorithmic systems do not knowingly recommend content that infringes intellectual property rights i.e. there are no interest categories relating to specific intellectual property.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have a notice procedure to flag and enable us to respond to intellectual property infringements.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>General trusted flagger measures.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Not applicable. We respond to reports of infringement on an individual basis.</p>
<p>Transparency</p>	<p>Yes, we warn users not to publish content that infringes on intellectual property rights</p>



Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	and we have an easily accessible reporting tool.
Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.	No specific measures relating to the protection of minors for this risk.
Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.	General content authenticity measures.

Conclusion

We consider the overall risk of the dissemination of IP infringing content to be significant. Snap has taken steps to mitigate these harms, which has substantially diminished the likelihood that Snapchatters will encounter IP infringing material. These mitigations include product and design measures like short content retention periods, some proactive moderation, and notice-and-takedown procedures.

We have concluded therefore that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of content that infringes intellectual property rights. Snap monitors this category to confirm whether further mitigating measures might be required.

4.1.6 Dissemination of adult sexual content

Estimates as to the volume of adult sexual content on the Internet vary, but some historical studies have considered that around 4% of websites, 13% of web searches, and 20% of mobile searches were related to adult sexual content¹⁰. As such it is conceivable that, without mitigations, this content could also appear on any of Snapchat’s in-scope services including, in particular, videos on Spotlight, For You, promoted on Snapchatter Public Profiles, features as part

¹⁰ Ogas, O. and S. Gaddam (2012), Boston University, *A Billion Wicked Thoughts: What the Internet Tells Us About Sex and Relationships*; and Google Inc, Columbia University and Carnegie Mellon University, *A Large Scale Study of Wireless Search Behaviour*, 2005.



of our Lenses or as places on the Snap Map, and could be the subject of advertisements via Snap Ads.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms). As shown in our [Prevalence Testing](#), adult sexual content had a low prevalence of in August 2023. Although this is significantly lower than the prevalence of adult content on the Internet in general, and low prevalence in absolute terms compared to the total volume of content on Snapchat, it is higher likelihood than the other risks that we have tracked. Further, as reported in our [Transparency Reports](#), in the first half of 2022, Sexually Explicit Content accounted for 76.6% of the total content we enforced. This was reduced to 67.9% in the second half of 2022 and, although we expect this to decrease further in our [Transparency Reports](#) for 2023, it is still likely to remain the largest category of content enforcements. As a result, we place adult sexual content in our **Low Likelihood** category, however, we are not aware of notable volumes of adult sexual crimes and have assessed this to fall within our **Extremely Low Likelihood** category.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and third party sources and considers that the severity of this risk varies depending on the nature of the content as follows:

- If information relating to adult sexual offences were to materialise on an online platform, the risk of harm would fall within our **'serious harm'** category due to the significant threat to human life and well-being, abusing people's fundamental rights and dignity and involving the criminal exploitation of vulnerable people.
- If information relating to sexually explicit content or depictions of nudity were to materialise on an online platform, the risk of harm would fall within our **'significant harm'** category. This content, while significant, does not pose the same severity of risk as adult sexual offences.

Overall potential risk prioritization

The overall potential risk of this adult sexual content depends primarily on severity of the issue. Overall, we consider the dissemination of sexual crimes and offenses on Snapchat's in-scope services to be a **Level 2** potential overall risk prioritization. We consider the dissemination of sexually explicit content or depictions of nudity to be a **Level 3** potential overall risk prioritization.



Snap's Mitigations

Highlights

Snap's [Terms](#) prohibit promoting, distributing, or sharing pornographic content, as well as commercial activities that relate to pornography or sexual interactions (whether online or offline). Breastfeeding and other depictions of nudity in non-sexual contexts are generally permitted. As this can be a challenging area, we make available additional guidance on sexual conduct and content that violates our Community Guidelines [here](#).

As explained in the [Moderation](#) and [Enforcement](#) sections of the Report, we have proactive and reactive moderation processes in place to detect and moderate adult sexual content. Our in-app reporting tool allows users to directly report adult sexual content which our teams will then remove if confirmed as violating. In addition to measures taken against sexual crimes and sexually explicit content (and nudity), Snap also takes steps to limit the prevalence and recommendation of sexually suggestive content.

Our [Prevalence Testing](#) has had a very significant impact in reducing the extent to which Adult Sexual Content is present on Snapchat to low levels. We expect this to decrease further as recent sampling data shows even further declines in violating views of this content.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect to Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, our online platforms have been designed to limit the prevalence of sexually suggestive content.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit adult sexual content and they are strictly enforced.
Moderation	Yes, specific proactive and reactive



<p>Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>moderation procedures to prevent and remove sexual content.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend adult sexual content i.e. there is no 'adult sexual content' interest category.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, for example we have specific prevalence testing and transparency reporting which we use to help detect and manage adult sexual content.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to Non-Consensual Intimate Image Abuse (NCII), notably Stop Fisha in France.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups, including in particular the EU Internet Forum (EUIF) which has expanded its remit to also tackle the trafficking of human beings (which is often driven by sexual crimes or pornography).</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools,</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to</p>



<p>tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

Adult sexual content is our highest prevalence, largest enforced content category issue on Snapchat’s in-scope services, and due to the potential for significant harm, falls within our Level 3 overall potential risk prioritization. We considered sexual crimes to fall within our Level 2 overall potential risk prioritization given the risk of serious harm, although this content falls within the Extremely Low Likelihood category. We have dedicated substantial resources and taken significant steps, including introducing an entirely new proactive detection mechanism, to ensure a low prevalence of adult sexual content. We expect the prevalence and reported content percentages for adult sexual content to decrease further: our most recent sampling data suggests there will be further declines in policy violating prevalence for adult sexual content as we continue to evolve our detection mechanisms. This significant proactive detection work was undertaken as a direct response to us detecting an uptick in the prevalence of adult sexual content on Snapchat, which demonstrates the effectiveness of our risk detection and management framework and procedures.

With these adjustments, and the other specific mitigations listed above, we have concluded that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of adult sexual content. However, we continue to carefully monitor this risk category to confirm the expected further fall in adult sexual content prevalence.

4.1.7 Dissemination of content regarding harassment & bullying

Unfortunately, harassment and bullying that have always been a persistent problem in schools and workplaces have crossed over to the online environment with the growth of the Internet. Without mitigations, this content could conceivably appear on Snapchat’s in-scope services.



Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms). Snap's internal [prevalence](#) measurement reveals that harassment (which includes bullying) has a very low PVP in August 2023.

Our [Transparency Report](#) shows that harassment is an issue that leads to a significant volume of content and account enforcements. In the first half of 2022, we received approximately 150,994 reports related to harassment in the EU, of which roughly 31% resulted in some enforcement action. In the second half of 2022, we have received significantly more reports, 566,708, relating to harassment. This significant increase was due to the introduction of new reporting options.

We also noted that in the second half of 2022, only 13% resulted in some enforcement action. This means that while the new reporting tools surfaced more reports, they appear to have reduced the overall quality of those reports. We are investigating the cause of this but we believe this is due to "harassment and bullying" being the first reporting option in the reporting menu. We believe that we receive a significant number of reports of "harassment" in bad faith, where an individual reports public, amplified content because they don't like it - for example a user may dislike that one prominent person is criticizing the actions or views of another prominent person, and report their content as "harassment." In other instances, we may receive a report alleging "harassment" where there is not sufficient information to take enforcement action. This accounts for significant differences between reporting and enforcement rates in general, and in particular for the new reporting tools where harassment and bullying may be selected as the first option regardless of whether the report concerns such behavior.

It is important to note, however, these enforcement metrics refer to user reports of content across Snapchat. The bulk of true harassment occurs in non-public surfaces **which are out of scope of our DSA risk assessment.** Only 6.7% of enforcements relate to Snap's public platforms (4.1% Snap Map, 2.5% Public User Story and 0.1 % Spotlight Snap).

As a result, in respect to Snapchat's online platforms within the scope of our DSA risk assessment only, we consider harassment and bullying to be of **Very Low Likelihood**.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.



Severity

Snap has assessed information published by governments and other third party sources and considers that if information relating to adult sexual offences were to materialise on an online platform:

- i. Where harassment and bullying content involves both (1) harms that risk significant damage to the physical or emotional well-being of Snapchatters, and (2) the imminent, credible risk of severe harm, including threats to human life, safety, and well-being, it would fall within our **'severe harm' category**; and
- ii. For other cases, we consider harassment and bullying content to fall within our **'serious harm' category**.

Overall potential risk prioritization

In general, we have assessed the dissemination of content regarding harassment & bullying to fall within our Level 2 potential risk prioritization category. However, all situations where there is risk of significant damage and an imminent risk of severe harm, are considered to fall within our overall Level 1 potential risk prioritization category.

Snap's Mitigations

Highlights

Snap's policies prohibit a range of content or behavior that harasses individuals, including: (1) harassment and bullying in general, (2) behaviour that constitutes or promotes sexual harassment; or (3) behaviour that constitutes non-consensual intimate content (i.e., production and/or distribution). When we consider whether to allow content for algorithmic recommendations, we also apply [additional rules](#).

In practice, where we algorithmically recommend content on our online platforms, we take proactive measures to stop the dissemination of content that includes harassment & bullying. We use a mix of automation (such as abusive language detection, image recognition models, and account history) and human review to enforce our guidelines as explained in the [Moderation](#) section of this Report. In our [Enforcement](#) section, we also explain the significant resources devoted to preventing the dissemination of content that includes harassment & bullying. Any content anywhere on Snapchat can be reported in-app or on our web site, and "harassment" is one of the reporting reasons offered, and as reported above, this includes new reporting options for user profiles.

Importantly, our [Transparency Report](#) shows that the median turnaround time for a harassment report is **6 minutes**. If content that constitutes harassment and bullying is reported to us, we respond very swiftly with appropriate action.



Additionally, when we learn of content suggesting that there is an emergency situation involving imminent danger of death or serious bodily injury involving any person, we will proactively escalate the report to law enforcement. We have established channels for referring such content to the FBI in the U.S. and Interpol in the rest of the world.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, fundamental design decisions mean that content constituting harassment and bullying can be easily reported.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit harassment & bullying. This is explained to users clearly in our Harassment & Bullying explainer with guidance on how we apply this policy.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, we have specific moderation procedures to prevent and remove harassment & bullying content. Our transparency report shows a very swift median turnaround time of 6 minutes.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend harassment & bullying i.e. there is no 'bullying and harassment' interest category.
Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the	Yes, other mitigations listed here also apply to our Advertising Systems.



<p>presentation of advertisements in association with the service they provide.</p>	
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for harassment & bullying. Our Safety Advisor Board also has several anti-bully experts which we call on for independent review and expertise.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with many trusted flaggers in the EU in relation to child safety, including for example E-Enfance in France.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>We are not working with other providers on harassment and bullying specifically. However, we work with several groups in relation to child protection in general, including in relation to the new EU AAD Code.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center. We provide local resources related to bullying and harassment. In France for example we direct users to E-Enfance. The new national number against digital violence, free for children and adolescents facing problems related to their digital use-- 100% anonymous, free and confidential.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents and teens, including the safety measures and resources highlighted in the Transparency mitigation section above, such as the Harassment and Bullying explainers.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons,</p>	<p>General content authenticity measures. We have applied an AI sparkle icon in specific situations, such as our Bitmoji Backgrounds. We assess whether to include such an icon</p>



objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.

on a case by case basis, considering whether generated images are photorealistic.

Conclusion

Harassment and bullying is the second most prevalent issue faced by Snapchat. However, it is still a low likelihood in absolute terms, particularly **in respect of Snapchat's in-scope services only**. Where there is a risk of severe harm, we consider bullying and harassment has a Level 1 overall potential risk prioritization. In general, we consider the dissemination of content on Snapchat's in-scope services that includes harassment & bullying is a Level 2 potential risk prioritization. In practice, we have taken significant measures to prevent harassment and bullying, including clear guidance on our rules and how we enforce them, easy to access reporting tools and very rapid response times to address violating content. New reporting options have resulted in a significant rise in reports combined with a fall in the enforcement rate and we are investigating the reasons for this.

As a result, we have concluded that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of harassment and bullying content. We are monitoring the quality of reports resulting from new reporting actions.

4.1.8 Dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders

The risk of young people encountering content that promotes eating disorders, body image dissatisfaction and distorted values and attitudes online and on social media in general has been identified in several studies. Without mitigations, this content could conceivably appear on Snapchat's in-scope services.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Our [Transparency Report](#) shows that "self harm & suicide" is an issue that leads to a moderate volume of content and account enforcements. In the second half of 2022, we received 129,785 reports related to self-harm and suicide, enforcing against approximately 20,054 pieces of content and 18,311 accounts. However, those figures relate to Snapchat in general. When we



consider Snapchat’s in-scope services specifically, i.e. the public parts of Snapchat which fall within the scope of the risk assessment obligations under Article 34, the likelihood of these public spaces being used for the dissemination of content that glorifies self-harm is even lower. In 2023, we rejected 4248 Snaps on Spotlight and 110 Snaps on For You. Our [prevalence testing](#) identified a very low prevalence of content that glorifies self-harm and this fell to a negligible amount by August 2023. Only an extremely low percentage of all Lenses have ever been found to include self-harm content and all of these were rejected before publication.

As a result, we consider the risk of dissemination of content glorifying self-harm to fall within our **Extremely Low Likelihood category** for Snapchat’s in-scope services.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if information relating to self-harm were to materialise on an online platform, we consider the severity of harm risked from such content (including content relating to self-injury, suicide or eating disorders) to fall within our “**serious harm**” category, Where the dissemination of content that indicates an imminent, credible risk of severe harm, including threats to human life, safety, and well-being, we consider the severity of harm risked to fall within our **severe harm category** (as explained in our [severe harm explainer](#)) In practice, we devote enforcement resources to preventing the dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders.

Overall potential risk prioritization

Although the prevalence of content that glorifies self-harm on Snapchat’s in-scope services is considered to be at the lowest level of all our risks, due to the potential for severe and serious harms to be caused, we have chosen to elevate the risk prioritization for these risks. Snap will always consider the dissemination of content that indicates an imminent, credible risk of severe harm, including threats to human life, safety, and well-being, as Level 1 overall potential risk prioritization (as explained in our [severe harm explainer](#)), and we devote significant resources to combatting this type of harm. Other content relating to self-harm (including content relating to self-injury, suicide or eating disorders) are also classified as a **Level 2** overall risk prioritization. As described in our [risk methodology section](#), we assess overall potential risk on a case by case basis and Snap reserves the option to deviate from the overall potential risk prioritization matrix we use as a guide. This is one of the cases where we have chosen to deviate.



Snap's Mitigations

Highlights

Snap's [Terms](#) prohibit the dissemination of content that promotes self-harm and suicide.

On Snapchat's in-scope services where we algorithmically recommend content, we take proactive measures to stop the dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders. We allow some discussion (such as news or public issue commentary) of self-harm, suicide, or eating disorders, when the discussion is not glorifying such behavior. Even so, we mark this content as "sensitive" internally and adjust our algorithmic systems to limit recommendations of this kind of content.

As described in our [Moderation](#) and [Enforcement](#) sections of this Report, we use a mix of automation (such as abusive language detection, image recognition models, and account history) and human review to enforce our guidelines. Any content anywhere on Snapchat can be [reported in-app](#) or on our web site, and "self-harm and suicide" is one of the reporting reasons offered. Our [Transparency Report](#) shows that the median turnaround time for a "self-harm and suicide" report is **24 minutes**.

When we learn of content suggesting that there is an emergency situation involving imminent danger of death or serious bodily injury involving any person, we will proactively escalate the report to law enforcement. We have established channels for referring such content to the FBI in the U.S. and Interpol in the rest of the world.

We also work with third-party mental health groups to surface supportive interventions in-app. A user who searches for certain terms related to self-harm or suicide may be routed to suicide helplines in their region. For example, Snap has established a self harm flow for Lenses, which includes escalation to the Trust & Safety team, sending help resources and escalation to the Law Enforcement Operations team. Lenses that are rejected, although few in number, include help resources within the rejection reason.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.



DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat’s in-scope services have been adapted to include proactive moderation for content that promotes self-harm.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, our terms prohibit content that promotes self-harm and they are strictly enforced. Our Transparency Report shows that the median turnaround time for a “self-harm and suicide” report is 24 minutes.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, we have specific proactive and reactive moderation procedures to prevent and remove content that promotes self-harm. Snap includes help resources within rejection reasons, for example, in the Lenses submission flow.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend content glorifying self-harm i.e. there is no ‘glorifying self-harm’ interest category. We mark non-glorifying discussion as sensitive and limit the volume of recommendations.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for self-harm and suicide.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to content that promotes self-harm.</p>
<p>Codes and Crisis Protocols</p>	<p>Yes, we cooperate with other providers through</p>



<p>Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>various industry groups.</p>
<p><u>Transparency</u> Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center. In Snapchat we provide a number of tools to users. For example, if a user searches for suicide related terms we will surface our Here For You tool.</p>
<p><u>Protection of Minors</u> Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p><u>Content Authenticity</u> Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

Content that glorifies self-harm content is categorized within the Extremely Low Likelihood category for Snapchat's in-scope services. However, this content falls within our 'serious harm' category and as a result we have decided to categorize it as a Level 2 overall potential risk prioritization, even though our risk matrix would suggest a lower category. We always treat content relating to suicide and other situations involving imminent, credible risk of harm as a Level 1 overall potential risk prioritization. In response, we have significant dedicated mitigation measures, including clear prohibitions, guidance, proactive and reactive moderation, reporting tools, sensitive content recommendation limits and cooperation with trusted flaggers. We respond rapidly to reports of self-harm, with a median turnaround time of 24 minutes.



As a result, we have concluded that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of content glorifying self-harm (including the promotion of self-injury, suicide or eating disorders). Snap monitors this category to confirm whether further mitigating measures might be required.

4.1.9 Dissemination of content encouraging or engaging in violent or dangerous behavior.

Without mitigations, content encouraging or engaging in violent or dangerous behavior could conceivably appear on Snapchat’s in-scope services.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

On Snapchat, our [Transparency Report](#) shows that Threats & Violence account for a relatively modest 2.6% of all enforcement actions we took across all categories in the first half of 2022. We received 547,914 reports related to Threats & Violence. In this category, we took action against 348,790 pieces of content and enforced against 113,887 accounts. In the second half of 2022, this climbed to 753,467 reports (with action against 167,811 pieces of content and 132,915 accounts). Threats & Violence continued to account for the same percentage of all enforcement actions, 2.6%. Our [prevalence testing](#) showed that violent or disturbing content accounted for an extremely low percentage of PVP in August 2023. Our proactive content moderation has successfully evolved to reduce the prevalence of violent and disturbing content. In total, a very low percentage of all Lenses submitted were found to include violent or dangerous behavior, which were all rejected.

As a result, although all of the risks we track on Snapchat have a relatively low prevalence compared to the prevalence of these issues elsewhere online and offline, we have chosen to categorize the dissemination of content encouraging or engaging in violent or dangerous behavior as falling within **our extremely low likelihood** category relative to other risks on Snapchat’s in-scope services.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.



Severity

Snap considers that the spectrum of “encouraging or engaging in violent or dangerous behavior” can vary considerably and covers a broad range of content types:

- Content relating to imminent, credible threats such as school or other mass shooting and bombing threats, although this is mainly a US-related risk and less relevant for EU users. Snap considers credible imminent threats to human life to constitute a severe harm.
- Viral “challenges” may cause injury (for example, the “Milk Crate Challenge” of 2021). Since well before the existence of social media, some people have sought out videos of other people getting hurt. This content ranges from horrifying shock content, to relatively tame comedic pratfalls and minor injuries.

Snap has assessed information published by governments and other third party sources and considers that if information encouraging or engaging in violent or dangerous behavior were to materialise on an online platform, **these issues can vary considerably in severity, from our ‘severe harm’ category to our ‘significant harm’ category.**

Overall potential risk prioritization

Content encouraging or engaging in violent or dangerous behavior is one of the **lowest likelihood risk** categories on Snapchat and runs the gamut from urgent, credible threats to human life which we consider a **Level 1** overall potential risk prioritization (in deviation from our standard risk matrix), to unfortunate or even silly “fails” which we consider to have a **Level 3 potential overall risk prioritization.**

Snap's Mitigations

Highlights

We devote significant resources to enforcing against truly harmful or shocking content encouraging or engaging in violent or dangerous behavior.

Snap’s [Terms](#) address the dissemination of content encouraging or engaging in violent or dangerous behavior. Our algorithmic systems have special rules in place to handle violent and dangerous content. When we consider whether to allow content for algorithmic recommendations, we apply [additional rules](#).

On public surfaces where we algorithmically recommend content, we take proactive measures to stop the dissemination of content encouraging or engaging in violent or dangerous behavior. We use a mix of automation and human review to enforce our guidelines.



Any content anywhere on Snapchat can be reported in-app or on our web site, and “threats and violence” is one of the reporting reasons offered. Our [Transparency Report](#) shows that the median turnaround time for a “threats and violence” report is **24 minutes**.

We take additional measures to protect the well-being of the Snapchat community. For example, where our Trust and Safety and Law Enforcement Operations teams identify credible threats to human life, we have protocols in place for alerting local officials.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat’s in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, Snapchat’s in-scope services have been adapted to include proactive moderation of content encouraging or engaging in violent or dangerous behavior and easy reporting.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit content encouraging or engaging in violent or dangerous behavior and we have provided a specific Threats, Violence & Harm explainer which includes guidance on how we enforce this content. Our Transparency Report shows that the median turnaround time for a “threats and violence” report is 24 minutes .
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, specific proactive and reactive moderation procedures to prevent and remove content encouraging or engaging in violent or dangerous behavior.



<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend content encouraging or engaging in violent or dangerous behavior i.e. there is no 'violent or dangerous' interest category. We mark certain shocking content as sensitive and limit the volume of recommendations.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for content encouraging or engaging in violent or dangerous behavior.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers who focus on child and digital safety.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>We are not working with other providers on violent and dangerous behavior specifically. However, we work with several groups in relation to child protection in general, including in relation to the new EU AAD Code.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on harms (including a specific Threats, Violence & Harm explainer) and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons,</p>	<p>General content authenticity measures.</p>



objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.

Conclusion

Dissemination of content encouraging or engaging in violent or dangerous behavior on Snapchat's in-scope services is one of our lowest likelihood category risks. We recognize that the potential harm arising from such content can be significant and we have therefore tracked this risk with an overall Level 3 potential risk rating. We devote significant resources to enforcing against truly harmful or shocking content encouraging or engaging in violent or dangerous behavior as summarized above. Our prevalence testing shows the prevalence of this type of content to be very low and failing on Snapchat's in-scope services.

As a result, we have concluded that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of content encouraging or engaging in violent or dangerous behavior. Snap monitors this category to confirm whether further mitigating measures might be required.

4.1.10 Dissemination of harmful false information

Without mitigations, "Fake news," (online) "disinformation" and "deep fakes" could conceivably be present in videos published on Spotlight and For You, promoted in Stories on Public Profiles, in places and Snaps featured on Snap Map and in Lenses published via Lens Studio. Harmful false advertising might include ads for content that mimics the appearance or function of Snapchat features or formats or political advertising with false statements and slogans regarding important societal issues.

Likelihood

In practice, the dissemination of harmful misinformation is not common on Snapchat. In August 2023, our [prevalence testing](#) showed an extremely low prevalence of 'harmful false information'. As explained in our [Transparency Reports](#), False information accounted for only 0.1% of the total of all content enforced on Snapchat throughout both halves of 2022. We track Impersonation separately, and it similarly accounts for a very low percentage of our enforcement actions (0.3% in the first half and 0.2% in the second half of 2022). Lenses with this type of information are rarely submitted (only 0.0021% of all Lenses include misinformation, all of which have been rejected). As



a result, we have chosen to place dissemination of harmful misinformation into our **extremely low likelihood** category relative to other risks.¹¹

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if information encouraging or engaging in violent or dangerous behavior were to materialise on an online platform, it would fall within our '**significant harm**' category relative to other risks.

Overall potential risk prioritization

Harmful false information is classified as a Level 3 overall potential risk prioritization given its extremely low prevalence on Snapchat, but is a risk that we apply significant resources to mitigate against, from the design of our platform to the ways we carefully review content before it has an opportunity to reach a wide audience.

Snap's Mitigations

Highlights

In practice, Snap has observed the negative societal effects of false information circulating on other platforms and is keen to avoid this being an issue on Snapchat. We devote significant enforcement resources to preventing the dissemination of harmful false information.

Snap's [Terms](#) address the dissemination of harmful false information. Under these policies, **harmful false information** includes false or misleading content that causes harm or is malicious. It includes impersonation, as well as disinformation, misinformation, malinformation, and manipulated media that causes harm or is malicious, such as denying the existence of tragic events, unsubstantiated medical claims, or undermining the integrity of civic processes.

The design of the Snapchat app has been made to be hostile towards the dissemination of false information. Our platform design architecture makes it difficult to spread misinformation. Snapchat has made conscious design decisions to restrict the ability for content to go viral, including limiting the remix functionality and applying short default retention to content. On

¹¹ This classification is also supported by the fact that Snapchat was not included in the report issued by the European Commission: European Commission, Directorate-General for Communications Networks, Content and Technology, *Digital Services Act – Application of the risk management framework to Russian disinformation campaigns*, Publications Office of the European Union, 2023 ([url](#)).



surfaces where a broader audience can potentially be reached our proactive detection makes it difficult for misinformation to reach a large audience. Moreover, our content platform, For You, features content from approved media publishers and content creators. Our entertainment platform for user-generated content, Spotlight, is proactively and a priori moderated before content can reach a wide audience.

On Snapchat where we algorithmically recommend content, we also take proactive measures to stop the dissemination of false information. When we consider whether to allow content for algorithmic recommendations, we apply [additional rules](#). Content is “Not Eligible for Recommendation” when it contains misleading or sensationalized headlines. Also of note: there are distinctions between the content that may appear on the 4th tab (“Stories”) and 5th tab (“Spotlight”). In the Stories “For You” section, we allow political content, but only from trusted media partners, Snap Stars, and certain popular accounts, to be algorithmically recommended beyond their subscribers. In Spotlight, we do not allow political content from anyone; this surface is pre-moderated to prevent political content from achieving reach.

Beyond content from media partners and users, misinformation may come in the form of advertising. Every political, health or sensitive issue ad is reviewed by humans on the ad review team. Our ad policies require that these advertisers provide supporting documentation for all claims. We reject ads that contain unsubstantiated or false claims. During election seasons, we contract third-party fact-checking organizations, such as Poynter, to support our work. We have also partnered with numerous governments around the world to inform Snapchatters about upcoming elections and to vote. Recent examples are German, Dutch, Swedish and French election cycles. Our ad policies also require that advertisers be transparent about the paying entity; this information is displayed to the end user in the “slug” onscreen to prevent advertisers from impersonating other entities. All political ads are logged in our [political ads library](#).

As explained in our [Moderation](#) and [Enforcement](#) sections, we use a mix of automation (such as abusive language detection, image recognition models, and account history) and human review to prevent and remove content violating our [Terms](#) relating to harmful false information. Our human review teams maintain training examples of recurring harmful false information; they are familiar with the most common unfounded conspiracy theories that circulate online. When they encounter new false or ambiguous information that may relate to politics, health, or tragic events, they fact check using trusted resources. Where necessary, they escalate emergent narratives to the Platform Policy team for review. Where misinformation is being spread from an account that has been taken over or is falsely claiming to represent someone, our team works tirelessly to restore accounts to their rightful owners, and to remove accounts or content that deceives others about one’s identity.

Any content anywhere on Snapchat can be reported in-app or on our web site, and “false information” and “Impersonation” are two of the reporting reasons offered. Our [Transparency Report](#) shows that the median turnaround time for a “false information” report is **18 minutes** and



for an “impersonation” report is **4 minutes**. When considering impersonation, we allow parody that is unlikely to cause confusion; when reviewing content, our teams are trained to distinguish between these permissible activities and harmful impersonation attempts.¹²

France’s regulatory authority for audiovisual and digital communication (“ARCOM”) annually inquires on information on Snap’s measures to fight against the dissemination of false information likely to disturb order. Snap has implemented several of the recommendations provided by ARCOM resulting from previous reports. For more information, see Snap’s 2022 statement on disinformation.¹³

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat’s in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, Snapchat is not an attractive platform for spreading misinformation, in particular because it is difficult to reach a broad audience and content deletes by default. Snap has made conscious design decisions to restrict the ability for content to go viral and limiting the remix functionality to specific content types and applying short retention to content.
Terms and Enforcement	Yes, our terms prohibit misinformation. We have a specific Harmful False or Deceptive

¹² Correctly and consistently enforcing against false information is a dynamic process that requires up-to-date context and diligence. As we strive to continually improve the precision of our agents’ enforcement in this category, we have chosen, since H1 2022, to report figures in the “Content Enforced” and “Unique Accounts Enforced” categories that are estimated based on a rigorous quality-assurance review of a statistically significant portion of false information enforcements. Specifically, we sample a statistically significant portion of false information enforcements across each country and quality-check the enforcement decisions. We then use those quality-checked enforcements to derive enforcement rates with a 95% confidence interval (+/- 5% margin of error), which we use to calculate the false information enforcements reported in the [Transparency Report](#).

¹³ ARCOM, ‘Snap’s 2022 statement on disinformation’, [url](#).



<p>Adapting their terms and conditions and their enforcement.</p>	<p>Information explainer which explains our approach to enforcement. Our Transparency Report shows that the median turnaround time for a “false information” report is 18 minutes.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, we use specific proactive and reactive moderation procedures to prevent and remove misinformation. In particular, For You features content only from approved media publishers and significant content creators. Our entertainment platform for user-generated content, Spotlight, is proactively and a priori moderated before content can reach a wide audience.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend content encouraging or engaging in misinformation i.e. there is no ‘misinformation’ interest category. We take steps to prevent content with misleading or sensationalist headlines.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems. Every political, health or sensitive issue ad is reviewed by humans on the ad review team. We reject ads that contain unsubstantiated or false claims. All political ads are logged in our political ads library.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reports for harmful false misinformation.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers, our trusted flaggers may also report misinformation, but this rarely happens because of the limited amount of misinformation on the platform.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Snap has not yet signed up to be a member of the EU disinformation code. We have limited exposure to the risk and use our limited resources to focus on other codes relating to risks more relevant to Snapchat’s in-scope services. However, Snap works closely with French regulator Arcom, which monitors industry action against misinformation.</p>



<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center. This includes a specific Harmful False or Deceptive Information explainer.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

We recognise a risk of significant harm that could arise from harmful false information. In practice, Snapchat’s in-scope services have very little exposure to harmful false information. It is one of our lowest likelihood categories of risks. As a result we track this risk as an overall significant potential risk. Snapchat has significant measures in place to prevent harmful misinformation, in particular the design and function of Snapchat’s in-scope services which limits the spread of content, limits the places where user generated can reach a broader audience and targets proactive moderation at those areas to prevent harmful misinformation from becoming viral. We have a rapid response time when harmful false information does slip through. Our [Transparency Report](#) shows that the median turnaround time for a “false information” report is **18 minutes** and for “impersonation” is **4 minutes**.

As a result, we have concluded that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of harmful false information.



4.1.11 Dissemination of fraud and spam

Without mitigations, content encouraging or engaging in violent or dangerous behavior could conceivably appear on Snapchat’s in-scope services.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Our [Transparency Report](#) shows that “Spam and Fraud” reports lead to a moderate volume of content and account enforcements on our content online platforms, such as Spotlight and For You. In the second half of 2022, Fraud and Spam accounted for 10.3% of all our enforcement actions (this is an increase compared to the first half of 2022, where it accounted for 6.1% of all enforcement action) and we’ve seen a significant increase in the number of enforcements: 657,077 in the first half of 2022, up from 348,790 in the first half of 2022. Our [prevalence testing](#) shows that Fraud and Spam had a low prevalence in August 2023. The dissemination of fraud and spam is **very low** in terms of relative likelihood on Snapchat’s content platforms.

Fraud is by far the most common reason for advertising on Snapchat to be rejected. For example, in January 2023, the top three rejection reasons were all fraud related and together accounted for 57% of all rejections . The trend has continued through 2023. As a result, we consider dissemination of fraud and spam to be the **Low Likelihood** risk for our advertising systems, relative to other risks.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show a relatively low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that, if spam and fraud were to materialise on an online platform, it would fall within our “**significant harm**” category relative to the other risks we track.

Overall potential risk prioritization

We consider Fraud and Spam to fall within the **Level 3** overall potential risk category compared to other more severe harms. We devote significant resources to protecting our users from fraud and spam in user-generated content and advertising.

Snap's Mitigations

Highlights



Snap devotes enforcement resources to preventing the dissemination of content that includes fraud and spam.

Content

Snap's [Terms](#) prohibit the dissemination of fraud and spam across the full range of service recipients that create content on Snapchat: users, media partners and advertisers. When we consider whether to allow content for algorithmic recommendations, we apply [additional rules](#).

Commercial Promotions

For commercial promotion within content from media partners or users, we apply our [Commercial Content Policy](#). The Commercial Content Policy also prohibits Deceptive Content. The Commercial Content Policy also outlines rules to protect Snapchatters from potentially misleading references to Snap. Commercial content must not suggest an affiliation with or endorsement by Snap or its products.

Advertising

[Snap's Advertising Policies](#) detail the criteria that our automation and human review teams apply while considering whether to allow or reject an ad on our platform. Our advertising policies prohibit Deceptive Content. The advertising policies for [financial products and services](#) add further detail about the kind of deceptive content that is prohibited.

As explained in Section 5 of this Report, we are vigilant in our [moderation](#) and [enforcement](#) of our [Terms](#), including against fraud and spam content and ads. On Snapchat's in-scope services where we algorithmically recommend content, we take proactive measures to stop the dissemination of fraud and spam. We use a mix of automation (such as abusive language detection, image recognition models, and account history) and human review to moderate and enforce our guidelines.

Content on Snapchat can be reported in-app or on our Support Site, and "spam and fraud" is one of the reporting reasons offered. Our [Transparency Report](#) shows that the median turnaround time for a fraud and spam content report is **4 minutes**.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the



specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat is not an attractive platform for spreading fraud and spam, in particular because it is difficult to reach a broad audience, and Snapchat has made conscious design decisions to restrict the ability for content to go viral.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, our terms prohibit fraud and spam, with specific rules for commercial content and advertising, and we strictly enforce these rules. Our Transparency Report shows that the median turnaround time for a “fraud and spam” report is 4 minutes.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, we use specific proactive and reactive moderation procedures to prevent and remove misinformation. In particular, For You features content only from approved media publishers and significant content creators. Our entertainment platform for user-generated content, Spotlight, is proactively and a priori moderated before content can reach a wide audience.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend fraud or spam i.e. there is no ‘fraud’ or ‘spam’ interest category.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems. Advertising is subject to moderation before publication, with most ads subject to a human review.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific transparency reports for fraud and spam. Prevalence testing is generally not used for ads since they are prescreened and there is a higher bar for bad actors for ads since it requires payment configurations.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the</p>	<p>Yes, we cooperate with trusted flaggers, our trusted flaggers may also report fraud spam, but this is not generally the focus of their efforts.</p>



<p>implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>We are not currently members of a dedicated group or code addressing the issue of fraud and spam online. However, we are members of several organizations and trade associations that tackle online issues facing the industry.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

We address Fraud and Spam as a very low Likelihood issue in general, and low likelihood for our advertising systems, relative to other risks faced by Snapchat's in-scope services. Given the risk of significant harm arising from fraud and spam, we categorize this issue overall as a Level 3 potential risk prioritization. We handle significant volumes of enforcement and rejections every month and our prevalent testing shows this is working, with reducing prevalence of fraud and spam on Snapchat's in-scope services. However, this is an area which we continue to monitor and improve and we would like to see the trend in reducing prevalence continue.

As a result, we have concluded that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for fraud and spam. We will continue to



improve our prevalence testing and proactive moderation to confirm a downward trend in the prevalence of fraud and spam.

4.1.12 Dissemination of information related to other illegal activities

As we allow users to publish content, we recognise that without mitigations it is possible that information related to other illegal activities not already captured by our other categories above may be disseminated on Snapchat's in-scope services.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Our [prevalence](#) measurement and [transparency reporting](#) track the prevalence of known significant issues that could potentially impact online platforms, including Snapchat, as informed by our work with [Trusted Flaggers](#), [industry groups](#) and our [safety advisory board and internal cross functional working groups](#). These categories are already addressed above and we are not currently aware of other significant issues. As a result we believe the dissemination of information related to other illegal activities to fall within the lowest likelihood category relative to other risks identified by Snap. With the introduction of the Digital Services Act, we have introduced a new reporting option to report 'illegal content' in general, and we expect to use data gathered from this option to provide us with greater visibility on the prevalence of information relating to other illegal activity on Snapchat.

Severity

The extent of harm that might be risked by information relating to other illegal activity would depend on the issue. Snap has specific categories for risks concerning the dissemination of information which are most relevant to online platforms. As a result, we categorize the risk of harm in general from information relating to other illegal activity within our '**significant harm category**'. Snap would consider the issue of illegal activity to fall within our '**severe harm category**' where the content includes a credible threat to human life, safety, or well-being.

Overall potential risk prioritization

In general Snap assesses the overall potential risk of the dissemination of this type of content to be Level 3 overall potential risk prioritization. As in other cases, where any issue arises that poses an imminent and credible threat to human life, safety, or well-being, Snap treats this issue with a Level 1 overall potential risk.



Snap's Mitigations

Highlights

Snap is sensitive to the issue of internet platforms being used to engage in illegal activity in general. We believe the steps Snap has taken to mitigate harm in general have substantially diminished the likelihood that Snapchatters will find other illegal activity on our platform beyond the categories assessed in the previous dissemination issues discussed above. Unlike many of our peers, Snap does not offer an open news feed where unvetted publishers or individuals have an opportunity to broadcast illegal content, nor does Snapchat offer a 'reshare' functionality that would encourage virality, and does not allow user-generated content to gain wide viewership without going through human review.

Snap's [Terms](#) prohibit users from posting content that's illegal in their jurisdiction or using Snap for any illegal activity. Snap's [Community Guidelines](#) and enforcement strategy are driven by Snap's values and desire to facilitate a fun, positive user experience. Snap's Community Guidelines therefore prohibit both illegal activity and activity which Snap considers harmful or against our values, but which is not necessarily illegal under EU law. For this reason, Snap is likely over-inclusive on its policies against illegal activity.

It is possible, despite Snap's terms and policies prohibiting such practices, as well as Snap's [moderation](#) and [enforcement](#) mechanisms, that malicious actors will find ways to circumvent Snap's enforcement mechanisms and practices in order to engage in illegal activity, which could then appear on Snap's public surfaces. Snap removes illegal content and activity as we become aware of it, cooperates with law enforcement, and disables the accounts of egregious or repeat violators.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **[link in the left hand column to a full summary](#)** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.



DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat's in-scope services have been adapted to include proactive moderation for other illegal activities.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, terms prohibit other illegal activities and they are strictly enforced. Our legal team, supported by external counsel as needed, reviews reports of new issues to confirm illegality and appropriate enforcement action.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, general proactive and reactive moderation procedures to prevent and remove other illegal activities.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend information relating to illegal activity i.e. there is no 'illegal activity' interest category.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>We rely on our Trusted Flaggers, industry groups and our safety advisory board and internal cross functional working groups to ensure we are prioritizing the right issues. With the introduction of the Digital Services Act, we have introduced a new reporting option to report 'illegal content' in general, and we expect to use data gathered from this option to provide us with greater visibility on the prevalence of information relating to other illegal activity on Snapchat.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the</p>	<p>Yes, we cooperate with trusted flaggers who are able to flag other illegal activities.</p>



implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.	
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	Yes, we cooperate with other providers through various industry groups on prominent issues facing online platforms.
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	General content authenticity measures.

Conclusion

We prohibit the dissemination of information relating to illegal activities and criminal activity in our [Terms](#). We specifically track the issues relating to the dissemination of information which we consider to have the most relevance to online platforms, such as Snapchat. We treat other dissemination issues as a Level 3 overall potential risk prioritization and have taken steps to mitigate these risks. We regularly review our risk categories using our Risk Detection and Management processes and we expect the new option to report ‘other illegal activity’ to provide us with further insight.

As a result, we have concluded that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures for information relating to other illegal



activities. We will continue to review online trends and establish new specific categories as needed.

4.2 Category 2: Negative Effects on Fundamental EU Rights

(Article 34.1.b / DSA Recital 81)

In this part of the Report, we explain the results of our assessment on actual or foreseeable negative effects of Snapchat’s in-scope services on our Fundamental EU Rights as required by Article 34.1.b and Recital 81 of the Digital Services Act. Those Fundamental EU Rights are set out in the Charter of Fundamental Rights of the European Union (the “Charter”)¹⁴. We have assessed in particular the rights to human dignity, freedom of expression and of information, including media freedom and pluralism, private life, data protection, non-discrimination and consumer protection. We also consider the rights of the child, including how easy it is for Teens to understand the design and functioning of the service, as well as how Teens can be exposed through their service to content that may impair Teens’ health, physical, mental and moral development. Such risks may arise, for example, in relation to the design of online interfaces which intentionally or unintentionally exploit the weaknesses and inexperience of Teens or which may cause addictive behavior.

Category 2 - Negative effects on Fundamental EU Rights					
Category	Relative likelihood of risk occurring on Snapchat	Harm classification industry wide	Risk Prioritization	Approach	Conclusion
4.2.1 Right to human dignity	Extremely Low Likelihood	Severe harm industry wide	Level 1	Mitigated	Low Risk / Reasonable, proportionate and effective
4.2.2 Right to freedom of expression	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective
4.2.3 Right to private life	Extremely Low Likelihood	Serious harm industry wide	Level 2	Mitigated	Low Risk / Reasonable, proportionate and effective
4.2.4 Right to	Low Likelihood	Severe harm	Level 1	Mitigated	Low Risk /

¹⁴ Charter of Fundamental Rights of the European Union ([url](#)).



data protection		industry wide			Reasonable, proportionate and effective
4.2.5 Right to non-discrimination and freedom of religion	Extremely Low Likelihood	Serious harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective
4.2.6 Children's rights	Extremely Low Likelihood	Severe harm industry wide	Level 1	Mitigated	Low Risk / Reasonable, proportionate and effective and we are actively participating in efforts to develop an EU wide AADC to assess if further industry measures are needed.
4.2.7 Right to consumer protection	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective
4.2.8 Right to property	N/A. Already covered under Category 1: Dissemination of content that infringes on intellectual property rights				

4.2.1 Right to human dignity

All public spaces displaying user generated content have the potential for the dissemination of content that may undermine human dignity. We recognise that without mitigations such content could conceivably appear in any of Snapchat's in-scope services displaying user generated content, from videos featured on Spotlight / For You, to Place Stories on Snap Map. Advertising could, for example, include hate speech or discriminatory elements.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Snapchat, as with other platforms that host user generated content, may be used to spread content that undermines respect for human dignity. Without mitigations, this could include content that promotes:

- Human trafficking and/or the sale of coerced sex;



- Child sexual abuse material;
- Terrorism;
- Self-harm, including the promotion of self-injury, suicide or eating disorders;
- Incitement to violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the Charter.

The likelihood of all of these risk categories fall within the **extremely low likelihood category**. See section 4.1 above.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that may undermine human dignity were to materialise on an online platform, it would fall within our ‘severe harm’ category.

Overall potential risk prioritization

Although the prevalence of content that negatively impacts users’ rights to human dignity is negligible due to robust mitigations, the severity of the potential impacts is such that the risks remain in the Level 1 overall potential risk prioritization category. As described in our risk methodology in Section 1, we assess overall potential risk on a case by case basis and Snap reserves the option to deviate from the overall potential risk matrix we use as a guide. This is one of the cases where we have chosen to deviate.

Snap's Mitigations

Highlights

Snap’s approach to protecting users’ rights to human dignity and mitigating the related risks is implemented through a robust framework of content moderation as described in the [Moderation](#) section.

Activity that may undermine human dignity is not permitted on our platform under our [Terms](#) and [Community Guidelines](#). We have tools within the app where individuals may report this type of activity to our Trust & Safety team. They will then investigate the report and take action. In the event the report concerns any imminent threat to life, Snap will alert the appropriate authorities.. We also maintain relationships with several entities on a global basis through our Trusted Flagger program and they may also report activity to our Trust & Safety team. These Trusted Flaggers are vetted and they possess an expedited means for contacting our teams.

In addition to effective [content moderation](#), Snap has additional mechanisms in place to enhance the right to human dignity for users. For example, because Snapchat is a platform designed for communications between real friends, it can play a unique role in empowering friends to help each other through difficult times. When our Trust & Safety team recognizes a Snapchatter in



distress, they can forward self-harm prevention and support resources, and notify emergency response personnel when appropriate. The resources that we share are available on our global list of safety resources, and these are publicly available to all Snapchatters.

Specific Mitigations

This table lists a number of specific mitigations to address risks to human dignity on Snapchat's in-scope services. To avoid duplication, this table includes cross-references to other sections of this Report.

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	<p>Yes, Snapchat's in-scope services have been adapted to include proactive moderation for CSEAI and other illegal content that undermines human dignity.</p> <p>We also have tools within the app where individuals can report this type of activity to our Trust and Safety team.</p> <p>When our Trust and Safety team recognizes a Snapchatter in distress, they can forward self-harm prevention and support resources, and notify emergency personnel when appropriate. For example, if a user searches for suicide related terms we will surface our Here For You tool.</p>
Terms and Enforcement Adapting their terms and conditions and their enforcement.	<p>Yes, our Terms prohibit CSEAI and other illegal content that undermines human dignity and they are strictly enforced.</p>
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to	<p>Yes, specific proactive and reactive moderation procedures to prevent CSEAI and other illegal content that undermines human dignity.</p>



<p>specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>We have terms in place to prevent Media Partners from publishing illegal or harmful content on For You. All Media Partners are vetted prior to being permitted to distribute their content broadly on Snapchat by a team of editors. Media Partners go through an editorial review of their content, a reputational search (to evaluate if a Media Partner has a history damaging press, legal actions, etc.), and compliance review before they're able to distribute content. Media partners are proactively moderated, and the content of their shows/editions are reactively moderated. Senior partner managers will relay feedback to Media Partners to remove or change content. If a partner refuses, we could just remove it ourselves, but partners typically comply.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not knowingly recommend content that would negatively affect the right to human dignity i.e. there are no interest categories that we consider to negatively affect human dignity.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency report for CSEAI, terrorist content, and other illegal content that undermines human dignity.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to CSEAI and other illegal content that undermines human dignity.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. EUIF.</p>
<p>Transparency</p>	<p>Yes, we provide guidance on our terms, harms,</p>



<p>Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center. For example, if a user searches for suicide related terms we will surface our Here For You tool.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

Snap considers risks to human dignity to have a Level 1 overall potential risk. In response it has put in place a range of mitigation measures. This includes in particular our proactive content [moderation](#) which is designed to detect and prevent CSEAI and other content that may undermine human dignity from appearing on each of Snapchat's in-scope services. For example, our automated and human review on Spotlight. Our prevalence testing has allowed us to improve this proactive content moderation. As a result, we've reduced the prevalence of CSEAI and other content that may undermine human dignity on Snapchat's in-scope services to the lowest likelihood level. See Section 4.1

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for risks to human dignity.

4.2.2 Right to freedom of expression and assembly

Snapchat is an app whose mission is to empower people to express themselves, live in the moment, learn about the world, and have fun together. By design, the app itself presents an opportunity to enhance the freedom of expression and assembly of Snapchatters. However,



without mitigations, Snap, alongside other digital platforms hosting user-generated content, presents some risk to these rights and freedoms. These risks could include: Algorithmic biases, content moderation bias, submission of abusive notices and self-censorship.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

It is difficult to quantify the likely risk of negative impact on freedom of expression and assembly. Algorithmic biases and self-censorship are difficult to detect. We rely on user feedback and testing to flag significant incidents. At present, we are not aware of any significant bias of self-censorship issues in the algorithms used by Snapchat's in-scope services. We monitor the number and nature of the general community support requests we receive and this data does not identify any trend that suggests Snapchat may be negatively impacting freedom of expression or assembly. Our [transparency reports](#) show that in general we receive low incidents of illegal content reports from recipients of Snapchat or authorities where we choose not to take enforcement action. . Based on the lack of reporting Snap has received and the [overall design of Snapchat](#) (which does not generally provide a platform for political content in general), we deem this to fall within the extremely low likelihood category.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that may undermine human dignity were to materialise on an online platform it would fall within our **significant harm category**. However, Snapchat generally is not a platform for political or activist content and so the impact on freedom of expression and assembly is unlikely to be severe on Snapchat compared with other spaces on the internet dedicated to such content.

Overall potential risk prioritization

Although it can be difficult to determine, the lack of reporting and Snap's overall design, indicates that the prevalence of issues relating to freedom of expression and assembly are low. As Snapchat's in-scope services do not generally amplify political or activist content, the severity of any freedom of expression risk is significant but not serious or severe. We consider that freedom of expression risks fall within the **Level 3** category overall.

Snap's Mitigations

Highlights

Our [Terms](#) clearly define certain topics which we prohibit, including false information that threatens public health (e.g. COVID-19 vaccinations), civic processes, or that denies tragic events



(like the Holocaust). We also have an [explainer](#) to help our community understand how we handle harmful false or deceptive information. This provides clarity on the limits we have when it comes to freedom of expression and assembly.

Our platform is generally not a place for political or activist content. Such content is not eligible for promotion on Spotlight and user content on For You is only from a small number of popular, entertaining community creators and their content is moderated by humans against our Content Guidelines. All Media Partners are vetted prior to being permitted to distribute their content broadly on Snapchat by a team of editors. These Partners include news organizations, which are subject to their own professional rules. Media Partners go through an editorial review of their content, a reputational search (to evaluate if a publisher has a history damaging press, legal actions, etc.), and compliance review before they're able to distribute content. As a result, we provide a balanced approach to political and activist content on Snapchat that is designed to limit the sources of such information to professional media partners.

Snapchat utilizes content [moderation](#) policies and systems to protect users' rights to freedom of expression and access to accurate information. As all of our user generated content is moderated by a mix of automation and human moderation, we proactively remove content that does not meet our policies before being broadly distributed. In some cases, content against our policies may make it past moderation by mistake. In those cases, we rely on Snapchatters to report the content for re-moderation. As explained when discussing the [dissemination of content that infringes on intellectual property rights](#), Snap respects the doctrine of "fair use," i.e., that there are certain circumstances (such as news reporting, social commentary on issues of public interest, criticism, parody, or education) where excerpts of copyrighted material could be distributed without permission from or payment to the copyright holder. This helps reinforce the rights of freedom of expression and the freedom of assembly.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function	Yes, Snapchat's in-scope services have been adapted to include proactive moderation to



<p>Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>protect users' rights to freedom of expression and access to accurate information.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, terms prohibit harmful false or deceptive information and they are strictly enforced.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to protect users' rights to freedom of expression and access to accurate information.</p> <p>On Snap Map, our editorial oversight protection for content showing up on Snap Map tries to strike the right balance between the need to preserve the public safety versus the free flow of information and expression. Examples of this include in February of last year when Russia moved into Ukraine, Snap Map developed tooling that allowed us to block all of Ukraine from creating content. This was in response to concerns that Russia was using it for their own strategic purposes (propaganda and tracking the movement of Ukrainians).</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, the pool of content recommended by our algorithmic systems does not generally include political or other important societal matters.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we consult with our safety advisory board to ensure Snapchat is set up appropriately and monitor community reports for issues relating to freedom of expression.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>No, we do not work with trusted flaggers for users' rights to freedom of expression and access to accurate information.</p>
<p>Codes and Crisis Protocols</p>	<p>Yes, our Crisis Protocols handle issues related to users' rights to freedom of expression and access to accurate information. We have</p>



<p>Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>recently exercised these protocols successfully during the French riots in June 2023.</p> <p>Note, we will continue to reassess and explore the opportunity to join the EU disinformation code.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms (including harmful false or deceptive information), moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to allow Teens to express themselves without the pressures of friends lists, comments and likes. We have community, ad and content guidelines that are specific to teens. We also offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authentication measures.</p>

Conclusion

Snap considers the overall risk to be within the severe category given the stakes and the severity of threats to freedom of expression, despite low prevalence and robust protections in place. Snap's mission is to be an expressive platform where users can be their authentic self, and we view our obligation to facilitate freedom of expression as foundational. While harms to freedom of expression are hard to detect, and we are not aware of any significant bias of self-censorship issues in the algorithms used by Snapchat's in-scope services, we provide avenues for our users to report these issues to us, and we value and respect user feedback. We continually evaluate and evolve our algorithms, including to reduce perceived biases, and monitor for and respond to events that could impact freedom of expression. We couple this with enforcement of our [Terms](#)



and our robust [moderation](#) practices to provide a platform where users feel free to express themselves in the world.

We have concluded therefore that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures to address risks relating to freedom of expression. Snap monitors its impact on this fundamental right category to confirm prevalence continues to decline, or whether further mitigating measures might be required.

4.2.3 Right to private life

We understand well that online platforms can be used to spread content that undermines respect for private and family life, and that such content can have traumatic consequences if not properly mitigated. On Snapchat, without mitigations, content that undermines private and family life and personal data privacy could conceivably appear in any of Snapchat’s in-scope services displaying user generated content, including information in videos featured on Spotlight / For You and Snap Map.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Snapchat’s platform architecture, combined with its commitment to responsible policy enforcement across our content surfaces, establishes safeguards against negative impacts to the private life of users. Our prevalence testing shows that “invasion of privacy” made up an extremely low percentage of Policy Violating Prevalence on Snap in August 2023 (see our [Prevalence Testing](#) chapter). Snap also receives low numbers of privacy-related queries from recipients. As a result, we deem this risk to fall within our **Extremely Low Likelihood category**.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that may undermine private and family life and personal data privacy were to materialise on an online platform it could fall within our ‘**serious harm**’ category.

Overall potential risk prioritization

Given the stakes and the severity of threats to private life, Snapchat assesses the overall risk to be within the **Level 2** category, despite low prevalence and robust protections in place. As described in our [risk methodology](#) in Section 1, we assess overall potential risk on a case by case



basis and Snap reserves the option to deviate from the overall potential risk prioritization matrix we use as a guide. This is one of the cases where we have chosen to deviate.

Snap's Mitigations

Highlights

Snap takes a multifaceted approach to mitigating negative impacts to private life and personal data protection, starting from the way Snap develops its own enforcement mechanisms. Privacy is the first of Snap's four core platform governance values, which remains paramount as Snap contemplates the development of supplementary enforcement mechanisms that could potentially impact users' personal data. Through Snap's Platform Governance Framework, efforts to mitigate or understand harm must advance one or more of the platform governance values, and be consistent with the balancing principles of necessary, proportional, and legitimate. The principles of necessity, proportionality, and legitimacy derive directly from established human rights law and jurisprudence, and have been adapted for application in many different contexts, including as guiding principles for safeguarding against digital surveillance. Incorporating these principles into Snap's framework anchors our approach in an internationally validated, rights-respecting methodology—one that is familiar to, and utilized by, policymakers and advocates in every region of the world. In developing this framework, we've drawn on a large body of principles and expertise from across the digital policy, human rights, and online privacy communities.

We also mitigate these risks through intentional product design choices. Privacy by Design is Snap's approach to building products that consider user privacy from inception. Each product is subject to a PASS Review (Privacy Assessment System) to ensure that our products do not misuse user-data. We also engage with experts in the privacy and human rights community to inform our approach. This includes collaborations and engagement with individual experts (such as expert on human rights, privacy, and online safety Brittan Heller and former ICO Commissioner Elizabeth Denham, and several others), as well as think tanks and research collaborations.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **[link in the left hand column to a full summary](#)** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.



DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Privacy by Design is Snap’s approach to building products that consider user privacy from inception. Each product is subject to a PASS Review (Privacy Assessment System) to ensure that our products do not misuse user-data.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, for example, our Community Guidelines prohibit impersonation, our Commercial Content Policy prohibits non-consensual sexual material and our Spotlight Terms requires “<i>you must have any necessary third-party rights including, without limitation, music copyrights and rights of publicity, for all content in your Snaps</i>”.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to protect the privacy interests of our community.</p> <p>Users have the ability to report Snaps and the reporting menu includes options “They leaked / are threatening to leak my nudes”, “It’s an inappropriate Snap of me”, “It involves a child”, and “They are pretending to be me”.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend content that violates users’ rights to private life.</p> <p>For example, we have terms, moderation and enforcement to prevent distribution of illegal / violating content. We also do not process sensitive category information.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems. For example, Snap ensures that ads shown are in line with its Snap Advertising Policies which states that advertisements do not collect sensitive information or special category of data. We also ensure advertisers are not targeting specific individuals on our platform and that users do not feel like their privacy is being compromised by our advertising.</p>
<p>Risk Detection and Management</p>	<p>Yes, we have specific prevalence testing and transparency reports for sexual content and intrusion of privacy.</p>



<p>Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>We also monitor privacy-related inquiries as detailed above.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to sexual content and Teen safety which may impact users' right to private life.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g CIPL, FPF.</p> <p>Our content moderation policies provide de facto content moderation crisis protocol.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p> <p>Our Privacy Center offers a suite of information on our products, users' choices to safeguard their privacy and how to contact us.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures in place for Teens. For example, Teens cannot create public profiles and if they post to Spotlight or Snap Maps their profile details are anonymized as an extra precaution. Our reporting menu also include the option to report "It involves a child".</p> <p>Our Family Center includes resources and guidance for Teens and their parents or trusted adults.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition,</p>	<p>General content authenticity measures.</p>



providing an easy to use functionality which enables recipients of the service to indicate such information.

Conclusion

Snap considers the overall risk to be within the Level 2 category given the stakes and the severity of threats to privacy life, despite low prevalence. However, Privacy is the first of Snap's four core platform governance values. We have robust protections in place, including clear terms and moderation. Snap enforces against these content violations robustly. We also mitigate risks through intentional product design choices and collaborate with experts, think tanks and researchers on human rights, privacy and online safety to inform our approach.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures to address risks relating to the right to private life.

4.2.4 Right to data protection

We understand well the importance of ensuring that personal data is collected, processed or secured appropriately. Depending on how and the extent to which Snapchatters use our platforms, significant volumes of the content published on Snapchat's in-scope services, including on Spotlight / For You and Snap Map, is user generated images and videos.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Snapchat handles a significant volume of personal data relating to individuals in the European Union. Depending on how and the extent to which Snapchatters use Snapchat, this could be limited to basic account information or it could extend, for example, to published images and videos and metadata about the Snapchatter's interaction with such content. Significant volumes of the content published on Snapchat's in-scope services is user generated images and videos which might be related to individual Snapchatter creators and/or others. It is therefore more likely than not that Snapchat's in-scope services could cause an impact on an individual's data protection rights if such personal data is not collected, processed or secured appropriately.

We monitor the number and nature of Privacy and Data Protection requests we receive and Snap receives relatively low numbers of privacy-related queries relating to the European Union. Despite low numbers, given the significant volume of personal data being processed by Snap in relation to Snapchat, we have chosen to include this risk within our **Low Likelihood** category.



Severity

Snap has assessed information published by governments and other third party sources and considers that if an online platform were to undermine the right to data protection, this could fall within our **'severe harm' category**.

Overall potential risk prioritization

Considering the extent of the personal data being processed by Snapchat and the serious nature of rights of data protection within the European Union to be a Level 1 overall potential risk prioritization, notwithstanding the low incident of [privacy related queries](#) from recipients of Snapchat's in-scope services. As described in our [risk methodology](#) in Section 1, we assess overall potential risk on a case by case basis and Snap reserves the option to deviate from the overall potential risk prioritization matrix we use as a guide. This is one of the cases where we have chosen to deviate.

Snap's Mitigations

Highlights

Privacy is central to Snapchat's values. When we first created Snapchat, we decided to build a platform with strong [privacy principles](#), pioneering data minimization and messages that delete by default. We believe that visual communication and messages that delete by default give young people the opportunity to express themselves without the pressures of public metrics and permanence. Online platforms may have normalized having a permanent record of conversations online, but in real life, friends don't break out their tape recorder to document every single conversation for public consumption or permanent retention. This makes Snapchat feel less like a permanent record and more like a conversation with friends—allowing people to express themselves in the same way they would if they were just hanging out at a park with their friends.

We put significant thought and consideration to ensure those principles are reflected into the architecture of our platform, and into the design and implementation of our products, policies, and enforcement actions. Since Snapchat's inception, we have embraced a [privacy and safety by-design](#) approach and decided that our platform architecture and product choices should play a major role in risk-mitigation. They can be found in our privacy and safety by design principles. We have a dedicated cross-functional group responsible for compliance with these principles. This group brings together Legal, Policy, Engineering and Product. Material product changes relating to Snapchat are reviewed by Legal and specialist engineering teams, as well as relevant members of the cross-functional group. We use our Safety and Privacy by Design principles to help mitigate risks to Teens. We maintain Data Protection Impact Assessments (DPIAs) of our processing of personal data to ensure we are confident this will not result in a high risk to the rights and freedoms of individuals.



As a result of the measures that Snap takes to protect personal data and provide users with actionable tools and transparent information, Snap receives low numbers of privacy-related queries.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **[link in the left hand column to a full summary](#)** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat is a platform with strong privacy principles. These principles are reflected into the architecture of our platform.</p> <p>Product changes are subject to privacy by design reviews and we maintain data protection impact assessments.</p> <p>For example, our Lenses only require object detection rather than facial identification. Lenses can tell what is or isn't a face, they do not identify specific faces, limiting data processing for the use of Lenses. Snap does not use any data collected by Lenses to customize the content that the user sees in Spotlight or For You, nor is any data collected for advertising purposes. Besides, voice data collection of Snapchatters in the EU is off by default; it is only used to provide the service.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, our Privacy Center provides a suite of policies, including our Privacy Policy and they are enforced.</p> <p>In our Content Guidelines for Recommendation Eligibility we inform creators “We inform these standards with proactive moderation using technology and human review” and “you must</p>



	<p><i>have any necessary third-party rights including, without limitation, music copyrights and rights of publicity, for all content in your Snaps”</i> This prevents any risk that users may not be aware that their content submitted to Spotlight is subject to automated and human review, and prohibits creators from depicting individuals in content without necessary rights.</p> <p>In our Snap Spotlight Submission and Revenue Terms we state “<i>You understand that Snaps you submit to Spotlight are Public Content and may be visible to all Snapchat users, as well as non-Snapchat users on other services and websites</i>” This prevents the risk of creators being unaware that their Stories submitted to Spotlight become public and informs users that their content may be saved off Snapchat.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to prevent and remove content that violates users’ right to data protection.</p> <p>For example, on For You Media Partners are proactively moderated and only a small pool of Snapchatters are shown in For You (“Snap Stars” or “Popular Users”).</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend content that violates users’ right to data protection.</p> <p>For example, users can opt out from personalized recommendations based on inferred interest and we do not process sensitive category information.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, our Advertising Systems has a suite of protections including:</p> <ul style="list-style-type: none"> ● No microtargeting ● We offer controls to turn off most personalized ads. Users can learn more about their choices here How to Adjust



	<p>My Advertising & Interest Preferences on Snapchat.</p> <ul style="list-style-type: none"> • We ensure that sensitive data is not being used for ad targeting • We continue to trial evolving privacy enhancing technologies, such as third party data clean rooms, to provide advertisers with options to further minimize the privacy impact of Snap ad services.
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we consult with experts and our community, and we also monitor and respond to privacy-related inquiries.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>No, we don't cooperate with trusted flaggers in relation to data protection violations.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. CIPL, FPF.</p> <p>We also have a well-established protocol to deal with privacy incidents, as well as a Security Incident Response Policy.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on privacy protection in our Privacy Center. For example, we explain to users How We Rank Content in Discover¹⁵, How We Rank Content on Spotlight – Snapchat Support and Snapchat Ads Privacy & Transparency.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit disclosure of Teens' data and we avoid nudge techniques to encourage Teens to change their privacy settings.</p> <p>We offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p>

¹⁵ These - and other - support pages will be updated to change the recent name change from Discover to For You.



	<p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

Snap considers the likelihood and the serious nature of the impacts on the right of data protection within the European Union to fall within our Level 3 overall potential risk of Snapchat's in-scope services. Depending on how and the extent to which Snapchatters use these platforms, significant volumes of the content published on Snapchat's in-scope services is user generated images and videos. It is therefore more likely than not that Snapchat's in-scope services could negatively affect an individual's data protection rights if such personal data is not collected, processed or secured appropriately, which is why Snap enforces its [privacy principles](#) robustly. Privacy is central to Snapchat's values. We put significant thought and consideration into our [privacy principles](#) and those principles are reflected into the architecture of our platform. We have a cross-functional group responsible for compliance with our [privacy and safety by design principles](#), we review product changes for impact to data protection rights and we maintain Data Protection Impact Assessments of our processing of personal data where appropriate to ensure we are confident this will not result in a high risk to the rights and freedoms of individuals. We receive a low level of data protection queries as a result of the robust protections in place.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures against the risk of negative effects on data protection rights. Snap monitors its impact on this fundamental right category to confirm prevalence continues to decline, or whether further mitigating measures might be required.

4.2.5 Right to non-discrimination and freedom of religion

We understand well that online platforms can be used to spread content that contains or promotes discrimination for example by using discriminatory characteristics for targeting ads, biased algorithms used for recommender systems and content moderation, the spread of discriminatory content, facilitating online harassment, disproportionately reporting accounts of individuals from marginalized (religious) communities based on user reports, etc. This risk poses a



serious threat to the rights of EU citizens who are already vulnerable to abuse and have encountered discrimination and marginalization historically. Without mitigations, content that undermines the right to non-discrimination and freedom of religion could conceivably appear in any of Snapchat's in-scope services displaying user generated content, including information in videos featured on Spotlight / For You and Snap Map.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms). Illegal hate speech accounted for a small fraction of violative content on Snapchat - see section 4.1.2 above. Given the relatively low prevalence, this risk falls within our **Extremely Low Likelihood category**.

Severity

Snap has assessed information published by governments and other third party sources and considers that if an online platform were to undermine the right to non-discrimination and freedom of religion, this could fall within our '**serious harm**' category.

Overall potential risk prioritization

Snap would consider the risk to the right to non-discrimination and freedom of religion to fall within the **Level 3** category overall. Although we consider this risk to fall within our serious harm category, there is a relatively low prevalence of hate speech on our online platform.

Snap's Mitigations

Highlights

To help ensure our policies against hate speech are enforced responsibly, our teams consult the expertise and work of civil society organizations, like Access Now, human rights experts, law enforcement agencies, NGOs, and safety advocates. We are constantly learning, and will calibrate wherever necessary to ensure that our products and policies function to keep Snapchatters safe.

Practically, our in-app reporting tool allows users to directly report hateful content or activities that support terrorism or violent extremism. On our high-reach surfaces, like Spotlight and For You, we take a proactive approach to moderating any content that may violate these rules. When hateful content is reported, our teams will remove any violating content and users who engage in repeated or egregious violations will have their account access locked. As an additional measure, we encourage Snapchatters to block any users who make them feel unsafe or uncomfortable.



As reported in our [Diversity Annual Report](#), our North Star at Snap is building equity of experience for all current and future Snapchatters. No matter where the user lives, what their background is, how they look, how they communicate, their socioeconomic status, they should feel Snap products are made for them. We know inclusive design is the right path forward - both to grow our business and uphold our values. Product Inclusion encourages us to look at how Snapchatters' needs are influenced by aspects of their identity, environment, abilities and life experiences. Product Inclusion helps us create equitable experiences by intentionally involving and considering marginalized groups at critical moments throughout the product development process. Below we explain our diversity and inclusion efforts for Spotlight, For You, Snap Maps and Lenses, to combat discrimination on our platforms.

For publicly available content on Spotlight, For You and Snap Maps:

- We survey a subset of our users quarterly to understand whether they find their time spent on our experience entertainment and satisfactory. We use this to track whether our product changes are improving viewers' overall perception of the app.
- We provide a diversity of perspectives. We have multiple programs to foster a more diverse content community and surface different perspectives (e.g. [Black accelerator program](#).)
- We ensure there is always a large mix of content from creators from viewers' home country and content in the language in which they have set their device.
- We add diversity to every viewer's feed in terms of the account they see, and the categories of content we surface to them. This prevents users from entering an echo chamber or filter bubble of seeing the same content repeatedly. We use machine learning to understand content categories and diversify it.

Modifying facial features or overlaying cultural elements in Snapchat's Lenses may reinforce discriminatory ideas based on appearance or ethnicity and promote harmful imagery. Also, Lenses incorporating cultural symbols or references might lack proper context and sensitivity. The [Lens Studio Submission Guidelines](#), reiterated our Community Guidelines and spell out that the following categories of Lenses are prohibited:

- Content that demeans, defames, or promotes discrimination or violence on the basis of any of the identities listed in our Community Guidelines
- Examples: slurs, stereotypes, hate symbols, the promotion of hateful conspiracy theories, the glorification of atrocities or historical hatemongers

Snap designs every Lens with race, gender, ethnicity and cultural norms in mind. Snap leverages its ever-growing diverse training datasets, as well as feedback from community members. If a Lens does not resonate with our community, as expressed through a high ratio of user reports, we take that feedback into consideration and will re-review the Lens with a goal to leave as-is, modify, or remove.



If a Lens is appropriate, but could theoretically be misused by someone, that alone is not sufficient to reject a Lens. Snap considers current and historical global events when releasing a Lens, and delays or denies amplification to Lenses that may be deemed insensitive due to broader social occurrences throughout the world. Lenses should not change a user's skin tone to mimic a different ethnicity or race. Snap does not modify facial or other features in a way that evoke racial, ethnic, cultural or religious stereotypes or stigmatized disabilities. Snap presents religious and cultural iconography in a respectful manner, with feedback solicited from internal and external subject matter experts. This means Snap is especially thoughtful around holiday or event-based content, including the geography in which a Lens will launch. Also, Snap ensures that a Lens is not deceptive. Snap uses signifiers and watermarks where there may be questions of creative authenticity. Snap tests Lenses on photos/videos of and in real life settings with diverse groups of people to accurately enforce our policies.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **[link in the left hand column to a full summary](#)** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat's in-scope services have been adapted to include proactive moderation for hateful content or activities supporting terrorism or violent extremism.</p> <p>We also work with civil society organisation to ensure our policies are enforced responsibly.</p> <p>Product Inclusion helps us create equitable experiences by intentionally involving and considering marginalized groups at critical moments throughout the product development process.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, Snap's Terms and Community Guidelines prohibit Hate speech or content that demeans, defames, or promotes discrimination or violence on the basis of race, color, caste, ethnicity, national origin, religion, sexual orientation, gender identity, disability, or</p>



	<p>veteran status, immigration status, socio-economic status, age, weight, or pregnancy status. We strictly enforce these rules.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to prevent and remove hateful content or activities supporting terrorism or violent extremism.</p> <p>We provide in-app reporting for hateful content or activities supporting terrorism or violent extremism.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend hateful content or activities supporting terrorism or violent extremism.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p> <p>In order to ensure we are not using discriminatory targeting models particularly when there is significant legal impact to the consumers, we offer special targeting models that do not include gender or age, which we require for advertisers who are advertising in the housing, credit or employment (HCE) spaces, so that discriminatory factors will not go into who sees these ads. We do not allow advertisers to build audiences for their ads based on their own data about our teenage users regardless of those user's own ad settings (i.e. activity data from the advertisers own online properties and the advertiser's own customer lists).</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for hate speech, terrorist and violent extremist content.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the</p>	<p>Yes, we cooperate with trusted flaggers in relation to illegal hate speech, terrorist and violent extremist content.</p>



<p>implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. EU Internet Forum.</p> <p>We have signed onto the EU hate speech Code.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p> <p>We provide in-app reporting for hateful content or activities supporting terrorism or violent extremism.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures in place for Teens. For example, Teens cannot create public profiles and if they post to Spotlight or Snap Maps their profile details are anonymized as an extra precaution. Our reporting menu also includes the option to report “ It involves a child”. We hope protections like these help protect Teens from hateful content.</p> <p>Our Family Center includes resources and guidance for Teens and their parents or trusted adults.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>



Conclusion

Snap considers the overall risk to be within the Level 3 category taking account of the harm that risks to the right to non-discrimination and freedom of religion may cause and the low prevalence for hate speech on the platform. In practice Snap has substantial protective measures in place. Snap works with civil society organizations, like Access Now, human rights experts, law enforcement agencies, NGOs, and safety advocates to make sure we are calibrating wherever necessary to ensure that our products and policies function to keep Snapchatters safe. Our in-app reporting tool allows users to directly report hateful content or activities that support terrorism or violent extremism. On our high-reach surfaces, like Spotlight and For You, we take a proactive approach to moderating any content that may violate these rules. Further, our diversity and inclusion efforts help us create equitable experiences by intentionally involving and considering marginalized groups at critical moments throughout the product development process.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures to protect users' right to non-discrimination and freedom of religion.

4.2.6 Children's Rights

We understand that online platforms can impact children's and Teen's rights. This is a risk we take seriously as Snap's priority is protecting the safety and wellbeing of our users whilst ensuring they have a positive experience online. Privacy, safety and security are key values of the company and at the core of our value proposition to our users.

The 'rights of the child' under the Charter¹⁶ comprises two elements that are relevant to Snapchat's in-scope services:

1. Children have the right to such protection and care as is necessary for their well-being; and
2. Children have the right to express their views freely and have those views taken into consideration on matters which concern them in accordance with their age and maturity.

In respect of element 1, we address the well-being of children when considering Category 4 of the DSA risks in particular parts of the [negative effects on children](#) and [physical and mental wellbeing](#) elsewhere in this Section 4. **This section therefore focuses on element 2 i.e. risks to children's rights of expression.**

Likelihood

As explained in [Snapchat Community](#) as part of our Introduction to this Report, Snapchat is used by a wide demographic, with 18-24 years making up the highest percentage of users of Snapchat.

¹⁶ Art 24, Charter of Fundamental Rights of the European Union (CFREU) ([url](#)).



Nevertheless, there is still a percentage of our users who are Teens (13-17). Therefore we consider that children using Snapchat are just as likely to be exposed to freedom of expression issues identified in this Report as other members of the Snapchat Community as follows:

Risk Category	Relative likelihood of risk occurring on Snapchat	Relative likelihood of negative effect on children
Right to freedom of expression	Extremely Low Likelihood	Extremely Low Likelihood

As a result, we conclude that the relative likelihood of a risk of negative effects on children and Teens for Snapchat's in-scope falls within the **Extremely Low Likelihood** category.

Severity

We assessed the risk of harm from the right to freedom of expression to fall within our significant harm classification. However, we take the safety and wellbeing of the youngest members of our community very seriously and recognise that this group is particularly vulnerable and if a particular risk materializes, there is an increased risk that the severity of the harm they suffer is higher. For freedom of expression, we consider this as follows:

Risk Category	Harm classification industry wide	Is the industry wide severity risk higher for children and Teens?
Right to freedom of expression	Significant harm industry wide	Yes, Snap considers that it is vital that children and Teens are able to access online platforms and participate in lawful online debate and dialogue to learn, have their views heard and develop their own values and identities, regardless of their ability to pay.

As a result, we have chosen to place the severity of harm arising from an issue that negatively affects children's rights in our **'severe'** category.

Overall potential risk

Although the relative likelihood for the negative effects on children's rights falls without our Extremely Low Likelihood category, Snap considers the risk of harm to fall within the severest category. Consequently, Snap considers this to be a **Level 1** overall potential risk for Snapchat's in-scope services.



Snap's Mitigations

Highlights

As explained in the [freedom of expression and assembly](#) part of this Section 4, we have put in place a number of mitigations to ensure that all users, including Teens, have the right to express views freely, where appropriate:

- Our [Terms](#) clearly define certain topics which we prohibit, including false information that threatens public health (e.g. COVID-19 vaccinations), civic processes, or that denies tragic events (like the Holocaust). We also have an [explainer](#) to help our community understand how we handle harmful false or deceptive information. This provides clarity on the limits we have when it comes to freedom of expression and assembly.
- Our platform is generally not a place for political or activist content. Such content is not eligible for promotion on Spotlight and user content on For You is only from a small number of popular, entertaining community creators and their content is moderated by humans against our Content Guidelines. All Media Partners are vetted prior to being permitted to distribute their content broadly on Snapchat by a team of editors. These Partners include news organizations, which are subject to their own professional rules. Media Partners go through an editorial review of their content, a reputational search (to evaluate if a publisher has a history damaging press, legal actions, etc.), and compliance review before they're able to distribute content. As a result, we provide a balanced approach to political and activist content on Snapchat that is designed to limit the sources of such information to professional media partners.
- Snapchat utilizes content [moderation](#) policies and systems to protect users' rights to freedom of expression and access to accurate information. As all of our user generated content is moderated by a mix of automation and human moderation, we proactively remove content that does not meet our policies before being broadly distributed. In some cases, content against our policies may make it past moderation by mistake. In those cases, we rely on Snapchatters to report the content for re-moderation.
- As explained when discussing the [dissemination of content that infringes on intellectual property rights](#), Snap respects the doctrine of "fair use," i.e., that there are certain circumstances (such as news reporting, social commentary on issues of public interest, criticism, parody, or education) where excerpts of copyrighted material could be distributed without permission from or payment to the copyright holder. This helps reinforce the rights of freedom of expression and the freedom of assembly.

We believe that it is also important that our business model supports the right for all users to use Snapchat, regardless of ability to pay, by paying for the cost of the service through balanced, and proportionate targeted advertising (as explained further in the [data protection rights](#) and [advertising systems](#) section of this Report). This has been made more challenging by the obligation in the DSA to prohibit all forms of targeted advertising to Teens, even if balanced with



reasonable, proportionate and effective mitigation measures in place. However, we continue to offer Snapchat's in-scope services to all, without charge, including Teens.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, Snapchat's in-scope services have been adapted to include proactive moderation to protect Teens' access to accurate information and provide an appropriate environment to meet, see new experiences and express themselves.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, terms provide clear rules to Teens on the boundaries of appropriate expression and prohibit harmful false or deceptive information.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, specific proactive and reactive moderation procedures that are fairly applied to protect users' rights to freedom of expression and access to accurate information. On Snap Map, our editorial oversight protection for content showing up on Snap Map tries to strike the right balance between the need to preserve the Teen's safety versus the free flow of information and expression. Examples of this include in February of last year when Russia moved into Ukraine, Snap Map developed tooling that allowed us to block all of Ukraine from creating content. This was in response to concerns that Russia was using it for their own strategic purposes (propaganda and tracking the movement of Ukrainians).
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, the pool of content recommended by our algorithmic systems does not generally include political or other important societal matters regardless of where they fall on the political



	spectrum.
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	Yes, other mitigations listed here also apply to our Advertising Systems.
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	Yes, we consult with our safety advisory board to ensure Snapchat is set up appropriately for Teens and monitor community reports for issues relating to freedom of expression.
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	No, we do not work with trusted flaggers for Teen's rights to freedom of expression specifically, however we are working with trusted flaggers on children's safety in general.
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, our Crisis Protocols balance Teen's rights to freedom of expression with access to accurate information. We have recently exercised these protocols successfully during the French riots in June 2023.</p> <p>Note, we are actively working to support efforts to agree an EU Age appropriate design code to protect children's rights.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	Yes, we provide guidance on our terms, harms (including harmful false or deceptive information), moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to allow Teens to express themselves without the pressures of friends lists, comments and likes. We have community, ad and content guidelines that are specific to teens. We also offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity</p>	Yes, for example we display an icon in some Lenses that manipulate an image of a Snapchat to



<p>Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>make them look younger.</p>
---	--------------------------------

Conclusion

Snap considers children’s rights to be a lower likelihood risk but one that has a risk for severe harm industry wide, without appropriate mitigations. As a result we treat this as one of our highest priority risks, with a Level 1 Risk Prioritization. Snap is designed to fairly apply rules on content publication and provide an appropriate environment for Teens to exercise expression and assembly on Snapchat’s in-scope services (and Snapchat as a whole). As explained in the [Freedom of Expression](#) and [Protection of Minors](#) section of the Report, this includes adapting our systems to limit the access of Teen accounts to higher risk features and content, like public profiles and sexually suggestive content, as well providing Teens and Families with accessible guidance and tools for the use of Snapchat and ensuring our [Terms](#), [Moderation](#) and [Enforcement](#) also operate fairly.

We have concluded therefore that Snapchat’s in-scope services have reasonable, proportionate and effective mitigation measures to protect against negative effects on children’s rights. In addition, we are actively participating in efforts to develop an EU wide AADC to assess if further industry measures are needed.

4.2.7 Right to consumer protection

We understand that without mitigations online platforms can be used to spread content that contains false or misleading information that can harm consumers. This risk to consumer protection rights poses a serious threat to the rights of EU citizens who may be vulnerable to deception or invasion of privacy.

Likelihood

Given our assessment of the low likelihood of other related risks in Section 4.1 above (such as harmful false information), Snap considers that the risk to the right of consumer protection falls within the **Extremely Low Likelihood** category.



Severity

Snap has assessed information published by governments and other third party sources and considers that if an online platform were to undermine the right to consumer protection, this could fall within our **'significant harm' category**.

Overall potential risk prioritization

The potential for negative impacts to consumer protection rights falls within our Level 3 overall potential risk prioritization category, given the Extremely Low Likelihood and significant harm categorization.

Snap's Mitigations

Highlights

To mitigate these risks, Snap takes a multipronged approach. Snap has invested considerable resources in developing and enforcing advertising policies that safeguard consumer protection rights. We have robust ad policies to prevent inappropriate and illegal advertising on our platform, and we use a combination of automated and human review to prevent ads that violate our policies or the law from appearing on Snapchat. This also includes ensuring inappropriate ads are not targeted at Teens. Additionally, all ads can be flagged by Snapchatters in the app as being inappropriate along with the reason for the violation.

Separately, to ensure users know when content is commercial in nature, we automatically place an "Ad" marker on all paid ads that run on Snapchat. Our Commercial Content Policy requires all organic content posted by influencers to be marked appropriately and we now offer a "Paid Partnership" tag tool that influencers and users may use when they post commercial content to help them comply with this policy and their legal obligations.

To address potential risks with targeted advertisements, and to ensure advertisers are not manipulating small audiences with micro-targeted campaigns, most of the ads on Snapchat, including all political ads, require a specific minimum audience of Snapchatters to be targeted. We also offer special targeting models that do not include gender or age for advertisers who are advertising in the housing, credit or employment (HCE) spaces and are subject to specific legal requirements relating to those ads. Lastly, to ensure that users have choice about use of their personal data for targeting ads, we allow users to control the data that's used to determine the ads they see. In the EU, we offer controls to turn off most personalization of ads and for other regions users can restrict our use of third party data and being included in advertiser supplied audience matches for ads targeting.

Our Community Guidelines prohibit spreading false information that causes harm or is malicious, impersonation, i.e., attempting to deceive people about who you are, and disallow spam and other deceptive practices. Our Commercial Content Policy also disallows false or misleading



content, including deceptive claims, offers, functionality, or business practices, promotion of fraudulent goods or services, products or services with false celebrity testimonials or usage, deceptive financial products, and other similar content. Through these mitigations, Snap has been able to effectively uphold users' consumer protection rights.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	<p>Yes, Snapchat's in-scope services have been adapted to include proactive moderation for ads that violate our policies, or false or deceptive content.</p> <p>Snap places a strong emphasis on its adherence to Article 25 DSA concerning dark patterns. Consequently, this constitutes a strategic mitigation measure aimed at mitigating the potential impediment to consumer protection. Snap is committed to ongoing monitoring of this aspect to ensure continued compliance and effectiveness.</p> <p>We also require a specific minimum audience of Snapchatters to prevent advertisers from manipulating small audiences.</p>
Terms and Enforcement Adapting their terms and conditions and their enforcement.	<p>Yes, our advertising policies safeguard consumer protection and they are strictly enforced.</p>
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where	<p>Yes, specific proactive and reactive moderation procedures to prevent and remove ads that violate our policies, false or deceptive content.</p>



<p>appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>We also don't allow user-generated political content from being promoted on Spotlight. We take these measures in order to circumvent the spread of harmful and false content.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend false or deceptive content.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, our Advertising Systems are set up to safeguard consumer protection. For example, we automatically place an "Ad" marker on all paid ads.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency report false information, impersonation, spam and other regulated goods.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to regulated goods.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. CIPL, FPF.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we have an ad marker on all ads and provide transparency on our privacy practices including ads on our Privacy Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, for example we prevent inappropriate ads for Teens and advertising based on profiling. We make available robust reporting; and we offer Family Center and provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will</p>



	provide additional guidance for parents and carers on risks and support.
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	General content authenticity measures.

Conclusion

Snap considers consumer protection risks to fall within our overall Level 3 potential risk prioritization category given the widespread availability of this false and misleading content on the internet. In response it has put in place a range of mitigation measures. This includes, for example, developing and enforcing advertising policies that safeguard consumer protection rights. Our ad policies aim to prevent inappropriate and illegal advertising and our review processes were designed to enforce these policies. Because of safeguards in the product design and policy enforcement, to effectively diminish the likelihood that consumer protection rights are violated on the platform, this risk falls within the lowest likelihood level.

We have concluded therefore that Snapchat’s in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on consumer protection rights.

4.2.8 Right to Property

The property right that has a significant risk of being impacted by Snapchat’s in-scope services is the right to intellectual property. This risk stems from the disclosure of such property contrary to the intellectual property rights of a natural or legal person. This is discussed above under [Section 4.1.5](#) (Dissemination of content that infringes on intellectual property rights).

In addition, we consider there is a potential risk that individuals may harm someone else’s property while under pressure to create content that others find entertaining or humorous. This risk is discussed above under [Section 4.1.9](#) (Dissemination of content encouraging or engaging in violent or dangerous behavior).



4.3 Category 3: Negative effect on democratic and electoral processes, civic discourse and public security

(Article 34.1.c / DSA Recital 82)

In this part of the Report, we explain the results of our assessment on actual or foreseeable negative effects of Snapchat's in-scope services on our public security as required by Article 34.1.c and Recital 82 of the Digital Services Act. We have assessed in particular negative effects on democratic processes, civic discourse and electoral processes, as well as public security.

Category 3 - Negative effect on Public Security					
Category	Relative likelihood of risk occurring on Snapchat	Harm classification industry wide	Risk Prioritization	Approach	Conclusion
4.3.1 Negative effect on Democratic and Electoral Processes	Extremely Low Likelihood	Severe harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.3.2 Negative effect on Civil Discourse	Extremely Low Likelihood	Severe harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.3.3 Negative effect on Public Security	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations

4.3.1 Negative Effect on Democratic and Electoral Processes

The role of digital platforms in helping to shape information environments establishes a significant nexus with democratic and electoral processes. As digital technologies such as Snap enable expression and access to information, the impact of these platforms on the free and fair exercise of political rights warrants careful attention, presenting risks to which Snap has long been vigilant.



Likelihood

Snapchat's platform architecture, combined with its commitment to responsible policy enforcement across our content surfaces, establishes unique safeguards against risks to democracy. We understand well that online platforms may have a negative effect on the electoral processes and the exercise of political rights by amplifying digital disinformation or deceptive content relating to political matters or processes. The low cost and efficiency of using digital platforms to wage influence operations or spread political propaganda means that the likelihood of negative impacts to democracy is probable on high-reach digital surfaces that do not take steps to reduce risks. However, the steps Snap has taken to mitigate threats to democracy mean that likelihood is substantially diminished.

Independent reports of electoral interference on Snapchat are vanishingly rare. In connection with a major, high-profile election in 2022, we onboarded Snap to the Election Integrity Partnership (EIP),¹⁷ a partnership among leading research centers and civil society organizations who monitor online harms to democratic processes; as participants in the EIP threat escalation program, our teams received only one single incident report from the researchers monitoring risks on Snapchat.

Snap's own reporting metrics confirm the limited occurrence of content harmful to democracy. Our most relevant reporting category on this topic is "harmful false information," which our policies define as including content that "undermines the integrity of civic processes." Our most recent data indicates that from July to December of 2022, the total number of enforcements globally for harmful false information represented just 0.1% of total content enforced, falling within the lowest likelihood category.

Our assessment is that Snap's product design and policy practices substantially reduce the likelihood of negative impacts on democracy. The facts and figures above go to show that the risk of potential negative impact on democratic and electoral processes falls into the **extremely low likelihood category**.

Note: There were a large number of elections in 2024, including the EU Parliamentary elections. Overall, the European elections unfolded in a [positive](#) online environment with no major threats. This was confirmed by the European Commission and independent observers, who confirmed that they did not observe major online threats.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that has a negative effect on democratic and electoral processes were to materialise on an online platform, this could fall within our '**severe harm**' category.

¹⁷ Election Integrity Partnership (2020), [url](#).



Overall potential risk prioritization

Taking into account the real-world examples illustrating the potential disruptive effects on democracy, this is a severe harm risk if not mitigated. However, there is extremely low prevalence and this results in a Level 3 overall potential risk prioritization.

Snap's Mitigations

Highlights

Snap takes a multifaceted approach to mitigating negative impacts to democracy, including policy enforcement, product design, and expert engagement.

Snap's policies expressly prohibit content that undermines the integrity of elections and civic processes. Drawing from expert research from the Election Integrity Partnership,¹⁸ we orient this policy around four pillars of risk:

- Procedural interference: misinformation related to actual election or civic procedures, such as misrepresenting important dates and times or eligibility requirements for participation.
- Participation interference: content that includes intimidation to personal safety or spreads rumors to deter participation in the electoral or civic process.
- Fraudulent or unlawful participation: content that encourages people to misrepresent themselves to participate in the civic process or to illegally cast or destroy ballots.
- Delegitimization of civic processes: content aiming to delegitimize democratic institutions on the basis of false or misleading claims about election results, for example.

In addition, our [political ad policies](#) ensure that any political advertisements are subject to review and fact-checking *before* they are eligible for placement on Snapchat.

We also prevent advertisers from manipulating small audiences with micro targeted campaigns, particularly for political ads. We do so by requiring a specific minimum audience of Snapchatters to be targeted (including [Dynamic Ads on Snapchat | Snapchat for Business](#)).

In 2021 Snap joined the Dutch Code of Conduct for political ads.¹⁹ Under this Code online platforms agreed to acknowledge a responsibility in maintaining the integrity of elections and avoid dissemination of misleading content and messages inciting violence or hate speech on their platforms, committed to making key data on online political advertising available publicly and help avoiding foreign interference in elections by banning political advertisements from

¹⁸ Election Integrity Partnership, 'Evaluating Platform Election-Related Speech Policies, October 2020, [url](#).

¹⁹ For more details [url](#).



outside the European Union, and putting in place a user-friendly response mechanism to answer questions or solve problems related to the Dutch elections.

We also mitigate these risks through intentional product design choices. Our platform does not, for example, provide an unvetted feed of algorithmically curated political information; we disallow *all* political content from Spotlight (our broadcast platform for User Generated Content) and pre-moderate that surface to ensure that such political content is not distributed.²⁰ This safeguard ensures that Snap is not algorithmically promoting political statements from unvetted sources, and generally reflects Spotlight’s function as an entertainment platform. (Consistent with our commitments to fundamental rights of expression and access to information, Snapchat provides other, non-algorithmically amplified spaces for users to express their views and political observations, such as chat and My Story; users can also seek access to political information from known publishers and creators whom Snap has on-boarded for distribution on the Stories tab).

All Spotlight content goes through both automoderation and human review before it is eligible for distribution to a wide audience. Content that is approved for broader audiences must comply with our Community Guidelines and our Content Guidelines for Recommendation Eligibility. Any content that is reported will be reviewed against these guidelines again for compliance.

Political content is only eligible for broadcast (aka algorithmic distribution) on Snapchat on surfaces reserved for publishers or creators with whom Snap engages in partnership, or through advertising (where such content is pre-reviewed and fact-checked).

Snap does not allow Lenses that encourage a particular political perspective. In line with this approach, politically related Lenses are disabled in For You. Snap also rejects Lenses that perpetuate false information to elections (e.g. the wrong date). AR moderators are given strict guidance during elections to escalate misinformation.

We also engage with experts in the information integrity and democracy and human rights community to inform our approach. This includes collaborations and engagement with individual experts (for example, former US Ambassador to the UN Human Rights Council Eileen Donahoe, in addition to several others), as well as think tanks (such as the Atlantic Council’s Digital Forensics Research Lab) and research collaborations (such as the Election Integrity Partnership). Additionally, we partner with governments around the world to inform Snapchatters about elections and invite them to go vote. A recent example was the ‘23 Dutch provincial election cycle. With the Dutch Ministry of the Interior, Snap developed a lens where Snapchatters could place voting bins in their living room and answer questions about the election with ‘true’ or ‘false’.

²⁰ For these purposes, “political content” means content related to political campaigns and elections, government activities, and/or viewpoints on issues of ongoing debate or controversy. This includes content about candidates or parties for public office, ballot measures or referendums, and political action committees, as well as personal perspectives on candidate positions, government agencies/departments or the government as a whole.



By taking this quiz Snapchatters are increasing their knowledge about the elections and are reminded to go vote.

We take steps to explain our policy approach to safeguarding democratic information environments through our [Community Guidelines](#) and periodic [blog posts](#).

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, as outlined above our platform does not, for example, provide an unvetted feed of algorithmically curated political information; we disallow <i>all</i> political content from Spotlight (our broadcast platform for User Generated Content) and pre-moderate that surface to ensure that such political content is not distributed.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, we take steps to explain our policy approach to safeguarding democratic information environments through our Community Guidelines and periodic blog posts .
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, we algorithmically prevent political content from being promoted on Spotlight. Snap does not allow Lenses that encourage a particular political perspective. In line with this approach, politically related Lenses are disabled in For You. Snap also rejects Lenses that perpetuate false information to elections (e.g. the wrong date). AR moderators are given strict guidance during elections to escalate misinformation.
Algorithmic Systems	Yes, our algorithmic systems do not promote political content on Spotlight.



<p>Testing and adapting their algorithmic systems, including their recommender systems.</p>	
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, our political ad policies ensure that any political advertisements are subject to review and fact-checking <i>before</i> they are eligible for placement on Snapchat.</p> <p>We prevent advertisers from manipulating small audiences with micro targeted campaigns, particularly for political ads, by requiring a specific minimum audience of Snapchatters.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we review and monitor compliance with our internal terms, policies and procedures.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Our Trusted Flaggers are not typically focussed on this risk, but we welcome their input on this matter.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes. We engage with experts in the information integrity and democracy and human rights community to inform our approach. This includes collaborations and engagement with individual experts (such as former US Ambassador to the UN Human Rights Council Eileen Donahoe, global democracy scholar and Stanford Professor Larry Diamond, and several others), as well as think tanks (such as the Atlantic Council's Digital Forensics Research Lab) and research collaborations (such as the Election Integrity Partnership). Additionally, we partner with governments around the world to inform Snapchatters about elections and invite them to go vote.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we proactively encourage our users to go to vote through interactive campaigns.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we limit exposure to political content to Teens, but do educate Teens with trusted new sources on current events and inform users how they can participate in a democratic society. We offer Family Center; we make</p>



	<p>available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General Content Authenticity measures. We also label political advertisements, and maintain a political ads library.</p>

Conclusion

Snap considers the overall risk potential of negative impact on democratic or electoral processes to be in the Level 3 category, given severity of potential harm. However, as described above, Snap has put in place numerous specific mitigations, such as algorithmically preventing the promotion of political content in Spotlight, enforcing political ad policies, and disallowing Lenses encouraging political perspectives. Further, the design and function of Snapchat is such that it is not conducive for the widespread distribution of viral content and we provide robust in-app reporting, which further mitigates this harm. Snap recognizes the importance of democratic and electoral processes, and in fact has created interactive campaigns to raise awareness and encourage users to vote. Our prevalence data and our continuing monitoring efforts cited above show that our safeguards are effective at mitigating these risks on Snapchat.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on democratic and electoral processes.

4.3.2 Negative Effect on Civil Discourse

We recognize that without adequate mitigations, digital content platforms like Snapchat can contribute to negative effects on civil discourse. Across Snap's various products, these risks could include:

- The potential for personalized content and algorithmic biases lock users into echo chambers, reinforcing existing beliefs and potentially leading to polarized communities, which hinders open dialogue.



- The risk of amplified dis- and misinformation negatively impacting public opinion on important civic issues.
- The possibility of amplification of extreme or sensational content to retain user attention leading to heightened polarization and a hostile online environment.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

While it is rather difficult to classify the likelihood of such a comprehensive social issue, we can revert to and deduce from the reporting data available to us. We are unaware of any third-party reports identifying these risks on Snapchat. Our own reporting data suggests that policy violations related to harming civic discourse (i.e., our “harmful false information” and “hate speech” categories) are encountered rarely. Harmful false information represents just 0.1% of total content enforced in the second half of 2022. Our [prevalence](#) testing showed that hate speech made up an extremely low percentage of Policy Violating Prevalence (PVP) on Snap in August of 2023. Consequently, in terms of likelihood, we consider this risk falls within our **Extremely Low Likelihood** category.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that has a negative effect on civil discourse were to materialise on an online platform, this could fall within our ‘**severe harm**’ category.

Overall potential risk

Snap considers the dissemination of information with actual or foreseeable negative effects on civic discourse to fall within our severe harm category. Given the apparent low prevalence, overall, this risk falls within our Level 3 overall risk prioritization category.

Snap's Mitigations

Highlights

Snap’s policies prohibit the spread of “harmful false information,” which we define as false content that may result in broadly distributed harm, or is malicious. Referencing our internal policy



guidance, Snap enforces content as harmful false information if both of the following elements are present:

- Information is determined to be false
- The false information could cause “broadly distributed harm”. “Broadly distributed harm” refers to harms that undermine societal- or community-level safety or security; harms that undermine public health; harms that undermine civic processes or the exercise of political rights; and harms that denigrate the memory or history of peoples and tragic events.

In addition to our internal policies, Snap’s Community Guidelines also note that harmful false information is prohibited and includes denying the existence of tragic events, unsubstantiated medical claims, or undermining the integrity of civic processes – all of which could contribute to negative impacts on civic discourse.

Snap policies also prohibit the use of hate speech, hate symbols, and/or content that valorizes the perpetrators of, or denigrates the victims of, human atrocities such as genocide.

We define hate speech as content that demeans, or promotes discrimination towards, an individual or group of individuals on the basis of their race, color, caste, ethnicity, national origin, religion, sexual orientation, gender identity, disability, veteran status, immigration status, socio-economic status, age, weight, or pregnancy status. Our policies note that hate speech may include references to people that are dehumanizing or that compare humans to animals on the basis of these traits and categories. Hate speech also includes the valorization of perpetrators—or the denigration of the victims—of hateful atrocities (e.g., genocide, apartheid, slavery, etc.), as well as the promotion of hate symbols.

Under Snap’s policies, hate symbols include imagery that is intended to represent hatred or discrimination toward others, including those featured in the hate symbols database maintained by the Anti-Defamation League (ADL).²¹

Snap establishes additional safeguards against risks to civic discourse on our surfaces that help distribute content algorithmically. All Spotlight and For You content goes through both automoderation and human review before it is eligible for distribution to a wide audience. Content that is approved for broader audiences must comply with our Community Guidelines and our [Content Guidelines for Recommendation Eligibility](#). Any content that is reported will be reviewed against these guidelines again for compliance.

Snap has also made intentional product choices to mitigate risks to civic discourse; this includes the absence of algorithmically promoted groups, which have been shown to contribute to echo chambers and to be vectors for misinformation, with negative consequences for civil discourse.²²

²¹ The ADL database is available at: [url](#).

²²The Verge, ‘Facebook will stop recommending health groups’, September 2020, [url](#).



In addition, many of our surfaces are not ideal vehicles to cause risks to civil discourse. For example, unless saved to your Public Profile, Public Stories and Snaps on the Map are only available for a maximum of seven (7) days (and often much shorter), which limits their arc of influence. Similarly, there is considerable technical expertise required to create a Lens, making it a difficult surface (compared to other third party platforms) to navigate for the purpose of broadly distributed harm.

To remain vigilant against threats to civil discourse, Snap engages with experts from across civil society and the research community who study information integrity and resilience to online harms. These engagements include consultations and collaborations with online safety experts (including those represented on Snap's Safety Advisory Board), with organizations combating online hate (such as the Anti-Defamation League), and engagement with research organizations, including the Atlantic Council Digital Forensics Research Lab and the Digital Wellbeing.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, as outlined in the section on Democracy our platform does not, for example, provide an unvetted feed of algorithmically curated political information; we disallow <i>all</i> political content from Spotlight (our broadcast platform for User Generated Content) and pre-moderate that surface to ensure that such political content is not distributed.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, we take steps to explain our policy approach to safeguarding civil discourse information environments through our Community Guidelines and periodic blog posts .
Moderation	Yes, all Spotlight content goes through both automoderation and human review before it is



<p>Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>eligible for distribution to a wide audience. Content that is approved for broader audiences must comply with our Community Guidelines and our Content Guidelines for Recommendation Eligibility. Any content that is reported will be reviewed against these guidelines again for compliance.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not create echo chambers and ensure users are subject to different types of content and viewpoints.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, our political ad policies ensure that any political advertisements are subject to review and fact-checking <i>before</i> they are eligible for placement on Snapchat.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we review compliance with our terms and processes.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Our Trusted Flaggers are not typically focussed on this risk, but we welcome their input on this matter.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes. We engage with experts in the information integrity and democracy and human rights community to inform our approach. This includes collaborations and engagement with individual experts (such as former US Ambassador to the UN Human Rights Council Eileen Donahoe, global democracy scholar and Stanford Professor Larry Diamond, and several others), as well as think tanks (such as the Atlantic Council's Digital Forensics Research Lab) and research collaborations (such as the Election Integrity Partnership). Additionally, we partner with governments around the world to inform Snapchatters about elections and invite them to go vote.</p>
<p>Transparency</p>	<p>Yes, we proactively encourage our users to go to vote through interactive campaigns.</p>



<p>Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we limit exposure to political content to Teens, but do educate Teens with trusted new sources on current events and inform users how they can participate in a democratic society. We offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General Content Authenticity measures. We label political advertisements, and maintain a political ads library.</p>
<p>Other Mitigations</p>	<p>We prevent advertisers from manipulating small audiences with micro targeted campaigns, particularly for political ads, by requiring a specific minimum audience of Snapchatters.</p>

Conclusion

Similar to our conclusion on negative impacts to democracy and elections, Snap considers the overall risk potential of negative impact on civil discourse to be in the Level 3 category, given the severity of potential harm of dis- and mis-information and online echo chambers which can create a hostile online environment. In response, Snap has put in a range of mitigation measures that, in most cases, overlap with the mitigations for risks to democracy and elections. These mitigations include proactive content moderation, enforcement of our Community Guidelines and Terms, a restriction on political content which is a high risk area for dis- and mis-information, and engagement with outside experts and trusted flaggers. Snap also takes positive, rather than reactive or punitive mitigations, including encouraging Snapchatters to vote and participate in civil discourse, and audience minimums to preempt ad microtargeting. Although we do not document and report on civil discourse as a category, this would primarily fall within the dissemination of disinformation (Category 1), which has an extremely low prevalence, and



dissemination of hate speech (Category 1), which also has an extremely low prevalence. We take facilitating and encouraging civil discourse very seriously, and view this matter as important to the value of Snapchat to our users. As such, we continue to invest in measures to prevent any content that negatively impacts civil discourse from reaching a broad audience on Snapchat, which may undermine our goal of allowing users to live in the moment and enjoy the world around them. We also provide users with tools to report content and support resources online and in-App, and we hold our advertisers to standards that prevent false, misleading, or micro targeted advertising.

We have concluded therefore that Snapchat’s in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on civil discourse.

4.3.3. Negative Effect on Public Security

Without appropriate mitigations, we recognise that digital platforms may present risks to public security, particularly in the form of harmful, dangerous, or inciteful content; these risks may become compounded when such content may be amplified at great scale and distributed with high velocity. The design of Snap’s products and platform architecture scrupulously accounts for these risks; accordingly, we’ve implemented a number of key safeguards that help to advance both the safety of Snapchatters and the interests of public security across our services.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

While it is rather difficult to classify the likelihood of such a comprehensive social issue, we can revert to and deduce from the reporting data available to us. We assess that the volume of content presenting risks to public security is quite low on Snapchat. To inform this conclusion, we’ve looked at the [prevalence](#) (PVP) of harm categories that may contribute to undermining public security, which include the following: harmful false information (extremely low prevalence); hate speech (extremely low prevalence); violent or disturbing content (extremely low prevalence); and terrorism and violent extremism (extremely low prevalence). The low prevalence rate of these harms supports an assessment that our mitigations are effective and, consequently, it is uncommon to encounter these harms on Snapchat. In terms of likelihood, this risk would fall within our **Extremely Low Likelihood** category.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.



Severity

Snap has assessed information published by governments and other third party sources and considers that if content that has a negative effect on public safety were to materialise on an online platform, this could fall within our **'severe harm' category**.

Overall potential risk prioritization

Taking into account the real-world examples illustrating the potential disruptive effects on public security, this is a severe risk if not mitigated. However, we are encouraged—based on relevant prevalence data cited above—that our safeguards are substantially effective at mitigating these risks on Snapchat. The combination of low prevalence and severe nature results in a **Level 3** overall potential risk prioritization categorisation.

Snap's Mitigations

Highlights

We take several steps to ensure that we are addressing this risk across Snap's products and services, including enforcement of several relevant platform policies and internal crisis protocols for managing high-risk scenarios.

Snap's policies include several prohibitions that are enforced vigorously and equitably to support the interests of public security. These policies include a prohibition against spreading harmful false information. Internal policy guidance instructs that violations of these policies include risks to public security such as Snaps denying the holocaust or a school shooting, or information obtained illegally that is being shared to embarrass the person from whom the information was stolen.

Snap's policies also include prohibitions on content promoting terrorism or violent extremism, as well as "content that attempts to incite, glorify, or depict real violence that results in personal injury or death," and "depictions of human violence, child abuse, animal abuse, or gore."

We may also consider off-platform behavior when assessing risks to public security. Our Community Guidelines state expressly that "Snap reserves the right to remove or restrict account access for users whom we have reason to believe, in our sole discretion, pose a danger to others, on or off of Snapchat. These include leaders of hate groups and terrorist organizations, individuals with a reputation for inciting violence, perpetrating severe harms against others, or behavior that we believe poses a threat to human life."

Taken together, these several policy provisions provide a basis for appropriately actioning any content that poses an acute risk to public security.



In addition, we have internal operational protocols for responding to public crises. These protocols include the following steps:

- Our vendor teams carefully apply the [Community Guidelines](#) and [Content Guidelines for Recommendation Eligibility](#) to ensure the content is assessed appropriately against our rules (for example, routinely distinguishing between *documenting* violence and *advocating for* violence).
- When breaking news happens, such as ongoing violent protests, the vendor teams connect with our full-time content review team to summarize the kind of content they are encountering (e.g., violence, property damage, fires, expressions of criticism or support for various political positions), and summarize how they are currently actioning that type of content against our existing guidelines.
- That summary list comes to our Platform Policy team for review. Almost all of the time, Policy's answer is that they're actioning content correctly. (To cite a recent example, in the case of French protests over the course of this summer, our team determined that existing policies and procedures were working as intended.)
- In the event that the Platform Policy team determines that the policies are not being applied appropriately, the team will expeditiously draft clarifying guidance for vendors and content review teams. The draft guidance will be shared among relevant internal leaders for review before being distributed to operational teams.

In addition, as noted in the [Civil Discourse section](#), many of our surfaces are not ideal vehicles to cause risks to public security. For example, unless saved to your Public Profile, Public Stories and Snaps on the Map are only available for a maximum of seven (7) days (and often much shorter), which limits their arc of influence. Similarly, there is considerable technical expertise required to create a Lens, making it a difficult surface (compared to other third party platforms) to navigate for the purpose of broadly distributed harm.

Separately, we maintain tight internal protocols for escalating terrorist content or other imminent threats to the appropriate legal or emergency authorities. In such cases, vendors and review teams are trained to preserve relevant information and immediately send a report to Snap's Law Enforcement Operations team, who are professionally trained to appropriately engage with legal and emergency authorities.

This approach reflects Snap's deep commitment to public safety, and serves our community well to reduce negative impacts to public security.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many



of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat is not an attractive platform for spreading content that may have a negative impact on public security, including harmful, dangerous, and inciteful content, in particular because it is difficult to reach a broad audience, and Snapchat has made conscious design decisions to restrict the ability for content to go viral, including not offering a reshare functionality and applying short retention to content. On surfaces where a broader audience can potentially be reached our proactive detection makes it difficult for content that may have a negative impact on public security to reach a large audience. Moreover, our content platform, For You, features content from approved media publishers and content creators. Our entertainment platform for user-generated content, Spotlight, is proactively and a priori moderated before content can reach a wide audience.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, terms prohibit content that may have a negative impact on public security, including harmful, dangerous, and inciteful content, and they are strictly enforced.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to prevent and remove content that may have a negative impact on public security, including harmful, dangerous, and inciteful content.</p> <p>Content on Spotlight undergo rigorous moderation practices as reported in the Moderation section as well as under Risk Category 1. All content on Spotlight is subject to human review pursuant to our Broadcast UGC Policies that are further described in Moderation.</p>



	<p>All For You UGC content is moderated by humans, and we proactively remove content that doesn't meet our policies before being broadly distributed.</p> <p>Furthermore, all Media Partners are vetted prior to being permitted to distribute their content broadly on Snapchat by a team of editors. Media Partners go through an editorial review of their content, a reputational search (to evaluate if a Media Partner has a history damaging press, legal actions, etc.), and compliance review before they're able to distribute content.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend content that may have a negative impact on public security, including harmful, dangerous, and inciteful content.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for content that may have a negative impact on public security.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers, our trusted flagger may also report content that may have a negative impact on public security, but this rarely happens because of the limited amount of this type of content on Snapchat.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>



<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>

Conclusion

Snap considers the negative impact to public security to have a Level 3 overall potential risk given the potential disruptive effects of content that can, among other things, harm, put in danger, and incite the public at large. That being said, Snap has put in place a range of mitigation measures to bring the likelihood of this risk from coming to fruition into the lowest category. These measures include our proactive content moderation which is designed to detect and prevent hateful, dangerous, and inciteful content from reaching a broad audience on Snapchat's in-scope services. As noted in other sections, we continue to invest in measures that prevent this type of content from reaching a broad audience on Snapchat, as well as provide our users with tools to report content to Snapchat and law enforcement, and support our community via online and in-app support tools. As a result, the volume of content presenting risks to public security is low on Snapchat.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on public security.



4.4 Category 4: Negative Effects on Public Health

(Article 34.1.d / DSA Recital 83)

In this part of the Report, we explain the results of our assessment on actual or foreseeable negative effects of Snapchat’s in-scope services on our public health as required by Article 34.1.d and Recital 83 of the Digital Services Act. We have assessed in particular negative effects on public health, gender-based violence, Teens, as well as serious negative consequences to a person’s physical and mental well-being. We have considered risks relating to the design, functioning or use, including through manipulation such as by coordinated disinformation campaigns related to public health, or from online interface design that may stimulate behavioral addictions of recipients of the service.

Category 4 - Negative Effects on Public Health					
Category	Relative likelihood of risk occurring on Snapchat	Harm classification industry wide	Risk Prioritization	Approach	Conclusion
4.4.1 Negative Effects on Public Health	Extremely Low Likelihood	Significant harm industry wide	Level 3	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.4.2 Negative Effects on gender-based violence	Extremely Low Likelihood	Serious harm industry wide	Level 2	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations
4.4.3 Negative Effects on Children	Varies	Severe harm industry wide	Level 1	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations and Snap is participating in efforts to develop an EU wide AADC to assess whether further measures should be taken industry wide.
4.4.4 Serious Negative Consequences on physical and mental well-being	Extremely Low Likelihood	Severe harm industry wide	Level 1	Mitigated	Low Risk / Reasonable, proportionate and effective mitigations



4.4.1 Negative Effects on Public Health

We believe the health and wellness of the public and our users is paramount to our goal to be a platform of fun and freedom of expression. We recognize that, without adequate mitigations, digital content platforms like Snapchat could contribute to negative effects on public health.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

With regards to coordinated disinformation campaigns related to public health, as well as any dissemination of content that promotes harmful/unhealthy behavior (e.g., eating disorders or other self-harm content), we are encouraged that available data suggests prevalence of content on Snapchat related to these risks is quite low. As can be seen in the section on [prevalence testing](#), we have measured the prevalence of health misinformation and other harmful false information to be extremely low, and the prevalence of self-harm content is extremely low. For the potential negative effects on physical and mental well-being, see the relevant [section](#).

Therefore, we consider that currently the risk of negative effects on public health arising from Snapchat's in-scope services to fall within our extremely low likelihood category.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that has a negative effect on public health were to materialise on an online platform, this could fall within our **'severe harm' category**.

Overall potential risk prioritization

Snap assesses negative impacts to public health to present a systemic, severe risk that must be appropriately mitigated. Given the severe nature but extremely low prevalence, this would classify as a Level 3 risk in terms of Snap's overall potential risk prioritization matrix.

Snap's Mitigations

Highlights

Snap's [Community Guidelines](#) prohibit the spread of harmful false information, expressly disallowing content that includes unsubstantiated medical claims. Our policies elaborate that



such prohibited content includes any content that, for example, recommends untested therapies for preventing the spread of Covid-19; or that features unfounded conspiracy theories about vaccines.

Our [Community Guidelines](#) also prohibit “the glorification of self-harm, including the promotion of self-injury, suicide or eating disorders.” Our global approach to harm mitigation relies on teams, technologies, policies, and partnerships to help keep Snapchatters safe, healthy, and informed; however, content related to self-harm and suicide implicates unique sensitivities for which our efforts must account. We therefore take a tailored approach to this category of content – one that includes: (1) supportive interventions, (2) features promoting a culture of support, and (3) a considerate approach to policy enforcement and self-harm content removal. Each of these will be explained briefly in turn:

1. *Supportive Interventions*: In response to troubling search inquiries or content indicating mental or emotional distress, our products and teams intervene to surface mental health resources and support (either automatically, or at the discretion of Trust and Safety personnel). These resources are tailored to a user’s geographic region.
2. *Culture of Support*: Snapchat offers well-being features designed to educate and empower members of the Snapchat community to support friends who might be struggling with their social or emotional well-being. These features include “[Here for You](#)” content Snap has developed with the intention of educating Snapchatters about the importance of mental health, and ways to seek support.
3. *Considerate Policy Enforcement*: Especially given the risks of glorification, Snap’s policies prohibit content that depicts suicide or self-harm; however, since Snapchat is used for communication with friends and family, it is important to us that our enforcement actions do not deprive users’ friends and family of important distress signals and an opportunity to intervene. Accordingly, we instruct reviewing agents that:
 - Reported depictions of suicide or self-harm that reflect an emergency situation should be removed and possibly escalated to law enforcement or emergency authorities.
 - Content glorifying or inciting self-harm must be removed and is subject to an enforcement “strike.”
 - Depictions of self-harm or suicidal ideation that do not reflect an emergency situation are permitted so that the community of people around this person can offer help and support.

To inform a responsible approach to mitigating these risks to public health, Snap regularly engages with experts from across the field of online safety, health, and wellbeing. Our [Safety Advisory Board](#) includes several such experts (including, for example, Dr. Michael Rich, pediatrician, founder and director of the Digital Wellness Lab & Clinic for Interactive Media and Internet Disorders, with affiliations at Boston Children's Hospital and Harvard Medical School).



These experts have been consulted specifically on Snap’s approach to wellness and mitigating risks related to mental and emotional duress, eating disorders, and other forms of self harm.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat’s in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, Snapchat’s in-scope services have been adapted to include proactive moderation to reduce the spread of harmful false information, including unsubstantiated medical claims, and the glorification of self-harm, including the promotion of self-injury, suicide or eating disorders.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, Our Terms prohibit the spread of harmful false information, including unsubstantiated medical claims, and the glorification of self-harm, including the promotion of self-injury, suicide or eating disorders.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	Yes, specific proactive and reactive moderation procedures to prevent and remove harmful false information, including unsubstantiated medical claims, and the glorification of self-harm, including the promotion of self-injury, suicide or eating disorders.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not categorize or recommend harmful false information or content that glorifies self-harm
Advertising Systems	Yes, other mitigations listed here also apply to our Advertising Systems.



<p>Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for harmful false information and self-harm content.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to illegal content that harms public health.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. EUIF. We also coordinate with our Safety Advisory Board on issues related to public health, as it contains experts from the medical community.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>We seek to protect all users from these harms. We offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>



Conclusion

Snap recognises that without adequate mitigations, digital content platforms like Snapchat can contribute to negative effects to public health. In response, it has put in place a range of mitigation measures. These include in particular our proactive content moderation, which is designed to detect and prevent content that may contribute to negative effects to public health from reaching a broad audience on Snapchat. Given that we have measured the prevalence of health misinformation, self-harm content and other harmful false information to be extremely low we believe our mitigations have been effective. We continue to invest in measures that prevent this type of content from reaching a broad audience on Snapchat, as well as provide our users with tools to support our community via online and in-app support tools, such as Here For You and our Safety Center resources.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on public health.

4.4.2 Negative Effects on gender-based violence

We strongly oppose content that promotes gender-based violence. We recognise that without mitigations, a recipient of an online platform's services could promote content considered to be gender-based violence.

Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Snap does not track gender-based violence as a specific, separate category in its prevalence and its transparency reports. However, this type of content is captured within the scope of broader categories tracked by Snapchat including content relating to hate speech, harassment, disinformation and violence. We are encouraged by data suggesting the likelihood of encountering such risks on Snapchat is within the lowest level. For example, data indicate that the [prevalence](#) (PVP) of hate speech is extremely low; for disinformation and all forms of harassment (including NCII, sexual harassment, and sextortion). From this, we can deduce that content promoting gender based violence also falls within our **Extremely Low Likelihood** category relative to other risks we have assessed.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that has a negative effect on gender-based violence were to materialise on an online platform, this could fall within our **'serious harm' category**.



Overall potential risk prioritization

Although the prevalence of content within the scope of this potential risk on Snapchat is considered to be at a lower level, due to the potential for serious harm to be caused by this content, Snap considers this to be a Level 2 overall potential risk for Snapchat's in-scope services.

Snap's Mitigations

Highlights

Snap takes a multifaceted approach to mitigating risks that may negatively impact gender-based violence. Our policies include several prohibitions against content that may contribute to such risks, including sextortion, sexual harassment, NCII, harmful false information (which may include gender-based disinformation campaigns), hate speech, and human trafficking. See details of the measures we put in place to mitigate these risks in section 4.1 above.

We also undertake intentional efforts to understand these problems across our community. As part of our [Year Two Digital Well-Being study](#), we conducted a deeper dive into teens' and young adults' exposure to "sextortion" across platforms and services. Released in June 2023, those findings showed that in our same six target countries – Australia, France, Germany, India, the UK, and the U.S. – two-thirds (65%) of Generation Z teens and young adults said they or their friends had, at some point, been targeted for sextortion. Details can be found at [this](#) Snap blog post published by the WeProtect Global Alliance on June 21, 2023, to coincide with Snap's participation in the Technology Coalition's biennial Multi-Stakeholder Forum in Washington, D.C., which focused on the financial sextortion of Teens.

We believe our approach to these challenges reflects our commitment to responsibly mitigating harms that may negatively impact gender-based violence, and we are encouraged by evidence that our approach has contributed to a low prevalence of such content on Snapchat.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **[link in the left hand column to a full summary](#)** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.



DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat’s in-scope services have been adapted to include proactive moderation for illegal hate speech and violence, which includes gender-based violence.</p> <p>We also take gender-based violence into consideration across the design of our services. For example, some Lenses can be used with Friends. Snap has removed tips to “Try with Friends” to some Lenses where there is a risk for bullying or harassment, including in relation to gender-based violence. In risky cases, Snap won’t encourage users to try a Lens with friends or Snap disables the Lens for being used with the rear camera (e.g. disabling this for the Pride Lens limits the ability to out someone else). These restrictions only applies to Lenses created by Snapchat.</p> <p>For Lenses submitted to Snapchat, we reject harmful Lenses to reduce the likelihood that they are distributed on Snapchat.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes, terms prohibit gender-based violence and they are strictly enforced.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to prevent and remove violent content, and illegal hate speech, as further detailed in previous sections of this report.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend violent or illegal hate speech content, which would include gender-based violence, as further outlined in the previous sections.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p>



<p>presentation of advertisements in association with the service they provide.</p>	
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, as outlined in the previous sections on hate speech and violent content.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to illegal hate speech.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. EUIF.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide guidance to parents on the web.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	<p>General content authenticity measures.</p>



Conclusion

Similar to the related hate speech risk category, Snap considers gender-based violence to fall within our Level 3 overall potential risk category. In response it has put in place a range of mitigation measures. This includes in particular our proactive content moderation which is designed to detect and prevent illegal hate speech, including gender-based violence related content from reaching a broad audience on Snapchat's in-scope services. Although we do not specifically document and report on gender-based violence as a category, this type of content would primarily fall within our hate speech category, which has an extremely low prevalence (PVP) on Snapchat. We take this matter very seriously, and continue to invest in measures that prevent this type of content from reaching a broad audience on Snapchat, as well as provide our users with tools to report content to Snapchat and law enforcement, and support our community via online and in-app support tools, such as Here For You and our Safety Center resources.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for gender-based violence.

4.4.3 Negative Effects on Children

We understand that, without mitigations, online platforms could have a negative impact on children and Teens. This is a risk we take seriously as Snap's priority is protecting the safety and wellbeing of our users whilst ensuring they have a positive experience online. Privacy, safety and security are key values of the company and at the core of our value proposition to our users.

Likelihood

As explained in [Snapchat Community](#) as part of our Introduction to this Report, Snapchat is used by a wide demographic, with 18-24 years making up the highest percentage of users of Snapchat. Nevertheless, there is still a percentage of our users who are Teens (13-17). Therefore we consider that children using Snapchat are just as likely to be exposed to the issues identified in this Report as other members of the Snapchat Community.

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms). We conclude that the relative likelihood of a risk of negative effects on children and Teens for Snapchat's in-scope services compared with other risks **varies** depending on the underlying concern as we consider: (i) harassment and bullying content falls within our very low likelihood category; (ii) adult sexual content and data protection fall within our low likelihood category; (iii) and other relevant risk categories fall within our extremely low likelihood category.

Severity

As with likelihood above, this severity of harm caused if a particular issue arises depends on the harm caused. However, we take the safety and wellbeing of the youngest members of our



community very seriously and recognise that this group is particularly vulnerable and if a particular risk materializes, there is an increased risk that the severity of the harm they suffer is higher. In general, therefore, children and Teens may suffer a risk of greater harm from the issues we have identified as they may be more vulnerable and we have chosen to place the severity of harm arising from an issue that negatively affects children in our **'severe'** category.

Overall potential risk prioritization

Although the relative likelihood for the negative effects on children varies, Snapchat considers the risk of harm to fall within the severest category. Consequently, Snap considers this to be a Level 1 overall potential risk.

Snap's Mitigations

Highlights

As the risk of negative effects on children falls within our highest risk prioritization level, Snap has dedicated extensive resources to ensuring protections to safeguard the rights of Teens on the platform, greatly reducing the likelihood of rights infringement. These safeguards include, for example, those described on the following pages:

- <https://values.snap.com/privacy/teens>
- <https://parents.snapchat.com/safeguards-for-teens>

We also work with Trusted Flaggers in the EU, and globally, on child safety issues, as well as our Safety Advisory Board. For more information on this, see [Section 6](#) of this Report.

Taken together, these mitigations contribute to a safe and responsible environment for young Snapchatters.

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.



DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes, Snapchat's in-scope services have been adapted to include proactive safety by design features and content moderation for Teens. For example:</p> <ul style="list-style-type: none"> • We have created a different product experience for Teens and adults. For example, we don't show sexually suggestive content to Teens. • All content on For You has to be appropriate for 13+. • Regulated goods don't appear in ads to Teens. • Snap Map is designed to mitigate particular risks to Teens. For example, location sharing is off by default.
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes. For example, our Terms require that all content is appropriate for 13+, we require all users on our platform to declare they are over the age of 13. We strictly enforce our terms and if we discover that a user is under the age of 13 we will remove their account.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to age gate and ensure age-appropriate content (for example restricting Teens access to suggestive content), adjust content settings as designated in Family Center and remove reported content from view.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems take user age into account to provide age appropriate recommendations.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems.</p> <p>We have also launched changes to Snapchat's in-scope services so they no longer display advertisements based on profiling for our under 18 accounts in the EU.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of</p>	<p>Yes, we have specific prevalence testing and transparency reporting for violations, including for example in relation to CSEAI.</p>



<p>their activities in particular as regards detection of systemic risk.</p>	
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with numerous trusted flaggers for child safety who are able to flag other CSEA1 or other illegal and violating activities involving Teens.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. Technology Coalition, WeProtect Global Alliance, EUIF, Alliance and CIPL to better protect minors online. We are actively participating in efforts by the Commission and other stakeholders to develop a EU wide AADC.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p> <p>All information on our Privacy and Safety Center or our Support Center is drafted for 13+. For example, Privacy By Product - Privacy Features Snapchat Privacy provides Teens with ample opportunity to understand the Snapchat features.</p> <p>We also provide Family Center as a resource to Teens and their parents or trusted adults.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, Snapchat's in-scope services have been adapted to include proactive safety by design features and content moderation for Teens. We make available robust reporting and enforcement of our terms.</p> <p>Our Family Center - Parental Control For Teens Snapchat Safety provides Teens and their parents or trusted adults a suite of resources and guidance.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and carers on risks and support.</p>
<p>Content Authenticity Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events</p>	<p>General Content Authenticity measures We display an icon in some Lenses that manipulate an image of a Snapchat to make them look younger.</p>



and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.

Conclusion

Although the prevalence of public content that may have negative effects on children on Snap's in-scope services is generally very low, we recognize that Teens are at risk of greater harm if exposed and we take the safety and wellbeing of our community, particularly its youngest members, very seriously. As such, we have assessed this risk to be in our higher risk prioritization category, Level 1, relative to other risk categories.

In response, Snap has put in place a range of mitigation measures. This includes general platform safeguards such as our Teen friendly terms and support pages, our moderation and enforcement processes, our parental tools—Family Center, in-app reporting, and Teen specific content moderation and restrictions. Plus, additional safeguards have been put in place to help Teens understand and recognize Lenses and ensure that advertisers and advertisements on our platform comply with our requirements.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on Teens and children under 18. In addition, as explained in the [Codes](#) section of this Report, we strongly support and are actively participating in cross-stakeholder efforts to develop an EU wide AADC to assess if further reasonable, proportionate and effective measures are needed for online platforms, 'gateways' (such as device operating systems, app stores and web browsers) and other online services.

4.4.4 Serious Negative Consequences on physical and mental well-being

We've made it a point to build things differently from the beginning, with a focus on helping Snapchatters communicate with their close friends in an environment that prioritizes their safety and privacy. That's why Snapchat is purposely designed differently from traditional social media. It doesn't open to a public news feed powered by an algorithm with likes and comments. Instead, as outlined earlier in this report, Snapchat opens to a camera and has five tabs: Camera, Chat, Map, Stories, and Spotlight. Additionally, conversations on Snapchat delete by default to reflect real-life conversations. Before social media, our fun, spontaneous, and silly interactions with friends only lived on in our memories. Snapchat is designed to mirror that dynamic, to help people feel comfortable expressing themselves without feeling pressure or judgment. We will discuss these risks and our mitigations in more detail below.



Likelihood

All of the risks we track on Snapchat have a low prevalence compared to the prevalence of these issues elsewhere online and offline. To aid our prioritization, our methodology seeks to assess the relative likelihood between the risks we track (even though all are low in absolute terms).

Snap assesses that serious negative consequences on physical and mental wellbeing are high in likelihood in the absence of appropriate mitigations. Without mitigations, users of digital platforms may be exposed to content affecting their mental health, contributing to body dissatisfaction and low self-esteem. They may also be exposed to content inciting physically harmful activities, such as dangerous pranks or challenges.

While Snap assesses these risks across digital platforms to be high in the absence of safeguards and mitigations, over 90% of our community says they feel happy, connected, and comfortable while using Snapchat.²³ In addition, data related to relevant policy enforcements suggests a low prevalence of content associated with harm to well-being: for example, we measured the [prevalence](#) (PVP of self-harm content (including the promotion or glorification of unhealthy behaviors) and content promoting dangerous activities to be extremely low. The likelihood of encountering this content on Snapchat falls within our extremely low likelihood category. We also monitor the queries from consumers in general and behavioral issues are not a common complaint. Accordingly, while these risks are very real, we are encouraged by data indicating that our approach to mitigating these risks is effective at reducing the likelihood of such negative impacts on physical and emotional wellbeing.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

Severity

Snap has assessed information published by governments and other third party sources and considers that if content that has a negative effect on physical and mental well-being were to materialise on an online platform, this could fall within our **'severe harm' category**.

Overall potential risk prioritization

Snap's prevalence reports and community feedback suggest that the likelihood of this risk is relatively low. If we would follow our matrix, we would qualify the overall potential of this risk as Level 3, however, given the importance of this issue, especially in relation to younger users, we have decided to deviate from the matrix, and marked the potential risk as falling within our Level 1 overall potential risk prioritization category.

²³ 2022 Alter Agents study commissioned by Snap Inc. [url](#).



Snap's Mitigations

Highlights

Snapchat is intentional about addressing risks to the physical and emotional wellbeing of users. Our Community Guidelines prohibit a range of behaviors and content that may negatively impact wellbeing, including bullying and harassment; content or Lenses that glorify unhealthy behaviors or promote unrealistic beauty standards; violent or disturbing content, or content that promotes dangerous activities; and content that promotes suicide or self-harm.

To combat this, we have put in place a number of specific protections, we will highlight some of them here. On Spotlight, we put in place protections for both creators and viewers:

- **Creator protections**

- Users choose whether to post to Spotlight and can choose to disable comments.
- If comments are not disabled, Spotlight comments are auto-moderated for abusive language before they are viewed by the creator and all comments can be reported to human moderation. This protects the creator from seeing harmful comments.
- Teens are protected on Spotlight by not having their usernames displayed.
- We limit Teens' ability to reach a large audience on Spotlight to prevent older users from seeing content from younger users. This is to protect Teens from being contacted by older users.
- We provide users the ability to post content to Spotlight anonymously.
- Creators can choose to approve comments on their Spotlight Stories prior to publication.
- We do not show view-counts on Spotlight below a certain threshold. This is to prevent focus on low view numbers.
- We aim to distribute content created by Teens to Teens. This is to prevent Teens from building a following that is not their own age.
- Creators are in control of adding hashtags / topics to their videos. This provides creators some control over how their content is categorized.

- **Viewer protections**

- Content on Spotlight does not auto-play.
- We do not have public "favorites", i.e. a user's likes and interests are not public.
- Viewers can "hide" either content or a creator. Subsequently, the user will have a lower likelihood of seeing content of such nature or content from the creator that has been "hidden".
- We survey a subset of our users quarterly to understand whether they find their time spent on our experience entertainment and satisfactory. We use this to track



whether our product changes are improving viewers' overall perception of the app.

- We provide a diversity of perspectives. We have multiple programs to foster a more diverse content community and surface different perspectives (e.g. Black Creator Accelerator program).
- We ensure there is always a large mix of content from creators from viewers' home country and content in the language in which they have set their device.
- We add diversity to every viewer's feed in terms of the account they see, and the categories of content we surface to them. This prevents users from entering an echo chamber or filter bubble of seeing the same content repeatedly. We use machine learning to understand content categories and diversify it.

We believe the above measures contribute to the well-being of both creators and viewers and creates a more pleasant experience. Similar measures are in place for Public Profiles (which are currently only available for adult accounts). For example, we allow users with access to a Public Story to turn off all Story Reply messages so they don't see messages from users who reply to their Stories. We also give users control over Story Replies and filter out words they don't want to see. Users can input words that they don't want to receive in the Story Replies from their subscribers. If a Story Reply contains an inputted word, the user does not receive the story reply (and any other story replies) from the sender. Additionally, we allow creators to block repliers or report them.

We have also undertaken considerable efforts to stay apprised of users' wellbeing. On Safer Internet Day, 6 February 2023, we launched our inaugural [Digital Well-Being Index \(DWBI\)](#), a measure of Generation Z's online psychological well-being.

To gain insight into how teens and young adults are faring online – across all platforms and devices, not just Snapchat – and to help inform our Family Center and the broader online ecosystem, we polled more than 9,000 people across three age demographics in six countries. Not surprisingly, the research showed that social media plays a major role in Gen Z's digital well-being, with more than three-quarters (78%) of respondents saying social media had a positive influence on the quality of their lives. More information about that research can be found [here](#).

Specific Mitigations

In the table below we indicate the specific measures we have taken to mitigate this risk in respect of Snapchat's in-scope services, using the defined list of mitigations set out in Article 35 of the DSA. The primary purpose of the below table is to indicate whether each specific mitigation category applies to this risk and the descriptions are illustrative rather than exhaustive. As many of our mitigations apply to all of the risks assessed in this Report, to reduce duplication in this Report, each row in the tables provides a **link in the left hand column to a full summary** of the



specific mitigation in Section 5 of this Report which explains in more detail how each mitigation operates to reduce the risk.

DSA Mitigation	Applies to this risk?
<p>Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.</p>	<p>Yes. Snapchat has incorporated a host of design, features and functions to address this risk. Starting with the aforementioned decision to open to the camera and not a news feed. This encourages self expression, communication, and exploration through our AR Lenses.</p> <p>Snapchat and third parties have created Lenses centered on movement, fitness, yoga poses, breathing activities. In addition to this, there are several partnered lenses that prompt Snapchatters to talk about wellness, mental health and their experiences.</p> <p>Our user generated content feature, Spotlight, has both creator and viewer protections in place.</p>
<p>Terms and Enforcement Adapting their terms and conditions and their enforcement.</p>	<p>Yes. Our Community Guidelines prohibit a range of behaviors and content that may negatively impact wellbeing, including bullying and harassment, content or Lenses that glorify unhealthy behaviors or promote unrealistic beauty standards, violent or disturbing content, or content that promotes dangerous activities, and content that promotes suicide or self-harm. Our Community Guidelines are enforced.</p>
<p>Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation.</p>	<p>Yes, specific proactive and reactive moderation procedures to prevent bullying and harassment, content or Lenses that glorify unhealthy behaviors or promote unrealistic beauty standards, violent or disturbing content, or content that promotes dangerous activities, and content that promotes suicide or self-harm.</p>
<p>Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.</p>	<p>Yes, our algorithmic systems do not categorize or recommend content that our Community Guidelines prohibit.</p>



	<p>Our Content Guidelines for Recommendation Eligibility - Snap Inc. further describe how sensitive and disturbing content is demoted for distribution on Spotlight and For You. For example, glorification of violence is not suggested content to users on Spotlight or For You and any discussion on self-harm, including eating disorders is demoted to users based on their age, location, or personal preferences.</p>
<p>Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.</p>	<p>Yes, other mitigations listed here also apply to our Advertising Systems. For example, ads for diet and fitness products or services must not demean the user, or shame anyone on the basis of body shape or size.</p>
<p>Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.</p>	<p>Yes, we have specific prevalence testing and transparency reporting for harassment and bullying and self-harm and suicide and other prohibited content on Snapchat that may impact users mental wellbeing.</p>
<p>Trusted Flaggers Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21.</p>	<p>Yes, we cooperate with trusted flaggers in relation to illegal hate speech and child safety.</p>
<p>Codes and Crisis Protocols Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively.</p>	<p>Yes, we cooperate with other providers through various industry groups e.g. EUIF.</p> <p>In 2017 Snap joined FSM and has signed the FSM Code of Conduct which aims to protect users from content offered on digital services that could endanger or impair their development.</p>
<p>Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.</p>	<p>Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as how to and how to get help in our Safety Center.</p>
<p>Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.</p>	<p>Yes, we have protective measures to ensure age appropriate content and our Family Center offers resources and guidance.</p> <p>Planned improvement: New parents site will provide additional guidance for parents and caregivers on protections and support. Parents.snapchat.com is expected to be</p>



	launched in early September.
<p>Content Authenticity</p> <p>Ensuring that an item of information, whether it constitutes a generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces, and, in addition, providing an easy to use functionality which enables recipients of the service to indicate such information.</p>	General content authenticity measures. We are displaying an icon in some Lenses that manipulate an image of a Snapchat to make them look younger.
Other Mitigations	All mitigations covered above.

Conclusion

Given the heightened potential for negative consequences on physical and mental well-being inherent to online platforms, specifically social media, despite the prevalence on Snapchat being low, we consider the overall potential risk prioritization to be Level 1.

In response, Snap has made deliberate design and policy decisions to reduce the potential for harm on Snapchat. Snap has implemented numerous protections for both creators and viewers of Spotlight content and undertaken considerable efforts to understand users' wellbeing on Snapchat and other platforms.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate, and effective mitigation measures for the risk of negative effects on physical and mental well-being.



5. Specific Mitigations

Article 42(4)(b) of the Digital Services Act requires providers of Very Large Online Platforms to report on the specific mitigation measures that they have put in place pursuant to Article 35(1) of the DSA. Article 35(1) of the Digital Services Act requires providers of Very Large Online Platforms to put in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified pursuant to Article 34 of the DSA, with particular consideration to the impacts of such measures on fundamental rights, including where applicable the defined categories of measures set out in Article 35(1)(a)-(k).

In Section 4 of this Report above, we reported on our: (i) assessment of the specific systemic risks applicable to Snapchat's in-scope services; (ii) summary of the mitigation measures that Snap has in place tailored to those risks and (iii) conclusion as to whether those mitigation measures are reasonable, proportionate and effective. In this Section 5, we have provided details of the specific mitigations that Snap has put in place, as are summarized in Section 4, to comply with our obligation under Article 42(4)(b).

This section 5 uses the defined categories of measures set out in Article 35(1) to detail these measures, with the following subsections:

- [Snapchat Design and Function](#) (Article 35(1)(a))
- [Terms](#) (Article 35(1)(b))
- [Transparency](#) (Article 35(1)(i))
- [Moderation](#) (Article 35(1)(c))
- [Enforcement](#) (Article 35(1)(b))
- [Algorithmic Systems](#) (Article 35(1)(d))
- [Advertising Systems](#) (Article 35(1)(e))
- [Protection of Minors](#) (Article 35(1)(j))
- [Content Authenticity](#) (Article 35(1)(k))
- [Trusted Flaggers](#) (Article 35(1)(g))
- [Dispute Settlement Bodies](#) (Article 35(1)(g))
- [Codes and Crisis Protocols](#) (Article 35(1)(h))

Note that our measures to reinforce the internal processes, resources, testing, documentation and supervision of our activities as regards to the ongoing detection and management of DSA risks, as referred to in Article 35(1)(f), is set out in [Section 6](#) of this Report.

5.1 Snapchat Design and Function

Privacy and Safety by Design from day one

From the very beginning, we've focused on helping real friends connect when they're apart and to feel comfortable expressing themselves in the moment. Snapchat is deliberately built



differently from traditional social media, in ways that make it safer for our community. We believe our privacy-and-safety-by-design approach not only empowers authentic self-expression, but recognises that privacy is deeply intertwined with safety.

Adaptations and Mitigations

As a result of our privacy and safety by design approach, Snapchat has been designed with features and functionalities that mitigate the risks described in the [Risk Assessment Results](#) section above.

As we have made key foundational design decisions from day one, we don't have A/B metrics that demonstrate the benefits of these decisions, but we fundamentally believe that those metrics are not necessary because these design decisions were made to create an antidote to traditional social media. We hear from Snapchatters about the benefits of these choices all the time, and we believe that these foundational design decisions directly influence those results.

Although not all of the features listed below are in scope of the risk assessment, we have incorporated a summary of holistic mitigations that we have put in place to demonstrate our privacy and safety by design approach below:

Friends

First, by default users need to accept bi-directional friend requests or already have each other in their contact book to start communicating with each other. This design decision adds friction and prevents users from communicating with each other prior to accepting a friend request or being in one's contact book.

Private friend lists

Second, once users have accepted friend requests, the friend lists remain private. Snapchat does not disclose the friend lists of users to other users, nor do we expose the total number of friends that a user has. This protects the privacy of the user and their friends. On most other platforms friend lists are public by default or there is an option to share them publicly. These types of features create the ability for strangers to contact vulnerable groups (e.g. younger users).

Open to the Camera not a feed

Third, Snapchat opens to the Camera and invites people to express themselves. At the surface, this may sound like a small design decision, but it directly impacts the user behavior on the platform. Instead of inviting users to scroll a feed of content, the invitation to users is to express themselves, live in the moment and share a moment with their close friends.

Stories are by default set to be viewable by friends, not the public

Fourth, once users decide to share a Snap via My Story, by default only friends can view it. Snapchatters can choose to share to everyone, only to friends, or to a customized few. This



emphasis on sharing with friends and giving users controls over who can view their content are in line with how Snap takes into account privacy and safety when designing its features.

No focus on public vanity metrics

Fifth, once a user posts to their Story, we don't show vanity metrics, such as likes on that Story content. The goal is not to create a popularity contest around who has the most friends or likes. The design choice is to provide all users with a more authentic form to express themselves.

As a result of our privacy and safety by design approach, each of Snapchat's in-scope services has been designed with features and functionalities that mitigate the risks described in the [Risk Assessment Results](#) section above.

Spotlight

Spotlight offers creators at all stages of their career a variety of opportunities and tools to help them grow their audiences, build sustainable businesses and make content creation a full-time career. The content shown in Spotlight is personalized to provide viewers with a more relevant experience. We have made following design decisions to protect our creators and users:

- Creator protections
 - Users can post to Spotlight and choose to disable comments.
 - If comments are not disabled, Spotlight comments are auto-moderated for abusive language before viewed by the creator and all comments can be reported to human moderation. This protects the creator from seeing harmful comments.
 - Adults cannot comment on Teen's Stories on Snapchat.
 - Teens are protected on Spotlight by not having their usernames displayed.
 - We restrict Teens' ability to reach a large audience on Spotlight to prevent older users from seeing content from younger users. This is to protect Teens from being contacted by older users.
 - We provide users the ability to post content to Spotlight anonymously without fear of embarrassment or if they just want to protect their privacy
 - Creators can choose to approve comments on their Spotlight Stories prior to publication
 - We do not show views on Spotlight below a certain number of views. This is to prevent pressure over low view numbers.
 - We aim to distribute content created by minors to minors. This is to prevent minors from building a following that is not their own age.
 - Creators are in control of adding hashtags / topics to their videos. This gives creators control over how their content is categorized.
- Viewer protections
 - Content on Spotlight does not auto-play



- We do not have public “favorites”, i.e. a user’s likes and interests are not public
- Viewers can “hide” either content or a creator. Subsequently, the user will not see more content of such nature or content from the creator that has been “hidden”.

For You

For You is dedicated to Creator Stories, which includes Media Partner content, Snap Originals, and some user generated content created from Snaps by popular users (“Creator Content”). The UGC that appears on For You includes the Public Stories from Snap Stars and other users who meet a follower count threshold. Similar to Spotlight, we made following design decisions to protect our creators and users:

- Creator protections
 - There is no comments section for user generated comments
 - We do not show “views” on For You Stories. This protects For You creators from feeling embarrassed or being subject to ridicule due to low number of views.
 - Content published by creators has a limited publication duration (which may be changed by creators with a Snapchat+ subscription. This protects creators by ensuring their content is not available forever.
 - Creators are free to re-publish new and saved stories at any time, provided it does not violate the law or our Terms.
- Viewer protections
 - Content on For You does not auto-play
 - We do not have public “favorites”, i.e. a user’s likes and interests are not public

Public Profile

Users with a Public Profile can post Public Stories that are publicly viewable for all Snapchatters. Additionally, Snapchatters can permanently showcase their Public Snaps, Stories, photos or videos on their profile. Snapchatters can [Subscribe](#) to Public Profiles from Spotlight, Our Story, For You, or by using the subscribe button on a Public Profile. Unlike friend requests to non-Public Profile owners, Public Profile owners will not receive a notification for new subscribers. We have made following design decisions to protect users with Public Profiles:

- Users can easily delete all of their public content. We allow users to delete all of their public content in a single tap. We delete any and all content they added to their Public Profile and that is publicly viewable. Additionally, we give users more control over their content by allowing users to post public content while keeping their identity private. Our public options [are in fact options](#). If Snapchatters are not or no longer interested in being a creator and showcasing content publicly, they can simply choose to not add to their Public Profile, post to their Public Story, or share to Spotlight and to the Snap Map with their identity attached.



- We give users control over content that is publicly viewable by allowing users to hide or show their Spotlights on their Public Profile both at the time of submission and after submission.
- Public Profile Users can turn off remixes. We allow users to decide whether their public content can be remixed by other users.
- We educate users on their public options and attribution controls. When users tap on their public profile, public story, and spotlight/snap map posting, we show them educational modals that educate them about the public option.
- To ensure that users are aware when they become friends with another user so that they can control what data that user has access to, we send notifications to the user when they become friends with another user (bi-directional add has occurred).
- Only users with a declared age of 18 or older can have a Public Profile and Public Story. Viewers cannot distinguish between users without Public Profiles (under 18) and users with Public Profiles (18+) who have not edited the Profile in any way.

Note: Since completion of this 2023 Risk Assessment, we have announced that we are reviewing a [16-17 Public Profile experience](#) with specific mitigations for this age group. This product is not yet available in the EU. It will not be rolled out to the EU until we have finalized our review and completed an update to our risk and mitigation assessments as needed.

- We give users control over their ability to be contacted. We allow users with access to a Public Story to turn off all Story Reply messages so they don't see messages from users who reply to their Stories. We also give users control over story replies and filter out words they don't want to see. Users can input words that they don't want to receive in the story replies from their fans. If a story reply contains an inputted word, the user does not receive the story reply (and any other story replies) from the sender. Additionally, we allow creators to block repliers or report them.

We have also built in protections for users who engage with a Public Profile. For example, we inform users before they send a story reply to a creator that the creator could quote the reply and make it publicly viewable (with the replier's first name and Bitmoji). We also limit unwarranted connections between younger users and bad actors on the platform.

Snap Map

Users with new accounts cannot have their Stories featured on Snap Map. Further, content posted to Public Stories will only show on Snap Map if there are multiple users posting in a short time nearby and a percentage of those posters are non-Teen accounts. Teens in Europe also don't have the ability to post to Public Stories which means that their Snaps are not eligible for the Snap Map. When younger individuals (under 16 in EEA) use Snap Map, Snap collects and uses precise location data only for the purpose of providing the feature to the Snapchatter and only



stored for a short period. These controls prevent illegitimate use of Snap Map and protects Teens from exploitation.

Snap Map has numerous additional design choices in place to make Snap Map a safer space for our community. These include:

- **Opt-in and only Friend sharing.** Given the sensitivity of geolocation data, users must grant Snap location permission via a just-in-time choice option, and even after granting that permission users must additionally opt-in to sharing their location with others on the Snap Map. Users must affirmatively opt-in to share their location with friends. Location sharing is disabled by default, and sharing preferences with friends can be easily changed by users at any time in app settings.
- **No option to share location with strangers.** Further, Snapchatters can only share their location with other users that they are already friends with on Snapchat. By default, users are not sharing their location with any friends, as all users are defaulted to “Ghost Mode”. This was to ensure that location sharing would be understood by users before activation, in particular younger users so they could make informed choices about whether to use Snap Map, whether to share their location and, if so, with whom to share it. Snapchatters can update who among their friends can see their location at any time right from the settings gear in the Map. If users decide they would like to stop for any reason, they can simply toggle Ghost Mode on and they disappear from the Map within seconds.
- **Permission and prompts.** We want location sharing on Snap Map to be limited to engagement with friends on Snapchat. We also want to ensure user safety by not broadcasting a user’s location to others who are not friends of the user. Therefore, users cannot share their location with strangers. In Settings, users can choose to share their location with their friends, or a subset of friends only. There is no option to share their location with non-friends. Friendship must be bi-directional.
- **Reminders.** We want users, especially those who don’t regularly engage with the Map/Map settings, to be regularly reminded their location is visible and to which friends.
- **Device OS displays universally recognized icon.** We want users to know at the moment whenever Snapchat is accessing their device location data. Device OS automatically displays a recognizable icon for users to know whenever an app is accessing device location data. The icon is consistent across all apps on the device that accesses location data, so it should be familiar and instantly understood by users across age groups.
- **Auto-expiration of Last Active Location.** Auto-remove users' Last Active location from the Map if they have not opened the app after 24 hours.
- **Creator protections**
 - Currently there is no comments section for user generated comments
 - We do not show “views” for Stories on Snap Map. Protects creators from feeling embarrassed or being subject to ridicule due to low number of views.
 - Content published by creators has a limited publication duration (which may be changed by creators with a Snapchat+ subscription. This protects creators by ensuring their content is not available forever).



- Creators are free to re-publish new and saved stories at any time, provided it does not violate the law or our Terms.
- **Viewer protections**
 - We do not have public “favorites”, i.e. a user’s likes and interests are not public

Lenses

Lenses (in popular language often dubbed as ‘filters’) are created by a relatively limited number of community developers, and Snap’s internal Lens Team. Our Lenses are designed with privacy-and-safety-by design principles in mind. For example, Lenses require object detection rather than facial identification. Lenses can tell what is or isn’t a face, they do not identify specific faces, limiting data processing for the use of Lenses. Snap does also not use any data collected by Lenses to customize the content that the user sees in Spotlight or For You, nor is any data collected for advertising purposes. Besides, voice data collection of Snapchatters in the EU is off by default; it is only used to provide the service.

Snap also designs every Lens with race, gender, ethnicity and cultural norms in mind. Snap leverages its ever-growing diversity training datasets, as well as feedback from community members. If a Lens does not resonate with our community, as expressed through a high ratio of user reports, we take that feedback into consideration and will re-review the Lens with a goal to leave as-is, modify, or remove.

Advertising

We have also put in place risk mitigation measures for our advertising efforts. We prevent advertisers from manipulating small audiences with microtargeted campaigns, particularly for political ads. We do so by requiring a specific minimum audience of Snapchatters to be targeted (including [Dynamic Ads on Snapchat | Snapchat for Business](#)). This prevents microtargeting that can influence voters politically or push targeted misinformation to certain populations. Our advertising systems also do not use ‘special category’ personal data to target ads and we require advertisers to provide additional information for political ads.

Integrations with other mitigations

On Snapchat we have also adapted our features to integrate with our other risk mitigations described in this Specific Mitigations section of the Report, for example:

Terms

All public content must adhere to our Terms, for example the content **must be suitable for 13+**, in order to be featured or receive broad distribution on Spotlight and For You.

Content Moderation

We moderate content on Snapchat in a number of ways to mitigate the risks of harmful and illegal



content reaching a larger audience on Snapchat.

Spotlight / For You (UGC) - Spotlight and For You rely on a combination of automated and human moderation on all submitted content before any video receives broad distribution.

For You - Only approved Creators can have their Stories distributed in For You. Those that are approved have their Stories and the 'tile' art moderated.

Reported Content (Spotlight, For You) - All reported user-generated content is reviewed by human moderators.

Content Distribution

We have put in place risk mitigation measures to restrict the distribution of harmful content on Snap. For example:

- Content that is Sexually Suggestive and Sensitive (e.g., potentially-disturbing human body imagery, violence, horror, etc.) is not recommended to Teens.
- Spotlight Comments with abusive language are removed.
- Ranking avoids 'filter bubbles' through demotion, ensuring similar content isn't sequentially recommended to Snapchatters in For You or Spotlight.
- Unlike other platforms, we do not algorithmically promote political content in Spotlight, and in For You, we only amplify content from approved creators. We take these measures in order to circumvent the spread of harmful false information.

Conclusion

From day one, Snap has made conscious design decisions to mitigate systemic risks from occurring on its platform, including privacy and safety by design decisions such as shorter retention periods, default Story visibility to just friends, not promoting likes on a user's Story, not having public friend lists, and maintaining proactive content policies. Snap has implemented additional mitigation measures as further outlined in the remainder of this Report.

As explained in Section 4, we have concluded that the adaptations made by Snap to the design, features and functioning of Snapchat's in-scope services, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective mitigation measures for the risks identified.



5.2 Terms

All Snapchatters are required to agree to Snap’s [Terms of Service](#) before they can use Snapchat. Our Terms of Services include our [Community Guidelines](#), and other policies, such as our [Advertising Policies](#), [Privacy Policy](#), [Spotlight Terms](#), and [more](#). Snap’s terms and policies are jointly referred to in this Report as “**Terms**”. In addition, in response to the Digital Services Act we have provided short summaries in each section of our [Terms of Service](#), as well as easy to read explainers of key sections of our [Community Guidelines](#). The [Annex](#) to this Report contains a copy of the explainers that were last updated in August 2023.

Each of our Terms and how they mitigate each of the DSA risk categories is explained below.

Terms of Service

Snap explicitly states the following in Section 7 of its Terms:

“(…) you may not do, attempt to do, enable, or encourage anyone else to do, any of the following and doing so may result in us terminating or suspending your access to the Services:

(…)

- *violate any applicable law or regulation in connection with your access to or use of the Service”*

This Term addresses any use of Snapchat to conduct illegal activities as well as any dissemination of illegal content as defined in the Digital Services Act (i.e., Category 1, DSA Risks).

In addition, our Terms state that “All Public Content must be appropriate for people ages 13+.”

Community Guidelines

In our Community Guidelines, which are explicitly incorporated into our Terms of Service, we provide further guidance on the categories of illegal content, and content that Snap deems in violation of its Terms.

Our Community Guidelines are broken up into the following sections: Sexual Content, Harassment & Bullying, Threats, Violence & Harm, Harmful False or Deceptive Information, Illegal or Regulated Activities, and Hateful Content, Terrorism, and Violent Extremism.



These categories have been fine tuned over many years of content moderation on Snapchat, and encompass the illegal content that we have encountered on Snapchat over the years. For ease of reference we have incorporated a more detailed breakdown of each category (as at August 2023) below:

Sexual Content

- We prohibit any activity that involves sexual exploitation or abuse of a Teen, including sharing child sexual exploitation or abuse imagery, grooming, or sexual extortion (sextortion), or the sexualization of children. We report all identified instances of child sexual exploitation to authorities, including attempts to engage in such conduct. Never post, save, send, forward, distribute, or ask for nude or sexually explicit content involving anyone under the age of 18 (this includes sending or saving such images of yourself).
- We prohibit promoting, distributing, or sharing pornographic content, as well as commercial activities that relate to pornography or sexual interactions (whether online or offline).
- Breastfeeding and other depictions of nudity in non-sexual contexts are generally permitted.
- Additional guidance on sexual conduct and content that violates our Community Guidelines is available [here](#).

These Terms make clear to Snapchatters the extent to which sexual content is prohibited. This reduces the likelihood of several risks falling with categories identified by the DSA, including in particular: (i) the dissemination of child sexual abuse material and adult sexual content in Category 1, (ii) the right to human dignity and child rights in Category 2 and (iii) negative effects on public health, Teens and gender-violence in Category 4.

Harassment & Bullying

- We prohibit bullying or harassment of any kind. This extends to all forms of sexual harassment, including sending unwanted sexually explicit, suggestive, or nude images to other users. If someone blocks you, you may not contact them from another Snapchat account.
- Sharing images of a person in a private space — like a bathroom, bedroom, locker room, or medical facility — without their knowledge and consent is prohibited, as is sharing another person’s private information without their knowledge and consent or for the purpose of harassment (i.e., “doxxing”).
- If someone is depicted in your Snap and asks you to remove it, please do! Respect the privacy rights of others.



- Please also do not harass another Snapchatter by abusing our reporting mechanisms, such as intentionally reporting content that is permissible.
- Additional guidance on how bullying and harassment violate our Community Guidelines is available [here](#).

These Terms make clear to Snapchatters that the extent to which harassment and bullying is prohibited. This reduces the likelihood of several risks falling with categories identified by the DSA, including in particular: (i) the right to human dignity, private life and data protection, and child rights, in Category 2 and (ii) negative effects on public health, Teens, and gender-violence, as well as serious negative consequences to a person's physical and mental well being in Category 4.

Threats, Violence & Harm

- Encouraging or engaging in violent or dangerous behavior is prohibited. Never intimidate or threaten to harm a person, a group of people, or someone's property.
- Snaps of gratuitous or graphic violence, including animal abuse, are not allowed.
- We don't allow the glorification of self-harm, including the promotion of self-injury, suicide, or eating disorders.
- Additional guidance on threats, violence, and harm that violate our Community Guidelines is available [here](#).

These Terms make clear to Snapchatters the extent to which threats and violence are prohibited. This reduces the likelihood of several risks falling with categories identified by the DSA, including in particular: (i) the right to human dignity and property, and child rights, in Category 2, (ii) negative effects on civic discourse and public security in Category 3 and (iii) negative effects on public health, Teens, and gender-violence, as well as serious negative consequences to a person's physical and mental well being in Category 4.

Harmful False or Deceptive Information

- We prohibit spreading false information that causes harm or is malicious, such as denying the existence of tragic events, unsubstantiated medical claims, undermining the integrity of civic processes, or manipulating content for false or misleading purposes (whether through generative AI or through deceptive editing).
- We prohibit pretending to be someone (or something) that you're not, or attempting to deceive people about who you are. This includes impersonating your friends, celebrities, public figures, brands, or other people or organizations for harmful, non-satirical purposes.



- We prohibit spam, including pay-for-follower promotions or other follower-growth schemes, the promotion of spam applications, or the promotion of multilevel marketing or pyramid schemes.
- We prohibit fraud and other deceptive practices, including the promotion of fraudulent goods or services or get-rich-quick schemes, or imitating Snapchat or Snap Inc.
- Additional guidance on harmful false or deceptive content that violates our Community Guidelines is available [here](#).

These Terms make clear the extent to which harmful false or deceptive information is prohibited. This reduces the likelihood of several risks falling with categories identified by the DSA, including in particular: (i) the dissemination of harmful false information, fraud and spam in Category 1, (ii) the right to human dignity, private life and data protection, and child rights in Category 2, (iii) negative effects on democratic and electoral processes, civic discourse and public security in Category 3 and (iv) negative effects on public health, Teens, and gender-violence, as well as serious negative consequences to a person's physical and mental well being in Category 4.

Illegal or Regulated Activities

- Don't use Snapchat to send or post content that's illegal in your jurisdiction, or for any illegal activity. This includes promoting, facilitating, or participating in criminal activity, such as buying, selling, exchanging, or facilitating sales of illegal or regulated drugs, contraband (such as child sexual exploitation or abuse imagery), weapons, or counterfeit goods or documents. It also includes promoting or facilitating any form of exploitation, including sex trafficking, labor trafficking, or other human trafficking.
- We prohibit the illegal promotion of regulated goods or industries, including unauthorized promotion of gambling, tobacco or vape products, and alcohol.
- Additional guidance on prohibited illegal or regulated activities that violate our Community Guidelines is available [here](#).

These Terms make clear the extent to which illegal or regulated activities are prohibited. This reduces the likelihood of several risks falling with categories identified by the DSA, including in particular: (i) the dissemination of illegal content, including child sexual abuse material and other types of misuse of Snapchat for criminal offences, and the conduct of illegal activities, such as the sale of products or services prohibited by European Union or Member State law, including dangerous or counterfeit products, or illegally-traded animals in Category 1, (ii) the right to property and child rights in Category 2, (iii) negative effects on public security in Category 3 and (iv) negative effects on public health and Teens, as well as serious negative consequences to a person's physical and mental well being in Category 4.



Hateful Content, Terrorism, and Violent Extremism

- Terrorist organizations, violent extremists, and hate groups are prohibited from using our platform. We have no tolerance for content that advocates or advances terrorism or violent extremism.
- Hate speech or content that demeans, defames, or promotes discrimination or violence on the basis of race, color, caste, ethnicity, national origin, religion, sexual orientation, gender, gender identity, disability, or veteran status, immigration status, socio-economic status, age, weight, or pregnancy status is prohibited.
- Additional guidance on hateful content, terrorism, and violent extremism that violates our Community Guidelines is available [here](#).

These Terms make clear the extent to which hate speech and terrorism are prohibited. This reduces the likelihood of several risks falling with categories identified by the DSA, including in particular: (i) the dissemination of illegal hate speech and other types of misuse of Snapchat for criminal offenses and the conduct of illegal activities in Category 1, (ii) the right to human dignity, non-discrimination and child rights in Category 2, (iii) negative effects on public security in Category 3 and (iv) negative effects on Teens, as well as serious negative consequences to a person's physical and mental well being in Category 4.

We understand that each of the above categories can be nuanced and open to interpretation, that is why we have included explainers for each category (see the [Annex](#)).

Platform Specific Terms

In addition to the Terms described in detail above, we also have specific, publicly-available terms and policies that govern the use of additional aspects of Snapchat's features:

- **Spotlight:** Snapchat users who choose to contribute content to Spotlight agree to the [Snap Spotlight Submission and Revenue Terms](#), which are made available to all users prior to submitting a video to Spotlight and were last updated in August 2023. Snap also provides users who submit content to Spotlight with clear [Spotlight Guidelines](#), describing the policy, technical, and legal requirements for submissions to Spotlight, as well as reminding users of the Terms (including our [Community Guidelines](#)).
- **For You:** We have specific publishing agreements with our premium partners that post content on For You, such as media organizations and Snap Stars, that require them to abide by our Terms (including our [Community Guidelines](#) and [Content Guidelines for Recommendation Eligibility](#)).
- **Lenses:** Snapchat users who choose to develop and submit Lenses for publication on Snapchat via Lens Studio must agree to the [Lens Studio Terms](#). Lenses must comply with our [Lens Studio Submission Guidelines](#), which also remind users of the Terms (including our [Community Guidelines](#)).



- **Ads:** Snapchat users who choose to advertise to other users on Snapchat must agree to our [Snap Advertising Policies](#), including an obligation for advertisements to comply with applicable laws and rules in the European Union and each Member State where the advertisements will run.

Languages

Our [Terms of Service](#) have been translated into all official languages of the European Union as explicitly required by the Digital Services Act. We are in the process of updating our other terms to ensure they are available in official languages of the European Union.

Conclusion

Snap's Terms, including its Terms of Service, Community Guidelines, and platform specific terms are designed to mitigate DSA-related risks in a meaningful way. All users, including Media Partners, must agree and adhere to our Terms. These Terms, among other things, prohibit users from publishing illegal or harmful content, and provide a clear basis to moderate and enforce against individuals who violate them.

As explained in Section 4, we have concluded that Snap's Terms for Snapchat's in-scope services, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective measures to mitigate the risks identified.

5.3 Transparency

Snap is focussed on providing users with the right level of information, at the right time. We understand that our community does not always have time to read multi page documents. This is why we strive to provide users with bite-sized information that is easy to access and understand, while also giving them an opportunity to review more detailed information where appropriate.

Information provided to users can be divided into three categories:

1. Information we provide on our website;
2. Information provided in app stores; and
3. Information we provide in our application.

1. Information we provide on our website

At Snap we have a number of avenues to provide information to users. The two primary sources of information outside of our application are our [Privacy and Safety Hub](#) and our [Support Center](#).



Privacy and Safety Hub

Snap's Privacy and Safety Hub was launched last year and combines our Privacy Center, Safety Center and Transparency Center all under one umbrella. The rationale behind this change is that we believe there is a natural overlap between these areas, and that all the information provided in those domains contribute to providing awareness and building trust with our community and other stakeholders, such as parents, teachers, journalists, trusted flaggers, law enforcement, regulators, and NGOs.

The top navigation provides Privacy, Safety, and Transparency resources, as well as our latest News in those areas. In this section, we highlight a number of areas for illustration purposes, and refer to the [website](#) for further information.

Privacy Center

Our Privacy Policy provides a detailed description of our privacy practices, but we recognize that policies and terms can be overwhelming documents. That is why our long standing approach has been to provide additional, bite-sized information on our general practices (Privacy Center), our philosophy to privacy (Privacy Principles), what we do with user data (How we use your information), advertising (Snap and Ads), and specific products (Privacy by Product). In the introduction to our Privacy Policy, we state the following:

"We've done our best to write this Privacy Policy in a way that's easy to understand for all our users and free of difficult language and legal phrases. If you want to review something later on, you can always take a look at our [Privacy Center](#). We designed it to give you easy-to-digest summaries of our privacy practices. For example, our [Privacy by Product](#) page gives product-specific information and links to support pages with tips and tricks. Still have questions? Just [reach out](#) to us."

Privacy by Product gives users concise and easily understandable information about our products. For example, this webpage provides an overview of our approach to Snaps & Chats, as well as hyperlinks to more detailed information on specific aspects. Similarly, there's a section on Spotlight, Lenses, My AI, Stories, and many more products.

Safety Center

From the Privacy Center, users can easily navigate to the Safety Center, which provides an overview of our Safety resources, including tips on how to report content, the acknowledgment that safety is a shared responsibility, as well information on our Trusted Flagger Program, Safety Advisory Board, [Digital Well-Being Index](#) and more. Again, the goal here is to provide easy to navigate and process information.

Our community, trusted flaggers, and other stakeholders play a vital role in the safety of our platform. A primary way they do this is by reporting content. That's why we think it's crucial to raise as much awareness as possible about reporting. We have dedicated a page on our Safety








Center to reporting, which summarizes the various ways users can report content, and provides additional resources on how to report (e.g., a hyperlink to our [Safety Snapshot](#) episode on reporting). The page also links to our [Reporting Quick Guide](#) and contains a hyperlink to our web [reporting form](#).

Another important component of the Safety Center is our Safety Resources and Support page. The goal of this page is to provide users with additional resources, such as a hyperlink to [MindUp](#), information about our [Here For You](#) tool, and country specific information.

MindUp is a non-profit organization that supports children ages 3 to 14 by providing them with the tools and knowledge to manage stress and thrive in school all while maintaining optimism, resilience, and compassion.

Our Here for You search tool, which is accessible within the Snapchat app, shows resources from expert localized partners when users search for certain topics related to mental health, anxiety, depression, stress, suicidal thoughts, grief and bullying.

Our country-specific resources provide users with additional information about resources that are available to them in their country, such as children's helplines, suicide prevention hotlines, and more. See for example the below, for France:

Estonia (EE) 	∨
Finland (FI) 	∨
France (FR) 	∧
<u>E-Enfance</u>	
Call 3018	
The new national number against digital violence, free for children and adolescents facing problems related to their digital use-- 100% anonymous free and confidential.	
<u>Suicide Écoute</u>	
Call 01 45 39 40 00	
Suicide Ecoute helps those who are thinking about ending their lives or have decided to do so. Suicide Ecoute allows everyone, in complete anonymity, to express their suffering.	
<u>SOS Suicide Phénix</u>	
Call 01 40 44 46 45	
The SOS Suicide Phoenix France Federation aims to PREVENTION of suicide and PROMOTION of preventive actions in complementarity with the actors of the medico-social field.	
Germany (DE) 	∨
Greece (GR) 	∨



We recognize that not all caregivers, parents and teachers use Snapchat. Their lack of familiarity may create questions, and may also make it difficult for them to have a conversation with younger users. To address this concern, we have raised awareness with this specific group in a number of ways. First, we launched [Family Center](#), our in-app tool for parents and caregivers. To help develop Family Center, we worked with families to understand the needs of both parents and teens, knowing that everyone's approach to parenting and privacy is different. We also consulted with experts in online safety and wellbeing to incorporate their feedback and insights. Our goal was to create a set of tools designed to reflect the dynamics of real-world relationships and foster collaboration and trust between parents and teens. In the coming weeks, we will add a new feature that will allow parents to easily view new friends their teens have added.

In addition to Family Center, we also created a [Parent's Guide to Snapchat](#). The Parent's Guide helps parents navigate the Snapchat app, outlines Snapchat's and provides parents with additional information that empowers them and their family to safely express themselves, live in the moment, learn about the world, and have fun together.

Early September of this year, we will be launching a microsite: parents.snapchat.com, which will provide even more information for parents, and replace the above resources.

Transparency Center

Our Transparency Center provides additional transparency resources to our users and to the public at large, including our Community Guidelines (see Terms section), Transparency Reports and EU-specific information required under the DSA.

On our [EU](#) transparency page, we publish EU-specific information required under the DSA, including the number of Average Monthly Active Recipients of our Snapchat app in the EU, and information about our legal representative in the EU, how EU law enforcement agencies can submit requests to snapchat, and the regulatory authorities that regulate us under the DSA.

Since 2015, we have also been publishing Transparency Reports twice a year, to provide insight into Snap's safety efforts and the nature and volume of content reported on our platform. We are committed to continuing to make these reports more comprehensive and informative to the many stakeholders who care deeply about our content moderation and law enforcement practices, as well as the well-being of our community. As part of our DSA compliance, Snap will be adding new metrics and information to its Transparency Report. Copies of our most recent and previous Transparency reports can be found on our [Transparency Report](#) and [Previous Reports](#) webpages.

News Page

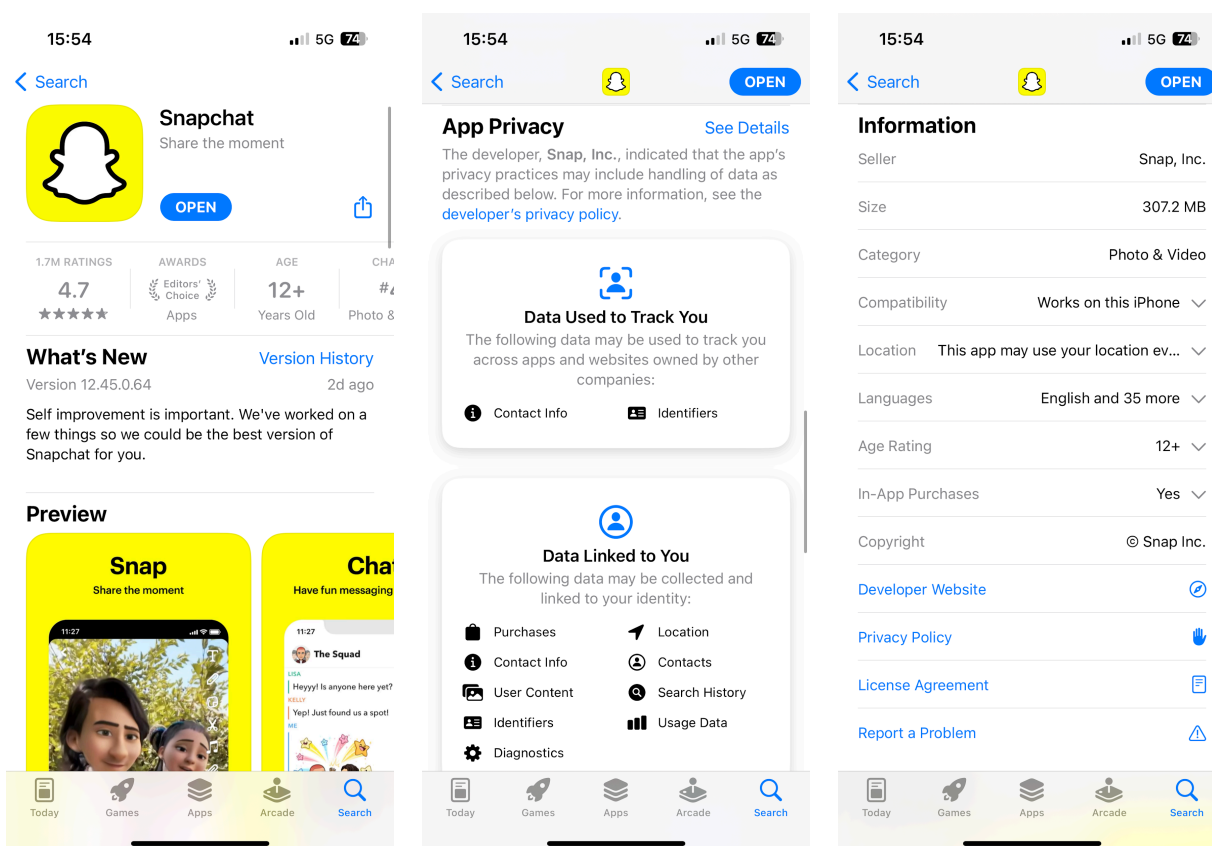
Snap also frequently publishes Privacy and Safety related information on the Hub's [News](#) webpage. The purpose of these news articles is to inform the general public about recent developments on issues relevant to privacy, safety and transparency on Snapchat. For example, a



recent article focused on the addition of AI experts to [Snap's Safety Advisory Board](#), another one announced our H2 2022 [Transparency Report](#), and earlier this year we published our first annual [Digital Well-Being Index](#).

2. Information provided in app stores

Prior to downloading Snapchat, we provide users with information about the Snapchat app in the Apple and Google Play Stores. This includes general information on the functionalities of the app, as well as information on our data collection practices, and links to our website, Privacy Policy and Terms. This way users are able to get a better understanding of the application ahead of using Snapchat.



3. Information we provide in our application

Once users have downloaded Snapchat, they are required to create an account before they can start using the application. At Snap, our philosophy is to provide timely notifications and generate awareness at points in time where we believe they will be most effective. We provide a high level



overview of our onboarding process and highlight examples of our “just-in-time” in-app notifications in this section.

Onboarding process

Step 1.

When users open Snapchat, they are invited to log in (if they have an existing account) or create a new account. The first set of notices users receive relates to notification settings, and the ability to connect their device’s contacts to find friends. Both are optional and are **out of scope of this Report**. However, for completeness, we note that the reason we prompt users to turn on notifications is that Snap is primarily a messaging service and notifications provide an essential utility when using the service. Snap is intended for real friends and family, and requires users to accept friend requests or be already existing contacts before they can start communicating with each other. Typically, users already have their close friends and family stored in their device contact book, so the “Find Your Friends” prompt is intended to make it easier for users to send friend requests to other users and to communicate with one another.

Step 2.

The second step of the onboarding flow requests basic account information such as the user’s first name, last name (optional), birthday and username. When asking for their birthday, we show users a neutral age screen, and if a user selects an age under 13, they are prevented from creating an account. We don’t notify the users the reason for a failure to create an account.

We have drawn on guidance from the UN Convention on the Rights of the Child²⁴ and UK Age-Appropriate Design Code²⁵ to adopt a risk-based approach to age verification in our age gating process. We considered the risks of the platform as well as the rights of younger user’s right to privacy, freedom to access information and freedom of expression under the Convention and balanced them against safety risks. We believe more invasive age gates come at a privacy cost for all users, and also disproportionately impact marginalized groups who may not have access to government IDs.²⁶ We have supported the UK Online Safety Bill amendment to require App Stores to play a more active role in sharing age signals to all app stores. We believe this is the better upstream solution to address any systemic risks associated with underage users accessing platforms.

If a user has inputted an age of 13 or older, they are prompted to provide a username. We check usernames against our Abusive Language Detection (ALD) models. If users type in an abusive username (i.e., one that does not comport with our Terms), they are prevented from creating an account and are asked to enter a username that adheres to our Terms.

²⁴ UN OHCHR, *Convention on the Rights of the Child*, [url](#).

²⁵ UK Information Commissioner’s Office, Introduction to the Children’s code, [url](#).

²⁶ See for example the report on age verification issued by the Australian eSafety Commissioner, [url](#).

**Step 3.**

The third step of the onboarding process is focussed on password creation and providing a phone number and / or an email address. These are standard steps to improve account security and provide Snap the ability to communicate with users.

Step 4.

Lastly, we offer users the ability to start finding friends on Snapchat, and the option to create a Bitmoji. Snapchat shows Bitmojis instead of profile pictures. Bitmojis protect the identity of users, and prevent abuse from predators who may use profile pictures as signals to reach out to their target victims.

Just-in-time notifications

Once a user has created an account, we create awareness at a feature-specific level, typically using Just-in-Time notices or “JITs”. We conduct user research and sentiment studies, and feedback we receive from users is that JITs or icons are more effective to inform users than long text. For example:

- **Snap Map**

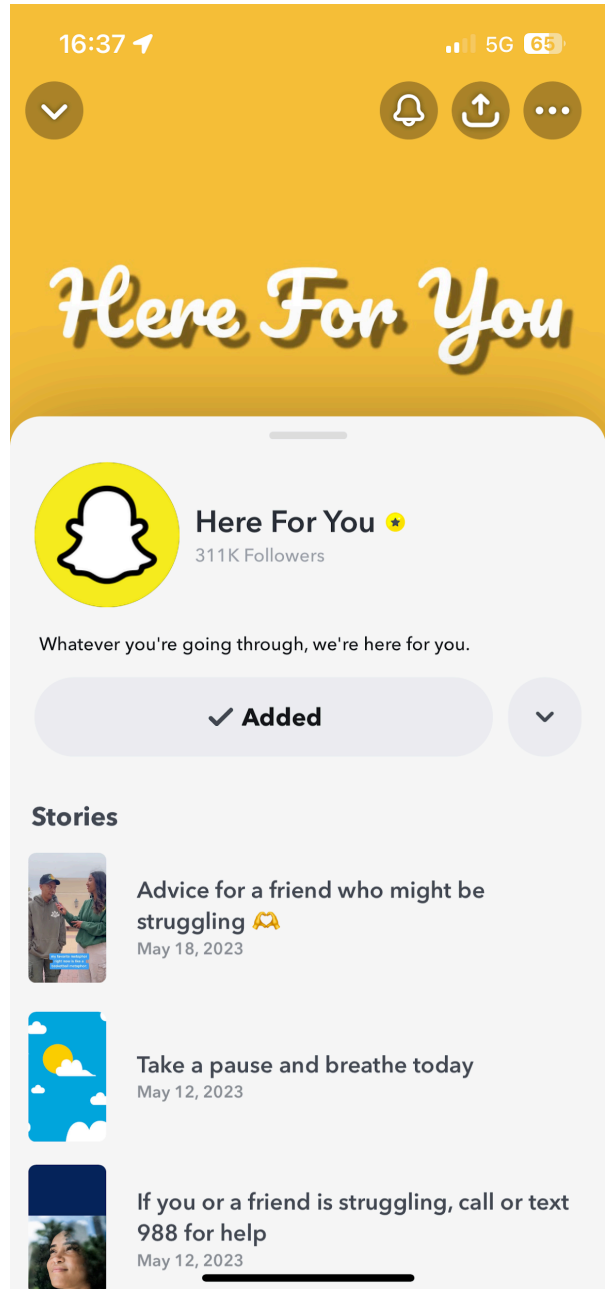
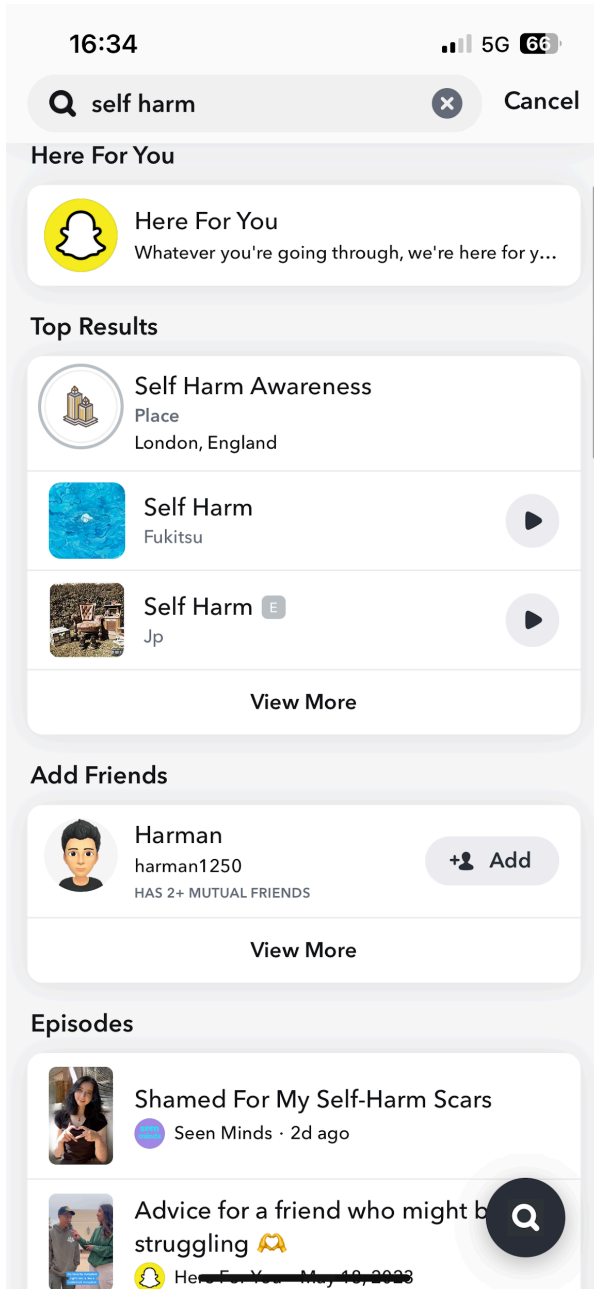
Snap Map is off by default and user location data is off by default. Users choose who can see their location. Users can choose to share their last active location while using the app with [all their friends they have added back, or just a group of select friends](#) and users may also decide to turn on [Ghost Mode](#) when they want to go off the grid.

- **Spotlight**

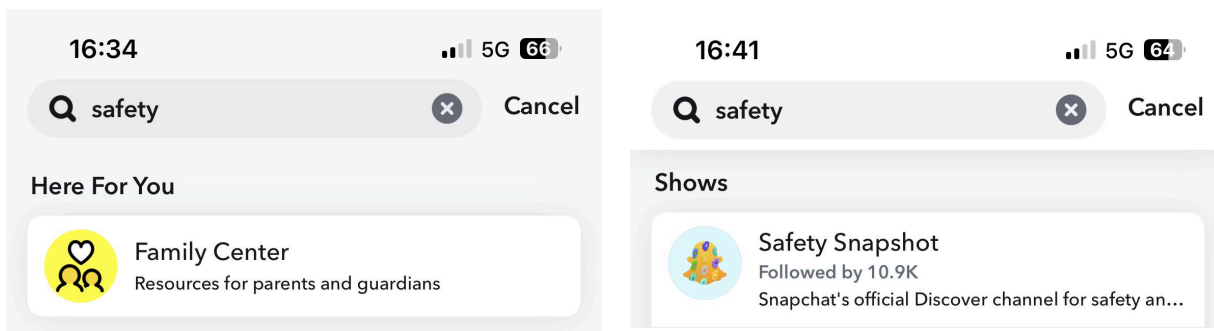
Before a user submits a Snap to Spotlight they are presented with a JIT informing them that Spotlight submissions are public. This is to create awareness that Spotlight is different from My Story submissions, which can be shared with friends only, unless the user actively chooses to share them with “Everyone”.

Thematic awareness and notices

Across Snapchat, we offer a number of resources to users to raise awareness on safety topics and protect them. For example, if a user types in “self-harm” or related terms in our Search functionality, we will show them relevant ‘Here For You’ resources as the first search result.



Similarly, search terms like “safety” will direct users to our relevant Here For You resources, such as information on our Family Center, and to our Safety Snapshots, our official channel for safety and privacy tips and tricks.



We also run campaigns on Snapchat to raise awareness about certain themes. For example, on [Global Data Privacy Day](#), we show users privacy and security information and tools, including interactive Lenses with tips on how to stay safe online.



Similarly, on [Safer Internet Day](#) 2022, we raised awareness around in-app reporting, and in 2023 we launched our Digital Well-Being Index. We also ran [campaigns](#) to raise awareness around the dangers of fentanyl, and continue to partner with organizations like [Song For Charlie](#) to combat illicit drugs on Snapchat.

Languages

As explained above, our [Terms of Service](#) have been translated into all official languages of the European Union as explicitly required by the Digital Services Act. However, Snapchat itself is only available in certain official languages of the European Union and not all. As a result, our in-app and publicly accessible information is also only available in certain official languages of the European Union. We consider it reasonable and proportionate and effective to offer our



mitigation measures in the same languages as Snapchat as we anticipate recipients only using Snapchat if they understand one of the available languages.

Conclusion

Snap offers a wide range of in-app and publicly accessible information to raise awareness around privacy, safety and security to its community and external stakeholders. Our approach is that these tools should be easily accessible, easy to use and understand, and provided in a timely manner. We believe that the awareness measures we have in place provide reasonable, proportionate and effective mitigations.

As explained in Section 4, we have concluded that Snap’s awareness raising information, in combination with the other mitigations explained in this Section 5, is a reasonable, proportionate and effective mitigation measure for the risks presented by Snapchat’s in-scope services.

5.4 Moderation

Overview

Across Snapchat, we’re committed to advancing safety while respecting the privacy interests of our community. We take a balanced, risk-based approach to combating harms — combining transparent content moderation practices, consistent and equitable enforcement, and clear communication to hold ourselves accountable for applying our policies fairly.

In summary, we’ve designed Snapchat with safety in mind, and this design is key in helping to prevent the spread of harmful and illegal content. Snapchat does not offer an open news feed where unvetted publishers or individuals have an opportunity to broadcast hate, misinformation, or violent content. We think about content on our platform in two categories:

1. “Broadcast content” is recommended for broad distribution on Snapchat. Broadcast content includes Spotlight, “For You” content on the Stories tab, Lenses, and Advertisements.
2. “Private content” is distributed to friends/followers. Private content includes Private Stories, Chat, Groups, Accounts.

All content everywhere on Snapchat must adhere to our [Community Guidelines](#) and [Terms of Service](#). Then, in order to be eligible for algorithmic recommendation beyond the creator’s friends or followers, content must meet the additional, higher standards described in our [Content Guidelines for Recommendation Eligibility](#).

We provide in-app tools for Snapchatters to report content that they find objectionable. We respond to user reports quickly, and we use feedback to improve the content experience for all Snapchatters.



We use a combination of automated tools and human review to moderate content. Proactive detection mechanisms or in-app reports may trigger a review, at which point, our tooling systems process the request, gather relevant metadata, and route the relevant content to the moderation team via a structured user interface that is designed to facilitate effective and efficient review operations. Moderators are trained on Snap's guidelines, relevant processes, and tooling. For more public information on our moderation practice, see our [transparency reports](#).

Conclusion

Safety is a priority across Snapchat, and we use a combination of in-app reporting, automation tools, and human review to combat harms on the platform. All content must adhere to our [Terms](#), including our [Community Guidelines](#) and [Terms of Service](#), and some content must also adhere to our [Content Guidelines for Recommendation Eligibility](#). We strive to be transparent and consistent in our practices and enforcement, while striking the right balance between privacy and safety.

As explained in Section 4, we have concluded that Snap's measures to moderate illegal or violating content, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective mitigation measures for the risks presented by Snapchat's in-scope services.

5.5 Enforcement

Introduction

As explained in the [Terms](#) part of this Report, Snap has carefully developed its Terms with a view to mitigating the systemic risks it has identified for the EU (see [Section 2 of this Report - Risk Assessment Results](#)). Integral to our risk mitigation efforts are Snap's policies and processes to enforce these Terms. Below, we explain how we enforce our Terms in a transparent, consistent and equitable manner, balancing our commitment to safety with respect for the privacy interests of our community.

In summary, we detect violations of our Terms through both proactive and reactive moderation. Our proactive moderation relies on technological tools (e.g., machine learning) as well as human review. Our reactive moderation processes are triggered when we receive a report of an alleged violation on Snapchat. Our reporting systems provide users and non-users in the EU the ability to easily report Snapchat accounts and content they believe violate our Terms. We review all flagged accounts and content against our Terms. When we determine that a user has violated our Terms, we may remove the offending content, terminate or limit the visibility of the relevant account, and/or notify law enforcement. Our policies and systems promote consistent and fair enforcement, and provide Snapchatters an opportunity to meaningfully dispute enforcement outcomes through Notice and Appeals processes that safeguard the interests of our community



while protecting Snapchatters' rights. We continually strive to improve our enforcement policies and processes and have made great strides in combating harmful and illegal content and activities on Snapchat. This is reflected in an upward trend in our reporting and enforcement figures and decreasing prevalence rates for violations on Snapchat overall. More public information can be found in our [Community Guidelines and related explainers](#).

Conclusion

Increases in reporting, enforcement and proactive law enforcement referrals over time do not mean that Snapchat has become less safe. On the contrary, these upward trends correlate with a drop in Policy Violating Prevalence (PVP) on Snapchat overall. In other words, as we get better at detecting and enforcing against an increased number of violations, the frequency of violations found on Snapchat decreases overall. Because we have prioritized rooting out the most Severe Harms, the drop in PVP rates on Snapchat for Severe Harm has been particularly steep. In the 90-day period between May 12, 2023 and August 9, 2023, Snaps labeled for Severe Harm categories represented only an extremely low percentage of PVP of the sampled Stories - our lowest likelihood category.

Note: This section relies on 2023 figures which are out of date at the date of publication of this 2023 Report. While they already show an extremely low prevalence on Snapchat, our 2024 Report shows a substantial further reduction.

We are committed to continuously improving the safety of our communities on Snapchat and beyond, and use prevalence testing to identify and adapt to changing abuse trends on Snapchat, so we are best equipped to detect and address any gaps in enforcement. More information on our risk detection and management can be found in [Section 6](#) of this Report.

As explained in Section 4, we have concluded that Snap's measures to enforce its Terms, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective mitigation measures for the risks identified for Snapchat's in-scope services.

5.6 Algorithmic Systems

Introduction

Snap provides a free personalized content experience that is intended to entertain and delight users in the same app they use to communicate with their friends and family. The content recommendation systems in Spotlight and For You are among Snap's most significant algorithmic systems. Users find new content on For You and Spotlight primarily through our algorithmic personalization/recommendation service. However, algorithmic recommendation systems in



general might give rise to, amplify and/or result in the rapid and wide dissemination of illegal content and/or other harms identified in [Section 4](#), if not adapted and tested appropriately. Without mitigations, these algorithmic systems have the potential risk of giving rise to the concerns referenced in Article 34 of the Digital Services Act.

How do our Recommender Systems work?

To help users discover content they will be interested in, Snap's content recommender systems seek to understand the types of content viewers are interested in and not interested in. See <https://help.snapchat.com/hc/en-gb/articles/17338132910484-Personalisation-on-Snapchat> for information on how our Recommender Systems work.

Benefits

Snap's recommender systems allow users to more easily discover interesting, entertaining, and relevant content. With over a million submissions a day of content, discovery methods like sorting by popularity, alphanumeric, timestamp, or curation are not practical.

Our recommender systems help viewers discover new interests they otherwise would have never found, and help creators who otherwise would not have been able to find an audience, allow users to learn, develop, play and have fun online. Users can explore different experiences, learn about topics of interest, and see what is happening around the world. Recommender systems are dynamic and responsive in that they can respond to viewers feedback.

We know users consider personalized recommender systems to provide significant benefit because:

- Viewers tell us (through their actions) that they prefer recommendations over other approaches and access to entertaining content is one of users' most frequent requests; and
- When we have tested removing personalization on Snapchat, we see a significant fall in user engagement (view time).

We also note that one of the reasons that traditional media services (i.e. linear television, newspapers, and magazines) are perceived to be in decline is because they are less entertaining to a diverse audience than the personalized alternatives provided by online platforms, such as Snapchat's in-scope services.

Adaption and Testing

In line with Article 35(1)(d), we explain in this part of the Report the extent to which we have adapted and tested our algorithmic recommendation systems to help address the risks identified in [Section 4](#). Snap has extensively adapted its algorithmic recommendation systems to ensure



our content experience is beneficial to users, and that the risks of algorithmic personalization are mitigated. Considering each risk and its mitigation(s) in turn:

Illegal or violating content

As explained in this [Terms](#) section of this Report, all content on Snap must comply with our Terms which requires all public content on Snapchat to be suitable for users as young as 13, including our Community Guidelines. Additionally, content personalized by our algorithmic recommendation system must also comply with our more restrictive Content Guidelines for Recommendation Eligibility.

As explained in the [Moderation](#) section of this Report, we have adapted our recommender systems and its processes to enforce our content policies with robust automated and human moderation. Our restrictive Terms and robust moderation help Snapchat mitigate the risk that illegal, false, or inappropriate content will be available to be promoted by our recommendation algorithms.

As explained in the [Enforcement](#) section of this Report, users may also easily report inappropriate and illegal content. Each piece of content in Spotlight and For You has a menu that allows users to report content. All reported user-generated content in Spotlight, For You and Ads is reviewed by human moderators. If the content violates our policies and somehow made it through our automated and human reviews, it is made ineligible for future recommendations by our algorithmic systems.

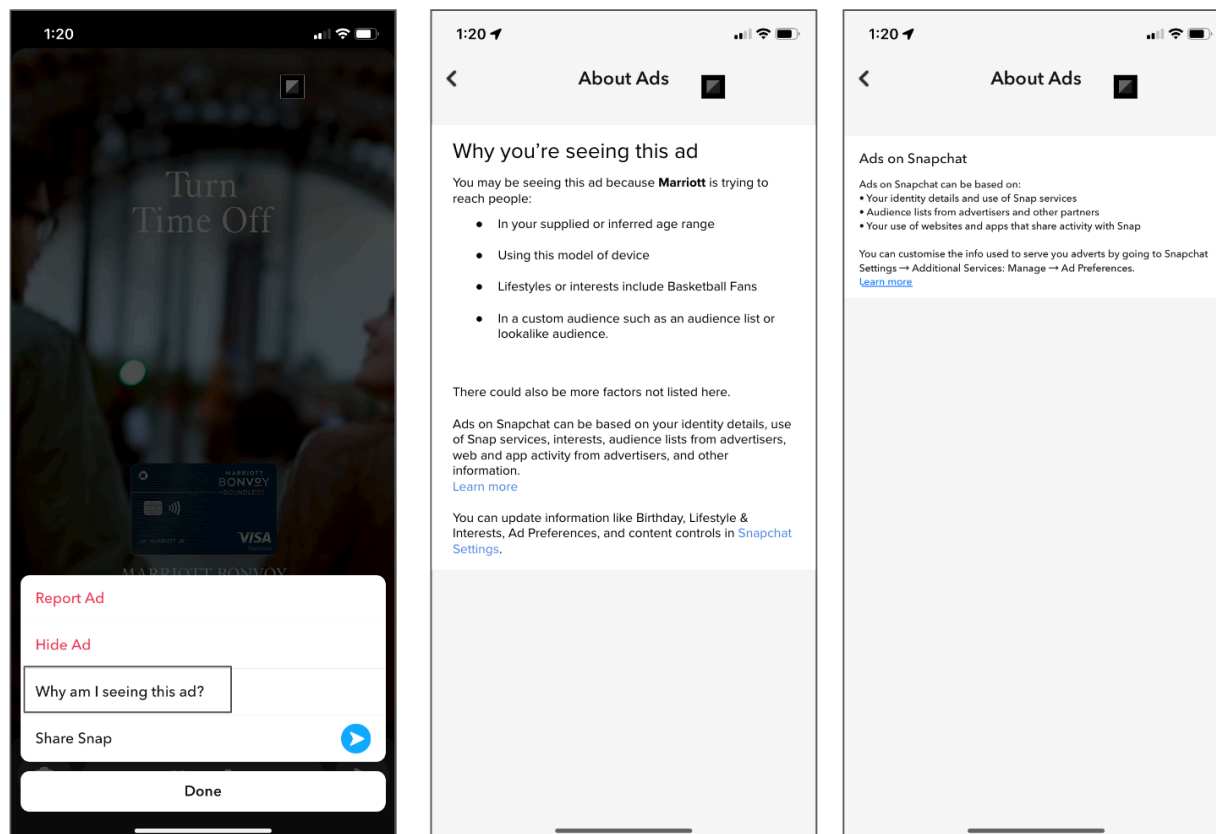
The effectiveness of these measures is tested through prevalence testing and by reviewing privacy and other consumer queries raised to our community support teams, our Data Protection Officer and our DSA Compliance Team.

Lack of user understanding

Our recommender systems are complex and the process, the signals used in ranking and how significant each signal is to the recommender system can be challenging for users to understand.

To help users and answer frequently asked questions, and as part of our DSA compliance, we have:

1. Adapted our content to include links to articles available explaining how we personalize content in Spotlight, For You and Ads [here](#). This includes a description of the main parameters used for our recommender systems, as well as the weighting applied to each signal.



- Users may also reach out to our Support team if they have concerns or questions about how our algorithms work. We test this is appropriate by reviewing privacy and other consumer queries raised to our community support teams and our Data Protection Officer.

Users may consider personalized recommendations based on inferred interests to be intrusive

We believe content is more relevant and entertaining when it's personalized to a user's interests, and not to someone else's. However, there is a risk that some users may experience personalized recommendations based on their inferred interest to be intrusive.

In For You and Spotlight users can disable personalized content by either tapping on '...' then 'Why am I seeing this content?' which will take the user to Settings or the user can navigate directly to Settings and 'European Union Controls'. When the user disables personalization, the For You and Spotlight experiences will be less personalized, and rely on essentials to determine what content to show the user, such as the language the user has set on their phone, their age, and country. Users will still see content, but it will be more random and less relevant to the user's interests (as required under Article 38 DSA). If the user wishes to enable personalization again, users can do so either by tapping on the favorite icon (❤️) in For You and Spotlight and then tapping 'Enable' in the 'Show More Personalized Content?' screen or by going to Settings in 'European Union Controls'.



Special / Sensitive Categories of Personal Data & Systems could be biased in a way that leads to discrimination

Algorithms that process special categories of personal data (as defined in GDPR) on a large scale are considered high risk and require explicit user consent. We have adapted Snapchat's recommender systems so they do not track or identify special categories of personal data, including for the purpose of recommending content and ads.

Rapid and Widespread illegal or false content & Exposure to crisis situations and unexpected events

There is a risk of rapid and widespread illegal or false content on Spotlight and For You, as well as exposure to crisis situations and unexpected events like riots. We combat this risk by prohibiting illegal or false content in our [Terms of Service](#) and [Community Guidelines](#) and allowing users to report violations. More importantly, Spotlight relies on a combination of automated and human [moderation](#) on all submitted content before any video receives broad distribution. On For You, only approved Creators can have their Stories distributed in For You. Those that are approved have their Stories and the 'tile' art moderated. We also monitor reporting and hide rates on both For You and Spotlight.

Filter bubbles

Our recommender system algorithms are designed to serve users with content that they will find engaging based on factors that include which categories of content they have previously watched. There is a risk therefore that, without safeguards, the algorithm will tag users who view content that may not be harmful on its own as being interested in that content and that repeated and frequent exposure to that content could be harmful. For example, while one piece of content related to dieting may not be harmful, if a user sees many or frequent videos about dieting, the user may feel inappropriately pressured to diet or may get a skewed perspective on how people manage their relationship with food.

We address this risk in a few ways. Firstly, we take significant steps to prevent and remove content that may become harmful when viewed frequently on Spotlight or For You, including as explained above and in the [Terms](#), [Moderation](#) and [Enforcement](#) sections of this Report. Secondly, our content categories do not include harmful content categories and so in the unlikely event that a user does view harmful content, this will not be used by our recommender system algorithm to recommend similar content. Thirdly, in our For You and Spotlight content recommendation systems, we have rules in place to ensure that a particular category of content will only be recommended occasionally to a given user. In other words, if a user is interested in makeup videos, we'll make sure to diversify the content by only showing makeup videos occasionally.

We evaluate our recommendations to users in terms of the number of categories of content we are introducing to them, while at the same time ensuring we do not overwhelm them with any



particular type of content. This helps reduce the risk of filter bubbles, since users will be served diverse content even if our models show they have a strong interest in certain types of content.

Automated moderation systems could erroneously exclude content.

There is a risk that our efforts to ensure appropriate content on Snapchat results in some content that is appropriate being mistakenly identified and incorrectly modified. This may create for example, a risk to users rights to freedom of expression.

To combat this ‘over-moderation’, we evaluate and work to improve our automoderation in terms of precision and recall, and currently have very high auto-approval precision for For You (previously Discover) and Spotlight. In addition, as explained in the [Enforcement](#) section of this Report, we have added additional moderation transparency messages (statements of reason) and a more comprehensive appeals flows for moderated creators and content as part of our efforts to comply with the DSA.

Viewers could be watching our content but not enjoying themselves

There is a risk that the recommendation systems and models we build end up optimizing only for short-term metrics like engagement (i.e. time spent) in the Snapchat app, rather than in support of Snap’s mission of “empowering people to express themselves, live in the moment, learn about the world, and have fun together”. Our long-term objective when recommending content to users therefore goes beyond time spent and is focused on whether our users are enjoying themselves and are entertained and satisfied with their experience.

Snap evaluates the effectiveness at achieving this objective in multiple ways, in particular ensuring that we evaluate our algorithmic performance using a wide range of factors and not solely relying on user engagement (i.e. time spent).

In addition, Spotlight has been designed not to distribute sensitive (i.e. shocking) content to 13-17-year-old users’ Snapchat accounts, which includes non-glorifying discussion of self-harm and suicide content (such discussion is not prohibited on Snapchat but may still be sensitive). For users under 18, we will remove all content labeled as sensitive. For users over 18, we will limit its distribution.

We evaluate our algorithms across the above dimensions because we believe they are the drivers to the ultimate outcomes we are attempting to deliver for users: that they be (1) satisfied with our experience - which we survey regularly (i.e. quarterly) across all tabs in our app and (2) continue to use it (i.e. user retention).

Conclusion

Users find new content on Snapchat primarily through our algorithmic personalization/recommendation service. While algorithmic content recommendation systems,



like the one Snapchat uses, present a number of risks, we've designed our systems to mitigate these risks. This includes the use of appropriate terms, robust automated and human moderation, sufficient transparency with our users about the functionality of these systems, the ability to opt out of personalization, and the other mitigations outlined above.

As explained in Section 4, we have concluded that our adaptation and testing of Snapchat's algorithmic systems described above, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective mitigation measures for the risks identified.

5.7 Advertising Systems

Introduction

Snap relies on online advertising to support its business. Snap recognises that without mitigations its advertising systems also have a significant risk of giving rise to the concerns referenced in Article 34 of the Digital Services Act. Snap Advertising is a digital ad product created for advertisers who would like to easily create and manage ads that target relevant audiences on Snapchat. We process user information about Snapchatters to serve them with ads within Snapchat that we think they might be interested in. However, advertising systems in general might give rise to, amplify and/or result in the rapid and wide dissemination of illegal content and/or other harms identified in [Section 4](#), if not adapted and tested appropriately.

How do our Advertising Systems Work?

An overview of Snap's ads services can be found [here](#) and [here](#). In essence, Snapchat collects data about our users as they register, log in and use Snapchat. As is described in our [Privacy Policy](#), this data is comprised of:

- Information the user provides us
- Information we collect as the user interacts with Snapchat
- Information we collect from third parties

Snapchat Ads Manager and its various tools allow advertisers to leverage this data for targeted advertising. Advertisers can use our [Audience Insights tools](#) to see the estimated demographics, including age, as well as locations, interests and device overviews of their targeted audience. User-level data is not directly available to advertisers.

Some of Snap's advertising tools allow advertisers to use data about their customers such as customer personal data provided by our advertisers and data collected from third-party services along with our users' personal data, to provide and improve ad targeting and measurement:

- Snap [Custom List Audiences](#) - An advertiser and/or their agent can use this service to upload customer list data to Snap via Ads Manager. See the [Custom List Audiences](#)



section of our Business Help Center. Customer list data provided by advertisers is used to create an ‘audience’ of Snapchatters matching the information in the customer list data. This allows advertisers to target ads to that audience, or similar audiences, on Snapchat. See the [Custom Audiences Overview](#) in our Business Help Center.

- [Snap Pixel](#) and [Conversion API](#) - An advertiser and/or their agent can also use this service to help target their ads on Snapchat:
 - For Pixel, advertisers install a piece of JavaScript within their web pages which sends data to Snap when those pages are accessed by website visitors. See the [Install Snap Pixel](#) section of our Business Help Center.
 - For Conversion API, advertisers install Snap API code on their servers that facilitates passing web, app and offline events directly to Snap via Server-to-Server integration. See the [Conversions API](#) section of our Business Help Center .
- [Advanced](#) and [Estimated](#) Conversion are examples of the additional services that we offer to advertisers to target and measure the performance of their advertising using advanced privacy enhancing techniques.

Snap acts as a data processor of data relating to EU data subjects received from advertisers via the Custom List Audiences, Pixel and Conversion API services. It processes the information in accordance with advertiser instructions subject to its data processing agreement (which follows requirements set out in Article 28 of the General Data Protection Regulation (GDPR)).

Our ad ranking algorithm determines which ads are displayed to a Snapchatter who is in the selected audience for those ads. The ad ranking algorithm uses various signals, including prior ad interactions and social signals, to determine which ads that user is more likely to interact with and then combines this with the results of advertiser ad action for that Snapchatter, to select an ad to display. Snap analyzes prior ad interactions to target advertisements. For example, we may determine that a user is likely to swipe up on certain types of ads or download certain types of games when they see an ad on Snapchat. We may then use this information to show that user similar ads. This is explained on our [Snap and Ads Privacy and Transparency](#) page.

Snapchatter interactions with the ad (i.e. impression data) is then logged to (a) attribute impressions to conversion events (such as a purchase on an advertiser website or download of an advertiser app) to demonstrate the performance of the ad and (b) to further train the ad ranking algorithm.

Benefits

Snapchat is used by millions of people in the European Union. They use Snapchat because it fosters fast and authentic communication with those who matter most to them. It is why our community continues to grow.



We consider it is in the best interest of all our users, including 13-17s, for them to have access to the best, most entertaining version of Snapchat possible, allowing them to exercise their digital rights (such as access to information, association with others, have a voice and to play and have fun) regardless of their financial background and ability to pay. We receive feedback everyday from our users; calling for new features, functionality and improvements. We are only able to do this by raising revenue from other sources. In common with many others in the industry, this has meant turning to advertising.

Our ability to raise revenue by selling targeted advertising opportunities to advertisers means that:

- Snapchat is maintained and improved for the benefit of Snap and all recipients regardless of their ability to pay. If Snapchat was only available for a fee, it would only be accessible to those who could afford to pay the fee, restricting access to Snapchat and raising risks to fundamental EU rights to information and to access to services, particularly for Teens.
- Snapchatters benefit from being able to exercise digital rights and association with others online through Snapchat regardless of their financial background. This includes developing their voice, having fun and access to entertainment and play. Balanced use of their personal data also benefits Snapchatters by avoiding seeing advertisements that are not relevant to them (which is one complaint we have received in the past). Although Snapchatters are given options to manually hide advertisements, through the use of personal data, Snapchatters benefit from targeted advertising by seeing more relevant, age and interest appropriate adverts²⁷. The greater the revenue Snap is able to generate the more resources Snap can dedicate to supporting access to the service and teens' development.
- Advertisers benefit from being able to promote their brand and products to a Snapchat audience most likely to be interested in them. This allows advertisers to focus their advertising and avoid spending on the display of advertisements to audiences that are not likely to be interested. Snap Ads also allows advertisers to better measure the success of their digital marketing campaigns so their quality can be continuously improved.

However, notwithstanding the benefits advertising systems bring to our users, to Snap and our advertisers, we recognise that our targeted advertising will only operate in the best interests of all our users provided that the processing of individuals' personal data (including by way of profiling) to facilitate the sale of ads that fund Snapchat does not result in our users being subject to 'economic exploitation'. Privacy and Safety are central to Snapchat's values. When we first introduced advertising to Snapchat, we ensured those advertising systems appropriately balanced the legitimate benefits explained above with individuals' fundamental rights and freedoms, in line with Snap's strong privacy and safety principles. We have continued to uphold

²⁷ N. Fourberg e.a., on 'Online advertising: the impact of targeted advertising on advertisers, market access and consumer choice', 2021, [url](#).



these values throughout Snapchat's life, adapting and testing our advertising systems to mitigate risks they may give rise to as identified in [Section 4](#) of this Report.

Adaptation and Testing

In line with Article 35.1(e), Snap has adapted Snapchat's advertising systems and adopted targeted measures aimed at mitigating the risks presented by its advertising systems, including by limiting or adjusting the presentation of advertisements on Snapchat, to help address the risks identified in [Section 4](#). Considering each risk and its mitigation(s) in turn:

Reasonable and Proportionate Targeting

We recognise that, as a platform, we have a responsibility to raise revenue in an appropriate manner, and we take this responsibility very seriously. We want to ensure advertisers are not targeting specific individuals on our platform and that users do not feel like their privacy is being compromised by our advertising. We also want to prevent advertisers from manipulating small audiences with microtargeted campaigns, particularly for political ads.

In order to mitigate this risk,

- Most of the ads on Snapchat, including all political ads, require a specific minimum audience to be targeted. This prevents adverts from being micro targeted.
- Snap generally has a short retention period for user data. We do not store data for excessive periods solely for monetisation purposes.
- Advertisers can only use our data indirectly via the targeting tools available on Snapchat. Amongst other things, this allows advertisers to target audiences based on a limited number of high level interest-based lifestyle categories (SLCs) audiences (none of which are available for targeting 13-17 year olds in EU and UK), which we have inferred a Snapchatter may be interested in. They are based on high level, non-sensitive categories inferences, such as Business News Watchers, Sports Fans, and Fashion & Style Gurus, that users can see and control in the app, as detailed in [this support page](#). The interest categories are intentionally short-lived (13 months), sufficient to allow a year-on-year comparison, and inferences are generally based over the previous 90 days. All users can view and manage their advertising interest categories in settings. None of these SLCs are aimed at 13-17s specifically and the user-level data is not directly available to advertisers.

We feel confident that our approach to advertising is reasonable and proportionate, as we have a low incidence of issues in relation to age targeting. Our approach to targeting minimums is based on mathematical analysis by our privacy engineering teams.

Advertising policies



As explained in [Section 4](#) above, our advertisers could use our advertising systems to disseminate information that is illegal or could otherwise harm users, impact their fundamental EU rights or negatively impact public security or health.

As explained in the [Terms](#) section of this Report, we ensure advertisers are clear about their obligations, we have robust ad policies to prevent inappropriate and illegal advertising on our platform. The systems used by advertisers to create and submit advertising (such as our Snap Ads Manager), have been adapted to require agreement to these Terms and provide easy access to guidance on what is required.

We test advertisers' compliance with these Terms using our Advertising Review process before advertising can be published. See below for more information.

Advertising Review

Notwithstanding that advertisers agree to our Terms, they may still deliberately or mistakenly seek to publish advertisements that violate our advertising policies or the law.

As explained in the [Moderation](#) section of this Report, in particular the part relating to [advertising moderation](#), we use a combination of automated and human review to prevent ads that violate our policies or the law from appearing on Snapchat. We reject hundreds of thousands of adverts globally each month. Fraudulent advertising accounts for the majority of these rejections and our advertising review teams are particularly vigilant for this form of violating advertising. This also includes ensuring inappropriate ads are not targeted at Teens. Our review takes account of the targeted audience i.e. if the ad is for alcohol and the selected demographic for the ad includes Teens, then it will be rejected. We use inferred age, as well as declared age, to help ensure Teen users see ads that are appropriate for their age. Inferred age is regularly checked to ensure it is up-to-date.

We monitor ad reporting and enforcement data to ensure our review process is catching a reasonable and proportionate level of violating adverts.

Advertising Reporting

Although we have an advertising review process in place to prevent the publication of advertisements with information that violates the law or our policies, it is possible that some of these advertisements may be missed or incorrectly reviewed and be published.

As explained in the [Enforcement](#) section of this Report, our advertising systems have been adapted with an easy mechanism for advertisements to be reported by Snapchatters from within the Snapchat app as being inappropriate along with the reason for the violation. Based on the number of reports, we will take down the ad or send it to human review for additional moderation.



All ads that are reported are reviewed by our human moderation team. Upon reporting the ad, Snapchatters are able to select a reason and write in comments. Both the reporting reason and the comment are provided in the moderation task, as well as the number of reports. We closely monitor sentiments of the ads on our platform and when ads are taken down, we inform the advertiser. We also monitor the aggregate number of reports for advertisements on a regular basis.

We monitor ad reporting and enforcement data.

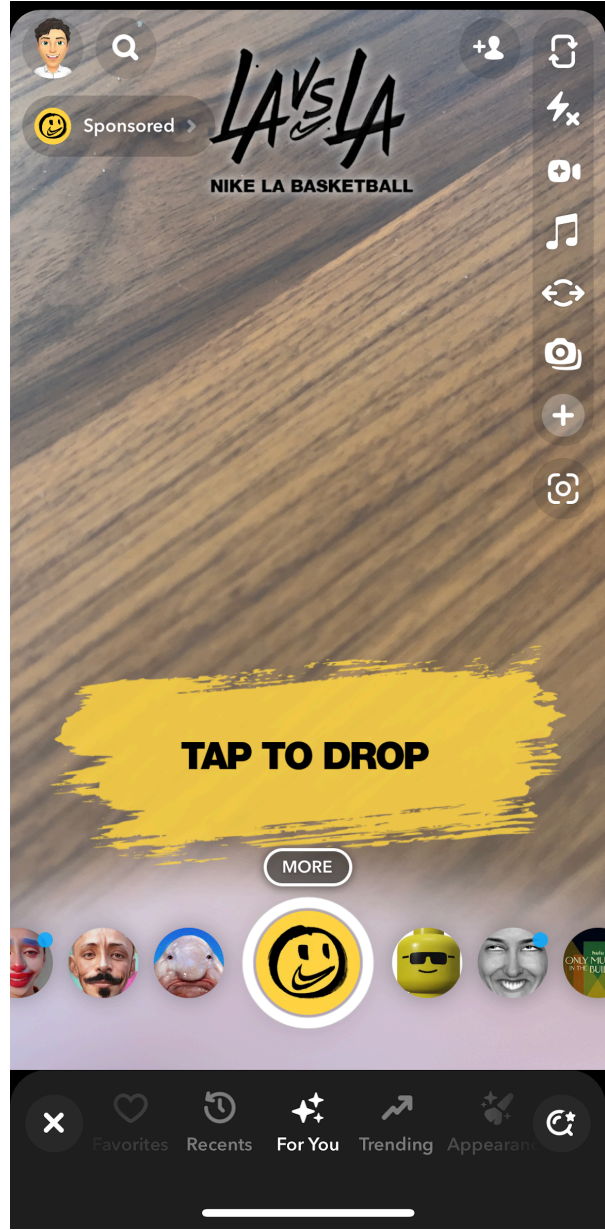
Ad Markers

If users are not aware when content is an ad or sponsored or other commercial content, there is a risk that without additional mitigations that this may lead to confusion, deception and exploitation.

We automatically place an “Ad” marker on all paid ads that run on Snapchat. Sponsored Lenses say “Sponsored”. Our commercial content policy requires all organic content posted by influencers to be marked appropriately. We now offer a “Paid Partnership” tag tool that influencers and users may use when they post commercial content to help them comply with this policy and their legal obligations. We automatically place an “Ad” marker on all paid ads that run on Snapchat. Sponsored Lenses say “Sponsored”. Our commercial content policy requires all organic content posted by influencers to be marked appropriately.

Ad marker example

Sponsored Lens example

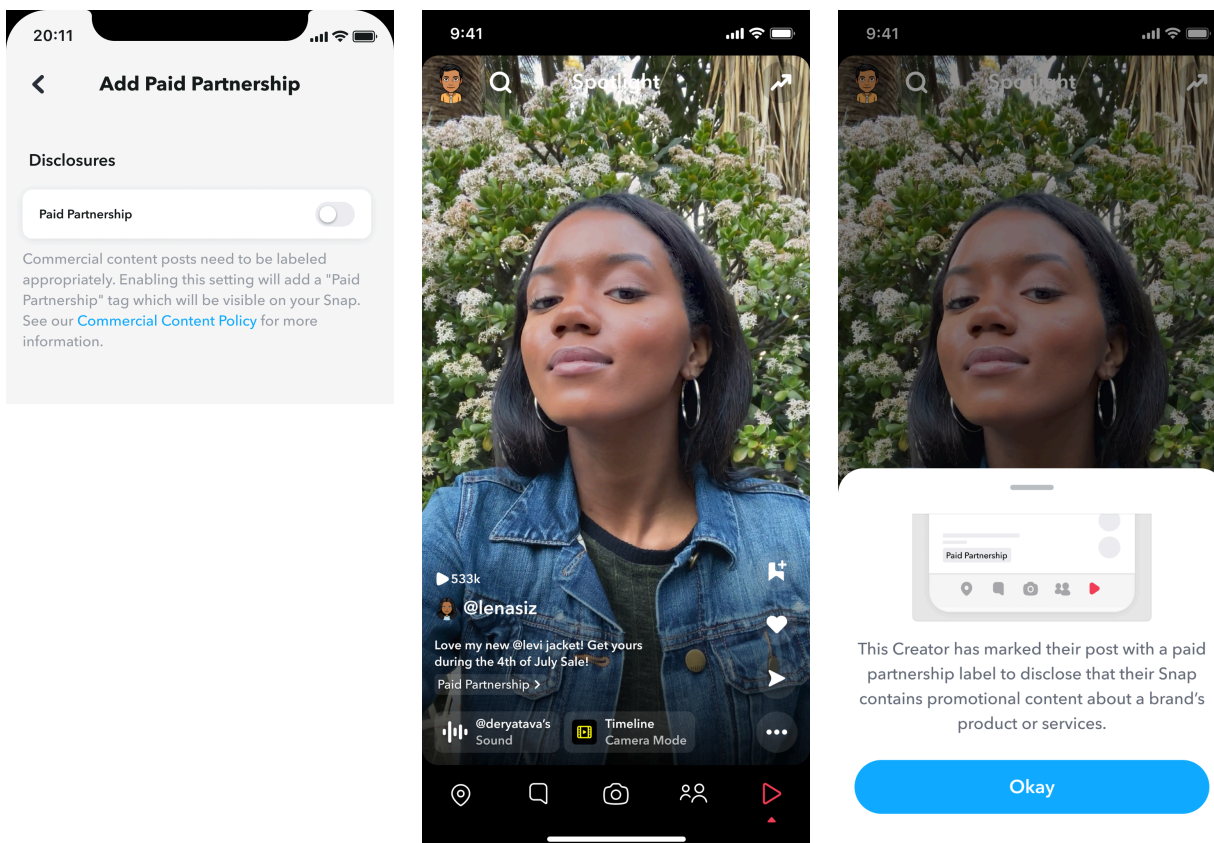


We now offer a “Paid Partnership” tag tool that influencers and users may use when they post commercial content to help them comply with this policy and their legal obligations.

[Add Paid Partnership](#)

[Paid Partnership label](#)

[Paid Partnership Explainer](#)



Transparency and Control

Some users may have specific vulnerabilities or other reasons to be concerned about any use of their personal data for targeting ads. If users do not understand how advertising works, they may not be able to confirm whether they should be concerned or exercise any choices they may have.

As explained in the [Transparency](#) section of the Report, our privacy center provides extensive information regarding our processing of personal information. This includes a [dedicated page](#) explaining how we use personal data for advertising purposes. We offer choices for users to control the data that's used to determine the ads they see. In the European Union, we have introduced controls to turn off most personalized ads except those based on real time location, language, age and device type, and this is always turned off for teen users in the European Union. All users can restrict our use of third party data and being included in advertiser supplied audience matches for ads targeting.

We use pre launch testing and our ad review process to help ensure these controls work as designed.

Special category data

Special categories of data concern information revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data,



biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. This information may have a greater risk of causing harm if used to target ads. A famous example of this concern relates to supermarket Target, which profiled purchase information to determine if a woman was pregnant, and revealed the pregnancy to a teenage girl's father by mailing vouchers for baby accessories²⁸.

Our targeting parameters such as Lifestyle Categories do not include special / sensitive categories. In addition we require in our [Terms](#) relating to advertising that advertisers do not send us this data from their sites. We do not allow advertisers to target audiences based on sensitive categories. When we discover advertisers sending us this data we remove it.

Our legal team has reviewed and confirmed our Lifestyle Categories do not include special / sensitive categories. Any changes must be reviewed by our Legal team as part of the mandatory product review which forms part of our privacy and safety by design processes.

Special targeting Models

We offer special targeting models for certain advertisers in regulated sectors (particularly in the United States) to ensure they meet rules to prevent discrimination, such as the housing, credit or employment sectors. We use pre launch testing and mandatory legal and privacy engineering review of any significant changes to these models.

Political ads

Political advertising presents a higher risk of misinformation and negative effects on democratic and electoral processes, as identified in [Section 4](#).

Snap ensures that ads shown are in line with Snapchat's [Advertising Policies](#), which contain a subsection on political and advocacy advertising policies. Snap allows political, election-related, advocacy and issue ads, but sets additional requirements and placed additional transparency safeguard in relation to publishing these types of ads:

- Political advertising must comply with all applicable laws and regulations, including all national election laws, copyright law, defamation law;
- All political advertising must include a "paid for by" message in the ad that is followed by the name of the paying person or entity. Snap may also require a "paid for by" disclosure on ad content that links to political content, ad content for political merchandise, or in other cases in Snap's sole discretion;
- Like all ads on Snapchat, political ads must comply with Snap's Terms of Service, Community Guidelines, and our Advertising Policies.
- We encourage political advertisers to be positive. But we don't categorically ban "attack" ads; expressing disagreement with or campaigning against a candidate or party is

²⁸ Drive Research: How Target Used Data Analytics to Predict Pregnancies ([url](#)).



generally permissible if it meets our other guidelines. That said, political ads must not include attacks relating to a candidate's personal life.

- Snap will review political ads on a case-by-case basis. To get started, political advertisers are required to fill out our [political advertiser form](#). Snap reserves the right to reject, in our sole discretion, or request modifications to ads that we believe violate the standards listed above or that are otherwise inappropriate. Our discretion will never be exercised with the intent to favor or disfavor any candidate, political view, or political party.
- Snap may publicly display and otherwise disclose information relating to political advertising, including ad content, targeting details, delivery, spend, and other campaign information.
- Snap has for some time provided transparency for political ads with its [political ad library](#).

As explained in the moderation section of this Report, higher risk adverts (including political adverts) are subject to human review.

Ads Library

There may be a higher risk that advertising will violate our terms or the law, in particular content misleading information, if the Snapchatter community and wider society does not have visibility into the history of ads over the past year that have run on Snapchat and some details about the targeting and reach of those ads.

Snap has recently introduced a new [ads library](#) (as required under Article 39 DSA) which provides increased transparency for ads - not just political - that are currently running, and historically have run in the past year, directed to EU users on Snapchat. This ads library is available to anyone, can be searched / filtered / sorted based on pre-defined parameters (e.g. country targeted, advertiser name, etc). This allows anyone to check who has paid for an advert and, if different, on whose behalf is the advertisement being published.

As the ads library is a brand new feature, it underwent pre launch testing to ensure it met design specs and will continue to develop based on further testing.

Freedom of Expression

The purpose of Snap Ads is to amplify advertisers' commercial messages, and as a result the content is rarely political or rather than expressing views. We have specific procedures for political ads. As a result, the risk of a negative impact on freedom of expression from Snap's other mitigations listed above is low.

Conclusion

Targeted advertising on Snapchat is necessary to ensure we can continue to provide a free service to all users regardless of their ability to pay. We have taken extensive steps to ensure our approach to targeted advertising appropriately balances the interests of Snapchatters, Snap and



advertisers. We have also put in significant measures to prevent fraudulent and other advertising that may be harmful or against the law. We human review most adverts and reject hundreds of thousands of them each month to keep Snapchat's community safe.

As explained in Section 4, we have concluded that our adaptation of Snapchat's advertising systems described above, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective mitigation measures for the risks identified for Snapchat's in-scope services.

5.8 Protection of Minors

How We Think About the Protection of Minors

Snap's utmost priority is protecting the safety and wellbeing of our users whilst ensuring they have a positive experience online. Privacy, safety and security are key values of the company and at the core of our value proposition to our users. We put significant thought and consideration to ensure those values are reflected into the architecture of our platform, and into the design and implementation of our products, policies, and enforcement actions. Since Snapchat's inception, we have embraced a [privacy and safety by design](#) approach and decided that our platform architecture and product choices should play a major role in risk-mitigation. They can be found in our privacy and safety by design principles.

Specific Safeguards for Teens

In addition to our defaults for all users, we have added protections in place for Teens, to help mitigate risks in a number of ways.

Family Center / Parent Tools

Our in-app parental supervision tool, Family Center, gives parents and trusted adults visibility into their teens' friends list and who they have messaged in the last 7 days. Parents can also access Content Controls, giving them the option to limit sensitive content on their teens' For You and Spotlight feeds. Parents are also able to easily report accounts that may be in violation of our Community Guidelines and have access to helpful resources directly from the app.

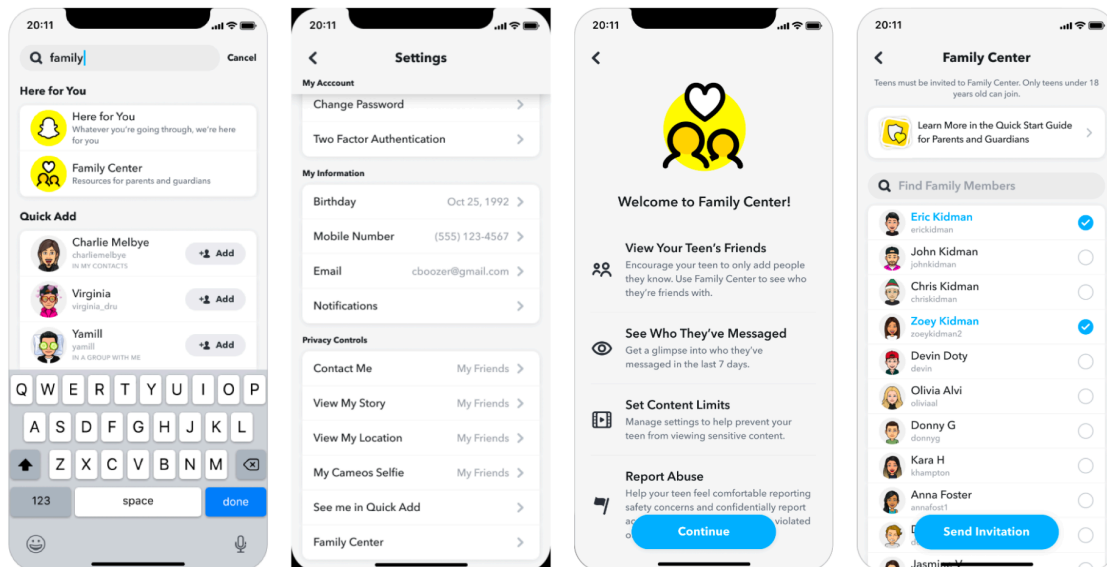
Our goal was to empower both caregivers and teens, balancing parents' desire for more insight with teens' desire for autonomy and privacy - notably ensuring that teens' messages remain private. We took our time to design this in a thoughtful way, engaging in user research and surveys, competitive research, focus groups and interviews with both teens and parents, feedback sessions with dozens of online safety experts and academics, reviews with our Safety Advisory Board, and extensive cross-functional internal reviews, including by our Product Legal and Privacy Engineering teams.



In their annual report,²⁹ Jugendschutz.net, the joint competence center of the German federal and state governments for the protection of children and young people on the Internet, highlighted Family Center as a positive example in the area of parental tools and support on social media platforms. The report concluded that Family Center can help teens and parents talk about negative experiences, contacts, or time spent on the platform. At the same time, it noted the opportunities for teen control, such as having to agree to parental guidance.



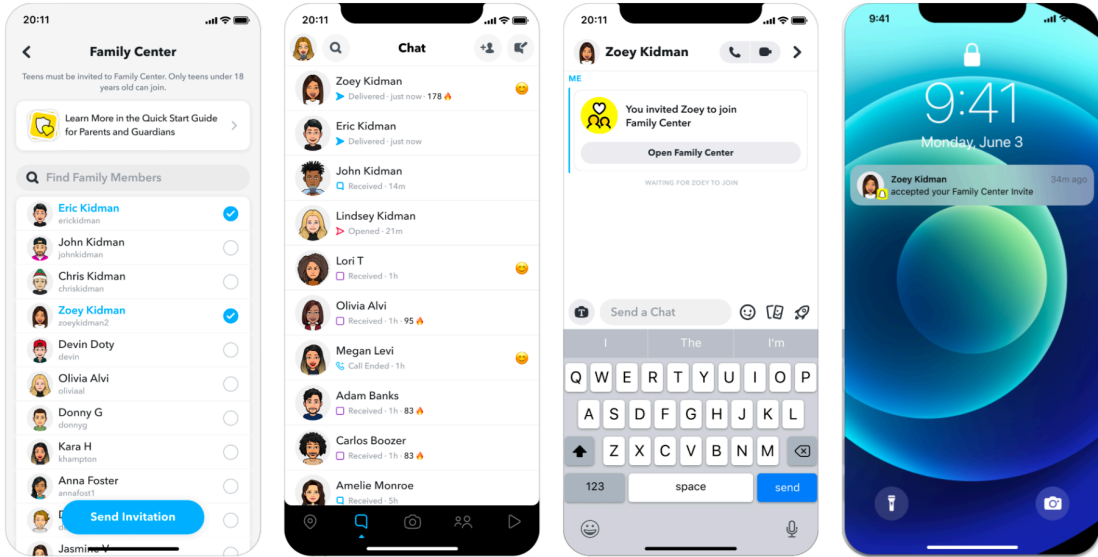
Finding Family Center



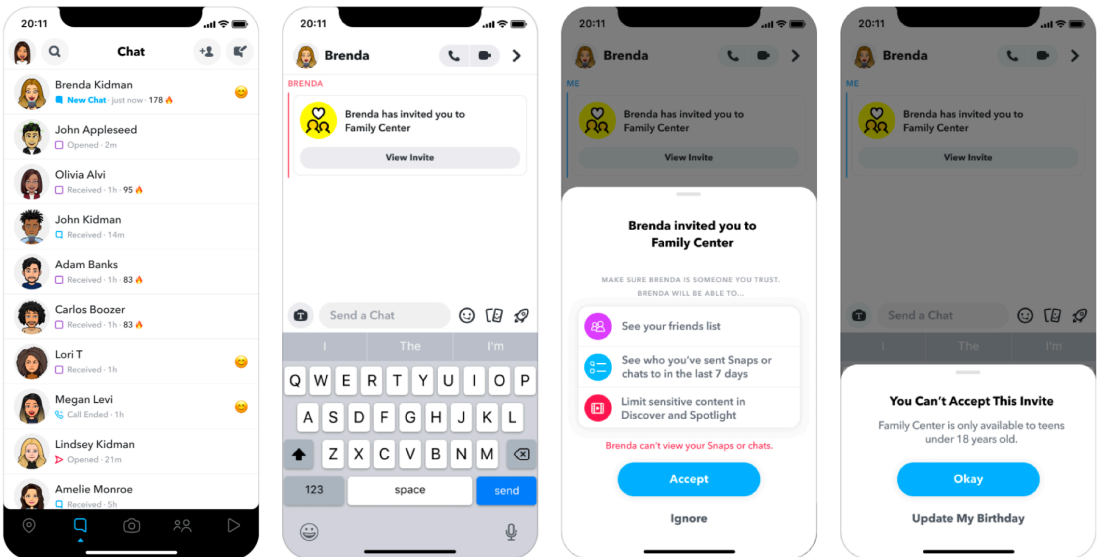
²⁹ Jugendschutz, 'Jugendschutz im Internet - 2022 Bericht', April 2023, [url](#).



Inviting Teens to Family Center



Joining Family Center





Using Family Center



Age Gating and Age Appropriate Content

As a central communications tool for young people, we take our responsibility to protect teens on our platform seriously. We know age verification is an industry-wide challenge everyone is trying to solve, and we are already working with industry peers, regulators, and third-party technology providers on possible solutions. We look forward to continuing these productive conversations to achieve solutions that work for everyone.

Snap currently takes a risk based approach, which relies on:

1. Declared age to limit access to Snapchat to its target 13+ audience:
 - a. Our declared age process has been designed to meet industry standards.
 - b. Users must add their birthday when registering for an account, and we do not allow users declaring they are under the age of 13 to create accounts.
 - c. We don't only rely on users' self-reported age, but have techniques to infer users' actual age, which may be a combination of their birthday, activity on the app, and the ages of people in their network.
 - d. If we determine an account belongs to someone younger than 13, we take action to terminate it.

2. A combination of declared and inferred age techniques for stronger age assurance to limit under 18 access to certain content and features targeted at more mature audiences:



- a. For You: Our publisher guidelines require content to be appropriate for a 13+ audience. If it's not, our guidelines require publishers to age gate their content.
- b. For Spotlight: We explicitly prohibit sexually explicit content on Snapchat. For Snapchatters between the ages of 13-17, we also have measures in place to help prevent sexually suggestive content from appearing. We have developed machine learning classifiers which work to identify sexually suggestive content and filter it from the experience before human intervention. In addition, our Spotlight content is evaluated by human moderators before being widely distributed. and
- c. Ads: We restrict ads based on the user's age. For example, ads for dating services must be targeted to users over 18 and must not be provocative, overtly sexual in nature, or reference transactional companionship. Similarly, ads for alcohol products must be age targeted to at least 18+, or the applicable minimum drinking age in the respective country where the ad is running.

We have a dedicated working group that is overseeing our age assurance efforts. As part of this work, the group is evaluating our approach, and discussing with industry contacts and third party age assurance vendors, to ensure we keep pace with industry practice. We are also supporting legislators and NGOs in the UK and France, to enhance the role of app stores, online devices, and web browsers in providing appropriate interfaces for age assurance and parental controls to facilitate consistent, effective and efficient solutions for the online ecosystem.

We hope to be able to significantly contribute to the EU Age Appropriate Design Code working group. We support its goals and believe it is important we develop effective industry wide standards for assessing a high level of privacy, safety, and security for Teens, in line with existing Age Appropriate Design Codes, Data Protection Impact Assessments and Privacy and Safety by Design obligations that already exist in Europe and other parts of the world. We believe such a code should consider both online platforms and the 'gateways' (such as device operating systems, app stores and web browsers) through which parents and Teens engage with such platforms.

Additional safeguards include those described for Snapchat's online platforms and other services on the following pages:

- [Teens | Snapchat Privacy](#)
- [Snapchat Safeguards for Teens](#)

Conclusion

We take the protection of Teens seriously at Snapchat. We've designed Snapchat to take privacy, safety, and security into account. Our key tenets include acting in the best interests of teens, strict default settings for all users, respecting teens' freedom to express themselves safely while recognizing their right to information about the world. We aim to achieve these tenets by positioning parents and guardians to guide teens in the responsible use of our platform, attaching



a heightened safety interest to Teens using our products, and establishing processing to ensure we develop products in a way that upholds these tenets. We have implemented these tenets through the use of Family Center, focusing on age appropriate content, reporting and blocking mechanisms, and putting in place appropriate protections and limitations on private messaging, friending, public content, and advertising.

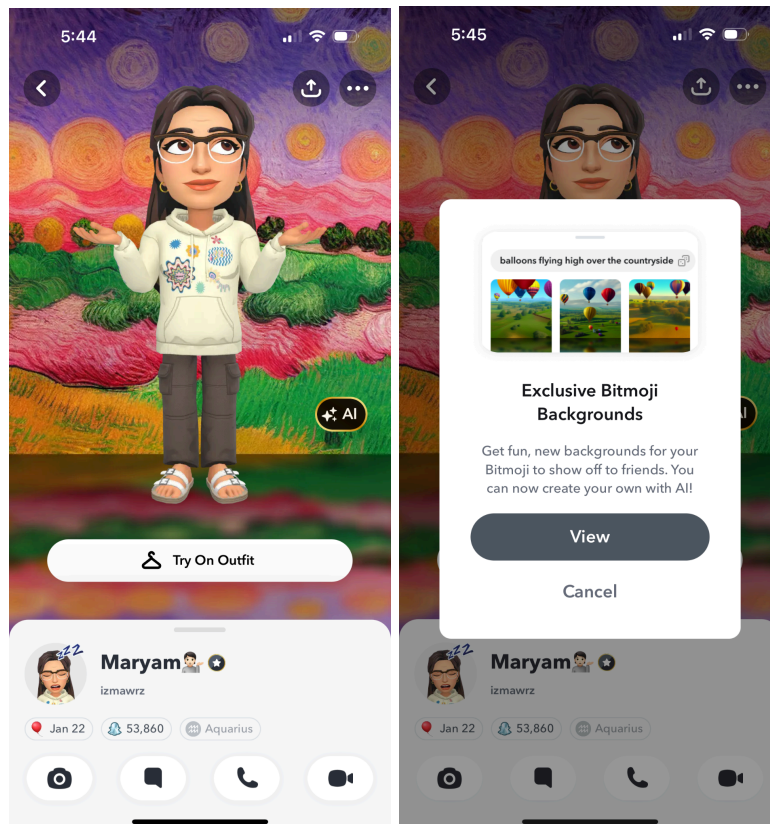
As explained in Section 4, we have concluded that the targeted measures we've taken to protect the rights of the child, including age verification and parental control tools, in combination with the other mitigations explained in this Section 5, are reasonable, proportionate and effective mitigation measures for the risks identified for Snapchat's in-scope services. We are actively working to support the efforts of the Commission and others to establish an EU Age Appropriate Design Code and consider whether further mitigation measures may be reasonable, proportionate and effective mitigation measures for online platforms, 'gateways' and other online services.

5.9 Content Authenticity

As identified in the [Risk Assessment Results](#), although we take harmful false information, fraud and spam or impersonation very seriously, there is not a high prevalence of harmful information of this nature on Snapchat.

We have applied icons to certain generated images. For example:

- Although out of scope of our risk assessment, we currently show the AI sparkle watermark when Snapchatters generate Bitmoji Backgrounds to be displayed when viewing their Friends' profiles (This is premium feature only available with Snapchat+)



- We display an icon in some Lenses that manipulate an image of a Snapchat to make them look younger.

However, the question of whether to use markings for generated or manipulated images is a challenging one for which there is no simple solution. It depends on the purpose of the watermark. Whether content is AI-generated or not isn't always a binary question -- as AI tools become more integrated across platforms, we should expect that "AI assisted" content--at different stages of the content's development--will be very prevalent; standards for provenance and watermarking should account for this, and society should not expect watermarks to provide a silver bullet solution.

Whether watermarking is appropriate may also depend on the nature of the generated or manipulated image. Most AI generation tools on Snapchat create images that are not photorealistic and it is obvious from reasonable observation that an image has been generated or manipulated using AI or other computational tools.

As a result, Snap is carefully monitoring the ongoing debate regarding whether and how best to use prominent markings in an attempt to identify generated or manipulated images, audio or video.



Conclusion

Content authenticity is a very challenging topic without a silver bullet. Snap is very conscious of the issues and is carefully monitoring this regarding whether and how best to use prominent markings and other measures to distinguish content that falsely appears to be authentic or truthful. Snap has used markers in a few cases to date in relation to content generated by tools made available on Snapchat. However, at present, most of the content on Snapchat that is generated or manipulated is not photo-realistic and is by nature distinguishable.

We have concluded that our position on the use of marks to distinguish content that falsely appears authentic or truthful is, in combination with the other mitigations explained in this Section 5, reasonable, proportionate and effective for the risks identified for Snapchat's in-scope services.

5.10 Trusted Flaggers

Trusted Flagger Program

Snap's Trusted Flagger Program was developed to help non-profits, non-governmental organizations (NGOs), select government agencies, and safety partners support the Snapchat community by leveraging a special channel to report content that violates Snapchat's [Community Guidelines](#). Trusted Flaggers send a report via email to a dedicated, confidential email address, containing the username(s) they wish to report and a completed report form with details of the violation. The email used is a high priority channel and reports are reviewed in 24 hours or less. Snap responds to the Trusted Flagger with details of enforcement action taken. This channel is reserved for urgently harmful situations and is designed to supplement in-app reporting, which is still very much encouraged.

Our Trusted Flagger Program allows us to gain insight from the Trusted Flaggers over the types of harm they are encountering, and the behavior of victims in these circumstances. In addition to providing a specific reporting channel, the Trusted Flagger Program also allows us to build strong relationships with Trusted Flaggers. Snap makes use of our strong relationship with Trusted Flaggers to give product safety updates, encourage the promotion of our safety tooling and provision of safety resources (like links to our [Safety Center](#)).

As at the date of this report, no Trusted Flaggers have been awarded the status of 'trusted flagger' pursuant to Article 22(2) DSA. We will be monitoring for the Commission publication of the entities that have been awarded pursuant to Article 22(5) in due course. We expect our Trusted Flagger Program will evolve to meet the DSA requirements through the creation of a reporting form to structure incoming reports and details of the metrics that must be collected by both Snap and partners in the DSA Trusted Flagger program.



Conclusion

Snap has an existing, carefully managed Trusted Flagger Program with valued member organizations from a wide array of countries including many in the European Union. Snap looks forward to evolving its Program to incorporate organizations that have been awarded trusted flagger status under the DSA.

As explained in Section 4, we have concluded that Snap’s Trusted Flagger Programme, in combination with the other mitigations explained in this Section 5, is reasonable, proportionate and effective for the risks identified for Snapchat’s in-scope services.

5.11 Dispute Settlement Bodies

We invest significant resources in our community support teams who work to resolve queries and complaints received from Snapchatters and others. Our Transparency Report provides a breakdown of the queries we receive.

Nevertheless, in line with DSA requirements (Article 21), we will proceed with informing our users in Section 19, Dispute Resolution and Arbitration, of our Terms of Service about the possibility of out-of-court dispute settlement in case they are not satisfied about the outcome of the internal complaint-handling system via the relevant dispute resolution body in due course following their certification by the Digital Service Coordinators (DSCs). We are sensitive to our obligation to provide information about the option of an out-of-court dispute settlement in an easily accessible, clear user-friendly manner.

Conclusion

Snap is committed to resolving user disputes effectively and in line with DSA requirements. Furthermore, we support establishing an EU-wide settlement body or an EU portal for better user interactions. This approach would ensure consistent application of rules across all EU member states and provide a simplified, single point of access for operators.

As explained in Section 4, we have concluded that Snap’s current approach to Dispute Resolution, in combination with the other mitigations explained in this Section 5, is reasonable, proportionate and effective for the risks identified for Snapchat’s in-scope services.

5.12 Codes and Crisis Protocols

Cooperation

Snap highly values cooperation with other providers and industry experts as a way to share best practices and learning experiences that can enhance risk mitigation strategies. We are highly



committed to industry partnership to steer progress in the fight against illegal and harmful content online. Snap is also actively involved in the work of a number of EU-based trade associations to contribute to the policy debate to support the development of a proportionate regulatory framework to promote online safety.

Codes of Practice

The DSA establishes that the Commission and the European Board for Digital Services (‘the Board’) shall encourage and facilitate the drawing up of voluntary codes of conduct at Union level to contribute to the proper application of the DSA (Article 45).

Snap welcomes the opportunity to support industry-wide efforts to promote risk-mitigation practices in the form of voluntary codes. As a company with limited resources, Snap is constantly required to prioritize and ensure its resources and efforts are focused on where the biggest risks and challenges for the company are. As we advance in our learning curve from our DSA risk assessment, we will continue to prioritize our engagement with voluntary codes where we see the highest relevance for Snap.

EU hate speech Code

As part of its long-standing commitment to fight harmful and illegal content, Snap signed onto the [EU Code of Conduct to counter illegal hate speech online](#) in 2018. Since joining the code, Snap has successfully passed all the evaluations and in the course of the last [monitoring exercise](#) (2022), and for the 5th consecutive year, **Snap did not receive any notification.**

Additionally, in the course of 2022 Snap has worked closely with the European Commission and other signatories to further strengthen some of the Code commitments by reinforcing and better framing the existing cooperation between IT companies and CSOs, beyond the remit of the monitoring exercises. This work led to the publication of an [Annex to the existing code](#) in December 2022.

Since February this year, Snap has been engaging with the European Commission team and regularly cooperating with other industry signatories with a view to contribute to the update of the EU Hate Speech code to bring it in line with the DSA by 2024.

FSM Code of Conduct

In September 2017, Snap joined 'the Freiwillige Selbstkontrolle Multimedia- Diensteanbieter e.V. (FSM), an officially recognized voluntary self-regulation association for the protection of minors in online media.



EU disinformation code

Snap has not yet signed up to be a member of the EU disinformation code. Being very resource constrained and considering our limited exposure to this type of risk, we have so far opted not to join.

EU AAD Code

Since our VLOP designation, we have engaged with different stakeholders at the European Commission³⁰ to flag our interest to contribute directly to the discussion and working group on the AAD Code.

Crisis Protocols

Snap has set up a number of crisis management protocols to help the organization swiftly tackle unexpected incidents and help minimize their impact on our service, users and operations.

Conclusion

Cooperation with external stakeholders is a very important element of risk mitigation for Snap. Knowledge sharing and best practice development with experts and peers are key to strengthen and increase the effectiveness of our internal risk mitigation measures. This is why the company has signed up to several voluntary codes and is actively engaged in many different international fora and associations to steer constructive debate and best practice development in areas like CSEAI, protection of Teens, and hate speech. We will continue to monitor our risks and prioritize interventions on the most severe risk areas. When it comes to dealing with unexpected events resulting in heightened levels of risks for the platform, we have processes in place, which we will continue to refine and formalize.

As explained in Section 4, we have concluded that Snap’s current approach to codes of practice and crisis protocols, in combination with the other mitigations explained in this Section 5, is reasonable, proportionate and effective for the risks identified for Snapchat’s in-scope services.

³⁰ Meeting with DSA DG Connect F2 team in February and June 2023; Meeting with DDG Renate Nikolay in May 2023; Meeting with Cabinet Suica in May 2023; meeting with EVP Commissioner Vestager in June 2023; exchanges with DG Connect team G3 in June and July.



6. Ongoing Risk Detection and Management

Snap has developed a number of practices to detect and manage risks to Snapchat’s in-scope services. This includes: (1) the establishment of a Platform Risk Framework based on Snap’s product values, established international human rights principles, and risk-based metrics such as prevalence and severity analyses; (2) the designation of a senior, cross-functional team responsible for applying the framework and assessing its outcomes; (3) development of a repository of internal resources to support the detection and management of risk—these include harm severity assessments; prevalence metrics; reporting data reviews and a library of Terms and policy resources; and (4) continual improvement and assessment through our Digital Well-Being Index (DWBI) Initiatives and Safety Advisor Board.

This Section of the Report provides further details of these practices pursuant to Article 42.4.(b) (reporting on the mitigation measures relating to Article 35.1.(f)) and Article 42.4.(e) of the Digital Services Act.

Platform Risk Framework

Snap has implemented a platform risk framework that draws on a combination of Snap’s product values, established international human rights principles, and risk-based metrics such as prevalence and severity analyses.

The framework is divided into two parts that borrow from relevant, longstanding elements of the international human rights framework: (1) identification of core platform governance values; and (2) a set of balancing principles for weighing those values against risks to our community and other harms. Reference to both of these elements in conjunction with one another provides a consistent approach for responsibly reviewing proposed harm mitigations with attention to foundational values.

As a result, we have a responsible, rights-respecting approach to platform governance and detecting and managing risk. We anticipate that our DSA Compliance Team and Cross-Functional Working Groups will review this approach periodically to ensure it is in line with DSA requirements and global best practices.

DSA Compliance Team and Cross-Functional Working Groups

The DSA compliance function is part of Snap’s long-established Data Privacy and Regulatory Governance and Risk Management team, overseen by the General Counsel and including senior personnel from the company’s Legal, Public Policy, Product, Engineering, Trust & Safety and Information Security teams – which is responsible for coordinating, overseeing, and implementing



the Snap privacy and regulatory program. These activities include an annual assessment of risks, certification of controls, evaluation of controls, and review of related policies and procedures.

Compliance Function

The compliance function significantly overlaps with the Data Compliance and Data Protection Officer function, including the requirements to closely cooperate with the EC, responding to inquiries and monitoring compliance with the DSA, and conducting risk assessments. The head of the compliance function reports to individuals who are members of Snap's exec meetings and frequently updates the Board and executives on DSA related matters. This satisfies the Article 41 requirement for the head of the DSA compliance function to report directly to the management of the company. The head of the DSA compliance function and our compliance officers were formally appointed during a board meeting on Monday July 24, 2023.

DSA Cross-functional Working Groups

In practice, Snap's compliance function extends well beyond our formal compliance officers. At Snap, privacy and safety by design decisions are typically made by cross-functional teams. We have a long standing cross functional team across Product, Eng, DSA Compliance Officer, Operations, Policy, Legal, Comms, Trust & Safety, and Privacy teams which meets very regularly to address risks flagged through various mechanisms such as industry reports, current events/news, internal data analyses and investigations, and feedback from regulators to assess problems, prioritize them and agree on strategy and execution plans to resolve identified risks. The DSA compliance officer is also part Internally the team is called the Safety XFN. Every quarter the team meets in person and virtually to align on priorities for the next quarter, and reflects on safety improvements that were made in prior quarters. These findings are also presented to senior leadership on a regular cadence.

This team has also established a dedicated DSA working group to ensure the cross functional teams activities continue to align with the requirements of the Digital Services Act. Based on the problem areas and findings we have instituted numerous types of changes including product experience and design changes, extension of detection and enforcement mechanisms, policy and operational process changes, introduction of support and educational resources for users and opportunities for users to flag certain types of harmful content or seek redressal mechanisms.



Integrating DSA into our Privacy and Safety by Design Process

Ongoing Risk Management

As explained in this Report, Snap has diligently identified, analyzed and assessed systematic risks relating to Snapchat's in-scope services, and has specified the mitigations it has in place to address them, as required by Articles 34 and 35 of the DSA.

Snap recognises that, as well as carrying out this assessment annually, it must also re-assess prior to deploying functionalities that are likely to have a critical impact on the risks identified (and therefore the mitigations specified to prevent them). This is required by Article 34 of the DSA, but is also an industry standard practice to ongoing risk management.

Risk assessments and mitigation obligations are being an increasingly common tool of digital service regulation. In Europe, such obligations are not only imposed by the DSA, but also GDPR (in the form of Legitimate Interest Assessments (LIAs) and Data Protection Impact Assessments (DDIAs)) and also in the UK and several EU Member States, the Age Appropriate Design Code (AADC) Assessments (or their equivalent) and shortly the UK Online Safety Act's safety assessments. It is important that Snap is able to manage these, often overlapping European requirements (in addition to other global requirements) in an efficient, effective and operationalised manner.

Existing Privacy and Safety by Design review process

Privacy and safety by design is a cornerstone of Snap's approach to designing and launching its products, and is built into Snap's compliance program. As highlighted above, at Snap, our mission is to empower people to express themselves, live in the moment, learn about the world, and have fun together. We believe that privacy and safety are foundational to the success of our mission.

Snap already has an extensive privacy and safety by design review process to assess privacy and safety risks in the design and development of Snapchat. As part of its privacy and safety by design program, Snap documents a review prior to new product and feature releases that materially affect how Snap collects, uses, handles, or discloses personal data of its users.

Products and features that process personal data go through a privacy by design and safety by design review process. Because the vast majority of features process personal data, this de facto means that most of the products go through this review process particularly when making changes that are likely to have a critical impact on privacy or safety risks. This process includes cross-functional stakeholders, including engineers, product managers, designers, legal, policy, and dedicated privacy engineers. Both legal and privacy/safety engineering have to sign off before a product can be launched. Where a product meets the threshold of a data protection impact assessment and/or an age appropriate design assessment, we require this to be completed and approved before the product is launched.



Integrating DSA Risk Assessments into our existing process

In order to make use of this process, Snap has designed a new Digital & Data Impact Assessment framework that combines our privacy, safety and security obligations into a single risk assessment, that is completed dynamically depending on the nature of the Snapchat feature being assessed. This new DDIA is embedded with our existing privacy and safety by design process and requires our cross-functional team to consider if a product change results in a significant impact on our existing consolidated DDIA assessment (including the DSA aspects). If so, this is required to be re-assessed before the product change can language.

As a result, we are able to detect and manage our DSA risk assessment and mitigation obligations on an ongoing basis.

Prevalence Testing

To more holistically measure abuse on Snapchat, we measure Policy Violating Prevalence (PVP) via random sampling of Public Stories to estimate the percent of policy-violating views. This is a continuous daily sampling exercise. We have used this to uncover blindspots and prioritized efforts to close those gaps such as improvements to our proactive detection mechanisms, infrastructure improvements and agent training. It has proven extremely valuable to advancing our mission to keep Snapchatters safe.

External Request Monitoring and Review

We produce a semiannual (every 6 months) [Transparency Report](#), that captures our Community Guidelines enforcement data, law enforcement operations data, and copyright & trademark data. Internally we also review additional breakdowns of this data and, in preparation for our DSA compliance, we have started to review data relating specifically to the European Union's individual Member States. We also monitor advertising review rejections, advertising reporting and enforcements, privacy and data protection requests and general community support requests, and have also used this data to support the conclusions reached in this Report.

The goal is to provide insight into our content moderation data, as well as our work with law enforcement and governments, in terms of how we work to keep our users safe. As we produce the report, we recognize shifts in our metrics (e.g., spikes or decreases in content and account reports and enforcements) and utilize these to inform heightened awareness from our moderation teams.



Digital Well-Being Index (DWBI) Initiative

Introduction

In the Spring of 2022, Snap launched a research project designed to gain insight into how Generation Z teens and young adults are faring online. Our inaugural Digital Well-Being Index ([DWBI](#)), a measure of Generation Z's online psychological well-being, was announced on Safer Internet Day 2023. The study asks about the risks and potential harms teens and young adults are encountering online across all platforms and devices, not just Snapchat. We conducted the research in six countries – Australia, France, Germany, India, the UK and the U.S. – and also included parents of teenagers between the ages of 13 and 19. Among other things, parents were asked to “predict” whether their teens had been exposed to online risks or potential harms.

We repeated this research in Year Two (2023), including a deeper-dive into the all-important topic of the online “sextortion” of Teens and young adults. We made results from this portion of the study available early, to coincide with our participation in the Technology Coalition's biennial Multi-Stakeholder Forum, which focused on the financial sextortion of Teens. [Our blog](#) detailing these specific research findings and Snap's relevant platform actions was published on the WeProtect Global Alliance website. The full results of our Year Two study, including the latest Digital Well-Being Index reading, will be released on Safer Internet Day 2024, February 6.

Benefitting all stakeholders

Snap invests in this research to glean insights about the overall online risk landscape, and we seek to share those learnings with other key stakeholders across the ecosystem. Researchers, academics, safety-focused non-governmental organizations (NGOs), governments, law enforcement authorities, parents, caregivers, and the general public, all stand to derive knowledge and intelligence from these findings.

Informing Snapchat features

These and other research findings helped to inform Snap's Family Center suite of tools, released globally in October 2022. Family Center enables parents to view their teens' friend lists and who they have been communicating with over the last seven days, while respecting teens' privacy and autonomy by not disclosing the content of their messages. Family Center also encourages supervising adults to report accounts they may be concerned about, and set content restrictions on suggestive material that may be available on the more public areas of Snapchat. (Content Controls are now turned “ON” by default for new members of Family Center). We plan to leverage these and other research findings and data to further improve Family Center and young people's overall safety experiences on Snapchat.

For more about Snap's Digital Well-Being Index and research, see: Our [blog post](#), [website](#), this [explainer](#), collection of [key research findings](#), the [full research results](#), and each of the six country infographics: [Australia](#), [France](#), [Germany](#), [India](#), [the United Kingdom](#) and [the United States](#).



Safety Advisory Board (SAB)

Snap launched a new Safety Advisory Board (SAB) in April 2022 with the aim of growing and expanding membership to include a diversity of geographies, safety-related disciplines and areas of expertise. In doing so, we initiated an application process, inviting experts and individuals from around the world to formally express their interest in providing guidance and direction to Snap on “all things safety.”

The SAB Board was developed to educate, challenge, raise issues with, and advise Snap on how best to keep the Snapchat community safe and counterbalance the online harms-dominated external landscape. When appropriate, the SAB provides feedback on new products, features, policies, and initiatives before they are launched or released. The SAB and its individual members do not act as a representative or spokesperson for Snap, but rather as a collection of independent voices. It helps to shape Snap’s approach to important safety issues and provides Snap with strategic safety-related advice and guidance as Snap grows.

Our Advisory Board has grown to 19 members, based in 10 countries and representing 11 different geographies and regions, 4 of the Members are EU-focussed. The board comprises 16 professionals from traditional online safety-focused non-profits and related organizations, as well as technologists, academics, researchers, and survivors of online harms. Members are experts in combating significant online safety risks, like child sexual exploitation and abuse and lethal drugs, and have broad experience across a range of safety-related disciplines. In addition, the Board has three members who are young adults and youth advocates. We selected these applicants to ensure the Board has ready access to the all-important “youth voice” and viewpoint; to make certain a portion of the Board includes committed Snapchat users; and to seek to balance professional views with practical perspectives from a core demographic of the Snapchat community.

In addition to the deep investment Snap is making in its SAB cohort, it regularly consults a cadre of some 50 safety experts from around the world on new product features and functionality, policies, and initiatives. This group offered essential feedback in advance of the launch of Family Center in the Spring of 2022, providing significant inputs that were reflected in the final Family Center feature-set and in subsequent updates. Members of this group were also consulted ahead of the addition of Content Controls to Family Center in early 2023.

Snap also conducts periodic internal trainings and learning-sessions, inviting external experts to help inform and educate Snap personnel working in a variety of safety disciplines about the overall risk landscape and Snap’s potential exposure. Collaborators have included the U.S.-based National Center for Missing and Exploited Children (NCMEC) on the topics of child sex trafficking and “self-generated” indecent imagery featuring youth, as well as the non-profit Protect Young Eyes on safety concerns of parents of teens. Snap will continue to invest in these and other



external partnerships and relationships to help bolster internal knowledge and awareness of the overall risk landscape.

Audit

Snap will be working with external auditors to ensure Snapchat's in-scope services are subject to an annual internal audit to assess compliance with Snap's obligations under Chapter 3 of the Digital Services Act.

Snap already has an annual audit process in place, as Snapchat must be subject to an independent third party audit pursuant to Snap's 2014 FTC Consent Decree. This assesses Snap's eight controls related to its data governance program. The preparation that goes into this audit is an ongoing endeavor throughout the year, and ramps up further in the months leading up to the audit. It requires the material time commitments from stakeholders across the company, including Ops, Product, Engineering, Customer Support, Privacy Engineering, Privacy Legal, Product Legal, and Commercial Legal.

Snap hopes to utilize this existing process to ensure an efficient and effective audit pursuant to pursuant to Article 37 of the DSA. The learnings from this audit will inform its DSA compliance programme and form part of its commitment to ongoing DSA compliance improvement.



7. Conclusion

This Report has been prepared to meet Snap's obligation under Article 42(4) of the DSA and sets out the results of: (i) the risk assessment conducted by Snap pursuant to Article 34; and (ii) the review of the specific mitigation measures the Snap has put in place to assess whether they meet the requirements of Article 35(1) DSA.

The risk assessment conducted by Snap identified, analyzed and assessed in accordance with Article 34 DSA any systemic risks in the European Union stemming from the design, functioning or use of Snapchat's in-scope services. Snap has also reviewed the specific mitigation measures that it has put in place to ensure they are "reasonable, proportionate and effective" for the specific systemic risks identified by its risk assessment as required by Article 35(1). The results of the risk assessment and mitigation review are set out in [Section 4 of this Report](#). The specific mitigation measures put in place by Snap are further detailed in [Section 5 of this Report](#) as required by Article 42(4).

The Report shows that we have reasonable, proportionate and effective mitigation measures in place and we continue to monitor a few areas to confirm that if additional measures are required Snap will act accordingly, as follows:

1. **Dissemination of illegal or violating content:** There is a low level of content that is illegal or otherwise violating Snap's [Terms](#) being disseminated on Snapchat's online services in general, compared to the prevalence of this content on websites and other online spaces . Within this low level of dissemination in general:
 - a. We have categorized three dissemination risk areas as falling within Level 1 risk prioritization for Snapchat's in-scope services: (i) child sexual abuse material, (ii) sale of drugs and (iii) credible imminent threats to human life due to risk of severe harm each may cause. We have confirmed we have reasonable, proportionate and effective mitigation measures for all three of these categories. As a result of these measures, all three have been assessed to fall within our extremely low likelihood category of the risks identified by Snap.
 - b. We have categorized five dissemination risk areas as falling within Level 2 risk prioritization for Snapchat's in-scope services: (i) sale of weapons, (ii) terrorism, (iii) adult sexual crimes, (iv) harassment & bullying and (v) glorification of self-harm due to the risk of severe or serious harm each may cause. We have confirmed we have reasonable, proportionate and effective measures in place for all five of these categories. As a result of these measures four have been assessed to fall within our extremely low likelihood category of the risks identified by Snap. We continue to monitor one of these risks, harassment & bullying content which has recently



seen a slight rise in prevalence and reporting, to confirm if this is a temporary issue due to the design of new reporting options, or if additional measures need to be put in place.

Note: Our 2024 Report shows a **significant fall in prevalence data relating to harassment & bullying content** on the in-scope services of Snapchat. While we continue to carefully monitor this content category, we confirmed this was a temporary issue and we have reasonable, proportionate and effective measures.

- c. We have categorized eight dissemination risk areas as falling within Level 3 risk prioritization for Snapchat's in-scope services: (i) illegal hate speech, (ii) sale of other prohibited products or services; (iii) intellectual property infringements, (iv) other adult sexual content, (v) violent or dangerous behavior, (vi) harmful false misinformation, (vii) fraud and spam and (viii) content relating to other illegal activities. We have confirmed we have reasonable, proportionate and effective measures in place for all five of these categories. As a result of these measures six of these have been assessed to fall within our extremely low likelihood category of the risks identified by Snap. The other two (adult sexual content and fraud and spam) fall within low and very low likelihood categories but with declining prevalence. We are monitoring these two categories to confirm prevalence continues to decline and further mitigating measures are not required.

Note: Our 2024 Report shows a **further significant fall in prevalence data relating to fraud and spam and adult sexual content** on the in-scope services of Snapchat. Both now fall within our extremely low likelihood categories. While we continue to carefully monitor this content category, we confirmed we have reasonable, proportionate and effective measures.

2. **Negative effects on EU Fundamental Rights:** We have categorized: (a) three risks to fundamental rights as falling within the Level 1 priority category for Snapchat's in-scope services: (i) human dignity, (ii) data protection and (iii) child rights; (b) one risk as falling within the Level 2 priority category: private life; and three risks as falling within the Level 3 priority category: (i) freedom of expression, (ii) right to non-discrimination and freedom of religion and (iii) right to consumer protection. We have confirmed we have reasonable, proportionate and effective measures in place for all of these categories. We are actively participating in efforts to develop an EU wide AADC to assess if further industry measures are needed to address risks to child rights.



Note: Since completion of our 2023 Report, the Commission has focused on producing Guidelines for Article 28 of the DSA rather than an EU wide AADC. We continue to actively support and participate in the Commission's efforts to establish EU wide guidance for online platforms to assess and mitigate risks to child rights.

- 3. Negative effects on Public Security:** We have categorized three risks to public security within the Level 3 priority category for Snapchat's in-scope services: (i) negative effect on democratic and electoral processes; (ii) negative effect on civil discourse and (iii) negative effect on public security. We have confirmed we have reasonable, proportionate and effective measures in place for all of these categories.
- 4. Negative effects on Public Health:** We have categorized: (a) two risks to public health within the Level 1 priority category for Snapchat's in-scope services: (i) negative effect on children; and (ii) serious negative consequences on physical and mental well being; (b) one risk within the Level 2 priority category for Snapchat's in-scope services: negative effects on gender-based violence; and (c) one risk within the Level 3 priority category for Snapchat's in-scope services: negative effects on public health. We have confirmed we have reasonable, proportionate and effective measures in place for all of these categories. As above, we are actively participating in efforts to develop an EU wide AADC to assess if further industry measures are needed to address risks to children.

Note: Since completion of our 2023 Report, the Commission has focused on producing Guidelines for Article 28 of the DSA rather than an EU wide AADC. We continue to actively support and participate in the Commission's efforts to establish EU wide guidance for online platforms to assess and mitigate risks to children.

As described in the [Ongoing Risk Management](#) section above, Snap continually monitors the effectiveness of its measures.

As well as the actions identified above, work to produce the risk assessments, and this report, has also identified steps we can take to improve our ability to assess risk and the effectiveness of our limitations specifically in respect of the DSA. Examples include: (i) some of the data we have relied on in support of our assessment is currently produced in respect of Snapchat as a whole, rather than in respect of each of Snapchat's in-scope services, and (ii) some risks might be tracked more granularly where there are varying severity levels, such as adult sexual content.

In conclusion, we have carried out a risk assessment of Snapchat's in-scope services and confirmed we have in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified. There are areas that we are monitoring, in particular, to ensure that remains the case.



8. Final Words

This is the first year in which VLOPs have had to produce a report on their assessment of risks and the specific mitigation measures they have put in place. In the absence of guidance, and in keeping with the spirit of the DSA, Snap has decided to take a comprehensive approach to the obligations in Articles 34, 35 and 42. We take the view that all the risks identified in the DSA are systemic to online platforms (which is why they have been identified in the DSA). We then look to ensure we have appropriate platform wide measures in place in general, taking additional steps for specific risks, certain services and high priority risks as necessary.

This reflects Snap's own internal approach to risk management and our core values to be kind, smart and creative. We have always taken the assessment of privacy and safety risks and mitigations seriously, and consider the DSA an opportunity to prove this.

This Report highlights the low risk profile that Snapchat represents, due to its unique design and function, and the efforts that have been taken to further reduce the risks specifically referred to in the DSA. Snap understands, however, that the work of addressing systemic risks on our platform is never done. We look forward to receiving feedback from the Commission as we continue to improve Snapchat for the benefit of Snapchat and our wider community.



Annex - Community Guidelines: Explainer Series

[Sexual Content](#)

Community Guidelines Explainer Series

[Harassment & Bullying](#)

Community Guidelines Explainer Series

[Threats, Violence & Harm](#)

Community Guidelines Explainer Series

[Harmful False or Deceptive Information](#)

Community Guidelines Explainer Series

[Illegal or Regulated Activities](#)

Community Guidelines Explainer Series

[Hateful Content, Terrorism, and Violent Extremism](#)

Community Guidelines Explainer Series

[Severe Harm](#)

Community Guidelines Explainer Series

[Snapchat Content Moderation, Enforcement, and Appeals](#)

Community Guidelines Explainer Series