Snap DSA Report

Risk Assessment Results and Mitigations

August 2024

Contents

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.

Since our 2023 Report, we have made a number of significant updates and these are summarized in a "What is New?" section that we have added to this Report. At a high level, these updates reflect three many themes that we have observed during the last year:

- The popularization of Gen AI - There has been intense interest and concern surrounding the ways in which advancements and access to generative AI technologies are impacting online platforms. Although our Gen AI creation tools are platforms generally not online platforms and are outside the scope of this Report, in the light of this focus, we have nevertheless provided details of our mitigation measures for both creation and dissemination of Gen AI content. For more information, see Section 5.9 Content Authenticity in this Report.

- Elections and democratic process - 2024 has been described as a Global Elections Super Cycle, with a reported 73 different national elections. For the EU, the European Parliamentary Elections in June were a particular focus. We published a blog[1] illustrating the steps taken by Snap to protect European democracy and the integrity of electoral processes, as well as the lessons learnt from the exercise. We have taken our experiences with these elections, as well as the Commission's Guidelines into account when assessing risks to democracy. For more information, see Section 4.3.1 Democracy and Electoral Processes in this Report.

- Minors - We have also seen intense discussion on the role of online platforms in the lives of young people. Over the last year, the world has witnessed CEOs of major technology companies in front of the US senate, the publication of draft codes of practice on

[1] https://newsroom.snap.com/snap-eu-election.

protecting minors under the UK's Online Safety Act and wide ranging debates in particular on screen time and age assurance. In the European Union, a new quantitative study from the Netherlands that compares online platforms on well-being, self-esteem and friendship closeness 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.[2] We have been working proactively to support the Commission's drive for an AADC code under the EU BIK+ strategy and the upcoming guidance for Article 28 of the DSA to provide a clear, harmonized direction to ensure a high level of privacy, safety and security across all online platforms. For more information, see Sections 4.2.6 (Children's Rights), 4.4.3 (Negative effects on Children) and 5.8 (Protection of Minors).

Over the next year, we look forward to continuing our constructive dialogue with the Commission, our Digital Service Coordinator and other stakeholders to further the objectives of the DSA, as well as encouraging continued close collaboration with those cover seeing the General Data Protection Regulation, Audio Visual Media Services Directive and the Digital Markets Act to ensure a holistic approach to privacy, safety and security across these digital platform laws.

[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)

What is New?

Snap is required to complete a report every year setting out the results of its risk assessment and details of its mitigations pursuant to Article 42(4).

Previous reports

Our first report was completed in August 2023 (the "2023 Report"). This was sent to the Commission without undue delay as required by Article 42(4). Snap's reports are published on its European Union Transparency page[3] 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) i.e. approximately 1 year and 3 months after each report is completed.

What is new in this Report?

Section 1 - Introduction

- Snapchat 101 - In all material respects, Snapchat's in-scope services remain the same since our 2023 Report. Our data shows that the vast majority of our users are still primarily using the messaging aspects of our platform, and we continue to believe this is an important lens through which to view Snapchat. **We have not deployed functionalities that were likely to have a critical impact on the risks identified pursuant to Article 34 of the DSA.**

- 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 (ii) 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 services of Snapchat to fall within the scope of our risk assessment and mitigation obligations in Articles 34 and 35. We have confirmed that Snap continues to

[3] https://values.snap.com/privacy/transparency/european-union

consider My AI and other similar generative AI tools made available by Snap to be out of scope of Snapchat's designation except for one advertising case identified below.

- In our short descriptions of the in-scope aspects of Spotlight, Discover, Public Profiles, Snap Map, Lenses, and Advertising services of Snapchat, we have noted in particular that:
    - We have removed reference to Spotlight Challenges in Spotlight as this feature has now been deprecated.
    - We have acknowledged that content created by generative AI tools (whether on third party platforms or Snapchat) could be disseminated via Spotlight and Discover and noted that Snap has taken this into account in this Report.
    - We have no significant updates, Snapchat+ features, or generative AI features relevant to Public Profiles.
    - We have flagged the launch of simplified location sharing and two new Snapchat+ features on the Snap Map, but confirmed that we have assessed these to be out of scope of Snapchat's designation.
    - We have noted that certain Lenses are only available to Snapchat+ subscribers and flagged the launch of GenAI Suite in Lens Studio (but confirmed that we have assessed the latter to be out of scope of Snapchat's designation).
    - We have flagged that our ad creation tools incorporate common, minor tools to generate assets such as text translations and background images. We have assessed that these are likely to be considered in scope.

Section 3 - DSA Risk Assessment Methodology

- DSA Risk Assessment Methodology - We have confirmed that there is no change to our risk assessment methodology since our 2023 Report. We have added an additional paragraph on how we deal with improvements identified in our risk assessment result and mitigation reports, and confirmed that we collate and retain supporting documentation for a minimum of 3 years.

Section 4 - DSA Risk Assessment Results

- We have updated the likelihood, severity, overall potential risk prioritization assessments, mitigation assessments and confirmed that there have been no changes to the conclusions we reached in our 2023 Report that we have reasonable, proportionate and effective mitigation measures for each systemic risk identified in Article 34 of the DSA:
    - Category 1 - Dissemination of content that is illegal or violates our terms and conditions i.e. dissemination of CSAM, Hate Speech, Sale of prohibited Goods, Terrorist content, IP Infringement, Adult Sexual Content, Harassment and Bullying,

Self-Harm, Violent and Dangerous, Harmful False Information, Fraud and Spam, and Other illegal activities
    - Category 2 - Negative Effects on Fundamental EU Rights i.e. Human Dignity, Freedom of Expression and Information, Private Life, Data Protection, Non-discrimination and Freedom of Religion, Children's Rights, Consumer Protection and Property.
    - Category 3 - Negative effect on democratic and electoral processes, civic discourse and public security
    - Category 4 - Negative Effects on Public Health, Gender-based Violence, Protection of Children and Physical and Mental well-being.

- We have seen a substantial reduction in prevalence rates across all of the illegal and other violating content categories that we monitor (see our update on prevalence testing below). As a result, we have been able to lower the relative likelihood of two of our risk categories:
    - Our Adult Sexual Content risk has been lowered from highest relative likelihood to our medium relative likelihood category.
    - Our Fraud and Spam risk has been lowered from medium relative likelihood to our lowest relative likelihood category.

  We had identified both categories as a focus for our ongoing monitoring and management of risk in the conclusion of our 2023 Report. We have noted that we will continue to monitor both categories, but are very pleased with the progress.

- The prevalence of violating content relating to regulated goods risk has also fallen from our medium relative likelihood category to our lowest relative likelihood category, to match the prevalence of 'illegal goods and activities' (which was already in our lowest relative likelihood category). This reinforces our assessment that our Sale of prohibited Goods risk continues to fall within our lowest relative likelihood category.

- In our 2023 Report, we also identified Harassment & bullying as focus for ongoing monitoring as it had recently seen a slight rise in prevalence and reporting. Since then, we have observed a significant reduction in prevalence of this content on the inscope services of Snapchat and confirm that reporting and enforcement rates relating to the inscope services on Snapchat are very low. We have therefore noted that we will continue to monitor this category, but maintain our conclusion that we have reasonable, proportionate and effective measures.

- We have made significant revisions to our assessment relating to the risk of negative effects on democratic and electoral processes to reflect 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.[4] We continue to conclude that we have reasonable, proportionate and effective mitigation measures.

- We have also updated our assessment relating to the risk of negative consequences on physical and mental wellbeing, in particular to take account of a new quantitative study from the Netherlands (currently in pre-print) that compares online platforms on well-being, self-esteem and friendship closeness.[5] 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.

- General - We have also noted in the specific mitigation summary for each risk the following general mitigation updates:
    - Regarding the measure that we have put in place with regards to the Protection of Minors, we have updated each summary to confirm the launch of our new parents site which provides additional guidance for parents and carers on risks and support.[6]
    - Regarding the measures that we have put in place with regards to Content Authenticy, we have updated each summary to confirm that Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.

Section 5 - Specific Mitigations

- Snapchat Design and Function - We have not noted any significant changes to Snapchat's Design and Function mitigations. We have provided additional screenshots relating to our launch of simplified location sharing on Snapchat, although this is out of scope of this Report. We have added additional information on the process and governance around the design of Snapchat's online interface.

- Terms - We have significantly expanded this section to provide additional information regarding the oversight and administration, content, accessibility and readability of our terms mitigations.

[4] Guidelines for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes, April 2024, url.
[5] 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)
[6] https://parents.snapchat.com.

- Transparency - We have not noted any significant changes to our transparency mitigations. We have provided some updates regarding updates to our Privacy Policy, additional safety resources added to our Safety Center, the launch of our dedicated microsite for parents, educators and other caregivers.

- Moderation and Enforcement - We have not noted any significant changes to our moderation and enforcement mitigations. We have made some presentational adjustments to ensure the information provided is located in the right section, as well as updating our statistics, providing more detail on oversight and administration, quality assurance, notification of criminal offenses and complaint handling.

- Algorithmic Systems and Targeted Advertising - We have not noted any significant changes to our algorithmic system and targeted advertising mitigations. We have provided some additional information regarding oversight and administration, model development and deployment, change management and monitoring and quality assurance.

- Protection of Minors - We have provided updates on the mitigation measures we put in place to protect minors, including clarifications regarding our use of inferred age techniques, readability of our terms and conditions, administration and oversight, overview as to how we apply the age appropriate design code, confirmation that we have disabled targeted advertising to minors, updates to the Family Center and other clarifications.

- Content Authenticity - We have made significant updates to this section to highlight how we have accounted for advancements in generative AI technologies in our risk assessments and our approach and measures in place to mitigate these risks both in respect to the use of generative AI in content creation (which we continue to consider to be out of scope of Snap's designation (save for certain commonplace ad creation tools) and are providing for context) and in dissemination to the public on Snapchat's inscope services.

- Trusted flaggers - We have provided additional information regarding Snap's trusted flagger program, as updating our statistics and trend analysis.

- Dispute Settlement Bodies - We have provided minor updates regarding the establishment of out-of-court settlement bodies and our approach, including noting that the Commission has not yet published a list of certified bodies.

- Codes and Crisis Protocols - We have provided updates regarding Snap's engagement with the European Commission and other stakeholders on updates to the EU Hate

Speech Code, and providing a recent case study where we have exercised our content crisis management protocol.

Section 6 - Ongoing Risk Detection & Management

- Platform Principles Framework - There have been no significant updates to our platform principles framework.

- DSA Compliance Team and Cross-Functional Working Groups - We have provided some additional information regarding roles and responsibilities, our compliance officers, independent compliance function, DSA management body, DSA Cross-Functional Governance Team, Points of Contact, Legal Representative and Supervisory Fee.

- Privacy and Safety by Design - We have updated this section in particular to explain how we conduct our critical impact checks, as well as confirm that since our 2023 Report, we have not identified any deployed functionalities that were likely to have a critical impact on our assessment of risks and mitigations pursuant to Articles 34 and 35 of the DSA.

- Prevalence Testing - We are extremely pleased with the progress we have observed from our prevalence testing over the last year, which demonstrates that the effectiveness of our proactive detection mechanisms, agent training and other content moderation and enforcement efforts has continued to increase significantly since our 2023 Report. In particular:

    - We have observed a significant decrease in our overall 'Policy Violating Prevalence' (PVP) rate.
    - All of the most prevalent violating content categories have significantly reduced PVP rates.
    - There are now no illegal or other violating content categories in our low likelihood category (which is a PVP of 0.5% and above).
    - Child Sexual Exploitation and Abuse Imagery (CSEAI) is no longer categorized among the most significant sources of harm

- External Request Monitoring and Review - We have confirmed that we continue to produce transparency reports and monitor advertising review rejections, advertising reporting and enforcements, 'privacy, data protection and DSA' requests and general community support requests, and use this data to support the conclusions reached in this Report. Since the DSA came into force, we have also been monitoring DSA queries raised via our dedicated email address and community support page.

- Digital Well-Being Index (DWBI) Initiative - We have updated this section to report on the results from the Year Three industry wide DWBI research study.

- Snap Advisory Groups - We have provided information on the progress of our work with the Snap Safety Advisory Board and the establishment of a new Snap Council for Digital Well-Being.

- 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, and that we will share the audit report in due course pursuant to pursuant Article 42.4(c) and (d).

Annex - Explainer Series

- Annex - Explainer Series - There have been a number of minor updates to our Explainer Series and references to the new versions are included.

Conclusion

- Conclusion - **We note that we have carried out a risk assessment of Snapchat's in-scope services and continued to confirm that we have in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified.** We have also reflected on the process with the areas for improvement identified in our 2023 Report, as well as identifying new areas for improvement over the coming year.

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.[7]

1.1 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

[7] See our Citizen Snap Report for more details.

generations use text messaging or their phone to stay in touch with friends and family. Since our 2023 Report, our data shows that the vast majority of our users are still 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.[8]

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.** There are a number of significant product changes that we are considering for the future, including in particular:

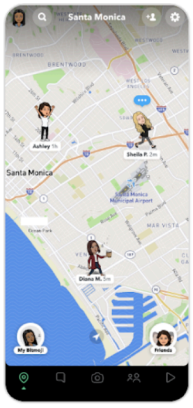
- We are considering simplifying the application from 5 to 3 tabs. The simplified version would still open to the Camera, but would move most friend related content left of the Camera to one tab (Chats, Snaps, Stories, and access point for the Map), and more public content right of the Camera in a unified content feed. We are running tests this year to assess how it performs. This is primarily a cosmetic change, and it should not impact the mitigations we have in place across our application, including content moderation. We have not assessed it to have a critical impact on the risks we have identified in this Report.
- We are also considering a Public Profiles experience that is suitable for 16-17 year olds, which we may start to roll out in select countries later this year. We have provisionally assessed that this change would likely have a critical impact on the risks we have identified in this Report. This experience will contain additional mitigations, on top of the existing Public Profiles mitigations. A roll out in the EU would be dependent on us first finalizing the risk assessment we have been carrying out pursuant to our DSA and other legal obligations.

As these product changes develop further, we would be happy to brief the Commission and provide a more detailed preview.

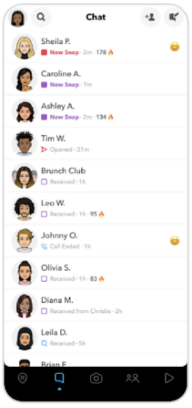
The next section provides a reminder of our platform architecture.

[8] View our Snapchat 101 video for more details.

Platform Architecture

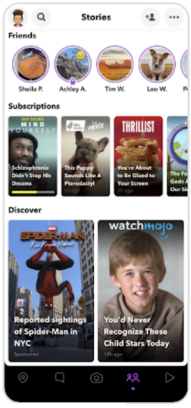


Map



Chat

Camera



Discover

Spotlight

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.

Discover
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.

1.2 Snapchat Community

We reach over 850 million[9] 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 432 million[10] daily active users globally.

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.

[9] Snap Inc. public data Q2 2024, see https://investor.snap.com.
[10] Snap Inc. public data Q2 2024, see https://investor.snap.com.

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. This has remained consistent since our 2023 Report.

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.

2. DSA Risk Assessment Scope

2.1 Approach

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,[11] 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.

[11] url.

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."

2.1.1 Scope in our 2023 Report

Taking account of the above DSA definition and guidance, and that fact that Snapchatters are automatically registered without a human decision or selection of whom to grant access, for our 2023 Report, Snap considered the Spotlight, Discover, Lenses, Public Profiles, Snap Map, and Advertising services of Snapchat to fall within the scope of risk assessment and mitigation obligations in Articles 34 and 35. These services entailed 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.

2.1.2 Scope for this Report

For this Report, (i) the DSA definitions and guidance remain the same and (ii) the functionality of Snapchat has not significantly changed and Snapchatters are still automatically registered without a human decision or selection of whom to grant access. **Therefore, Snap still considers the Spotlight, Discover, Lenses, Public Profiles, Snap Map, and Advertising services of Snapchat to fall within the scope of risk assessment and mitigation obligations in Articles 34 and 35.** These services still 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.

2.2 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, or auto-approve. If accepted or auto-approved, 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 your 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.

Snapchat+ Features relevant to Spotlight
None.

Gen AI Features being used by Spotlight
None. Snap recognises that content created by generative AI tools (whether on third party platforms or Snapchat) could be disseminated via Spotlight. Snap has taken this into account in this Report, including explaining how this activity impacts our risk assessment and what measures we have taken to address risks relating to the dissemination of content created by generative AI.

2.3 Discover

What is Discover?
Discover is part of the 4th tab in the Snapchat app, below your friends' Stories. Contrary to our 2023 Report, Snap ultimately decided not to rename this product from Discover to For You. Hence, any reference to For You that remains in this Report or in supporting documentation should be interpreted as a reference to Discover.

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

How does Discover work?
Discover displays personalized content to users. Discover achieves this using its ranking algorithm, which is described here. The intended purpose of this processing is to personalize Discover 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.

Discover also generates information about how Snapchatters interact with the content in Discover. 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 Discover 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.

Snapchat+ Features relevant to Discover
None.

Gen AI Features being used by Discover
None. Snap recognises that content created by generative AI tools (whether on third party platforms or Snapchat) could be disseminated via Discover. Snap has taken this into account in this Report, including explaining how this activity impacts our risk assessment and what measures we have taken to address risks relating to the dissemination of content created by generative AI.

2.4 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. For more information, see here.

How do Public Profiles work?

Currently, 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 Snappchatter can:

- Add a Photo, Bio, Description, Location, Stories, Spotlights, and Lenses to your Public Profile
- Be Followed by other Snapchatters
- Show their Follower 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.

Snapchat+ Features relevant to Public Profiles
None.

Gen AI Features being used by Public Profiles
None.

2.5 Snap Map

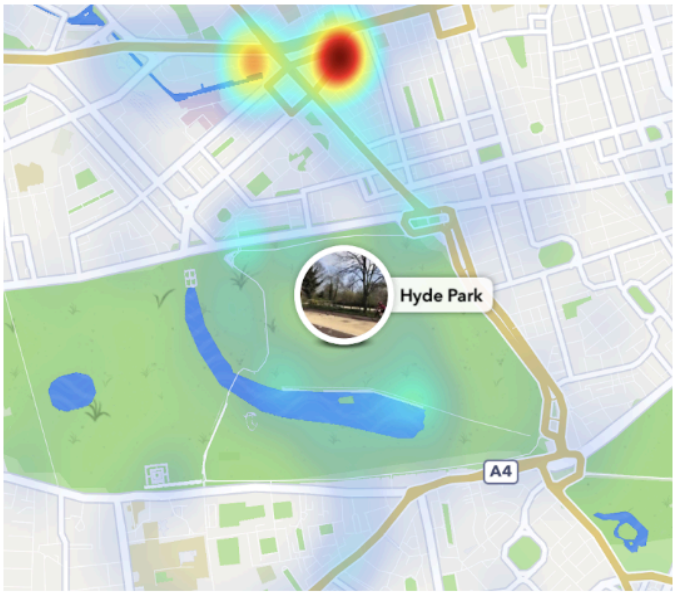
What is Snap Map?
Snap Map is designed to open up a world of possibilities for our community, enabling friends to experience something new in the world every day. Through an interactive map interface, Snap Map shows users what's happening nearby and around the world, anchored by the context of friends' 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

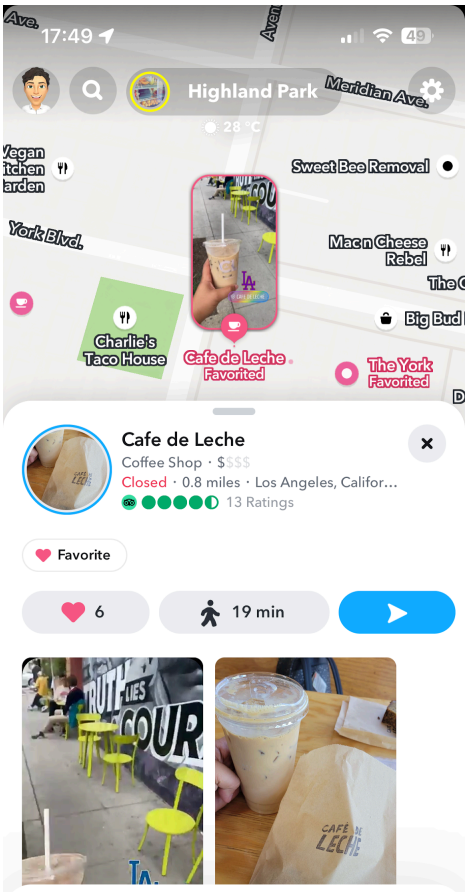
have their Snap shared on Snap Map when it's tagged to a place or venue. Snapchatters can also choose to share their location on the Map with friends while the Snapchat app is actively being used, or share their live location with them even when the app is backgrounded. Since our 2023 Report, we have simplified how location sharing works, which is highlighted in our blog post here. As location sharing is with Friends, rather than the public, and out of scope of Snapchat's designation, we have not detailed this change in the Report. Nevertheless, we have outlined safeguards we have implemented for location sharing in the Mitigations section below (Sections 5.1 and 5.3 specifically).

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

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

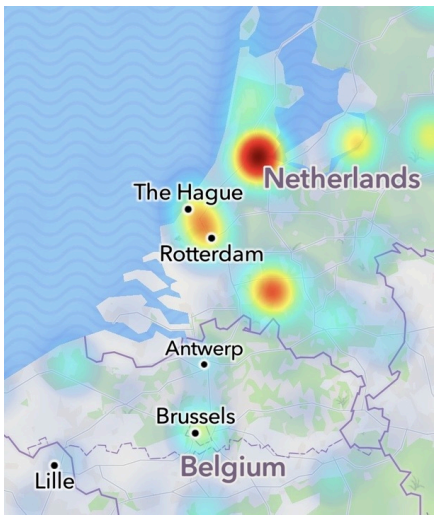


2. **Place Stories** appear on Place profiles. They contain Public Stories snaps explicitly tagged with the place, using either venue filters or place stickers.

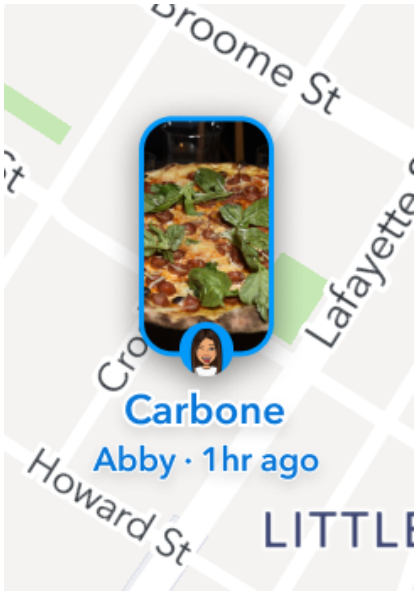


3. **City Stories** appear in the header of the Map and display the best snaps in that locality from the last 7 days. They can appear for cities and neighborhoods.

4. 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.



5. **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.

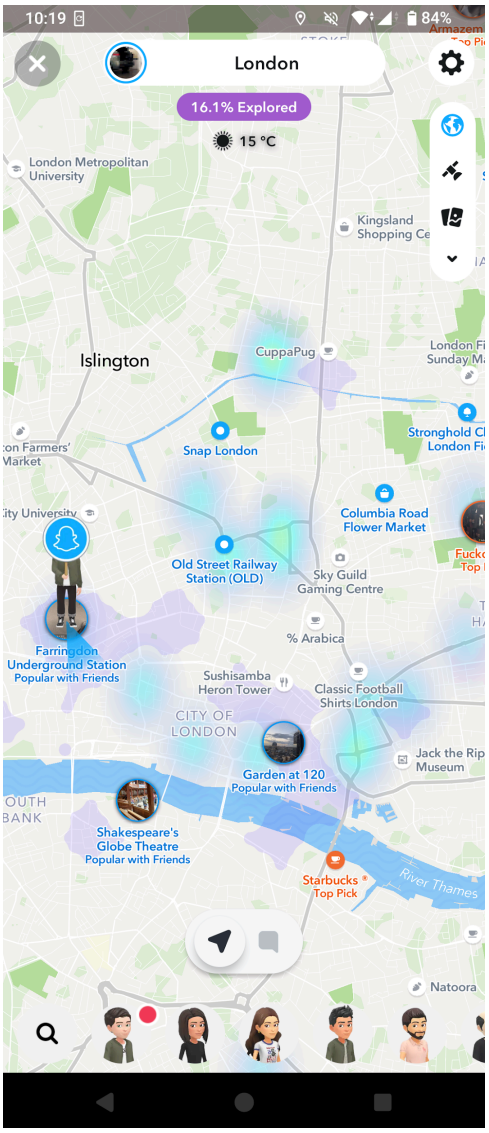


Snap Map submissions may be stored for a while and may be visible on Snapchat for long periods of time as explained to Snapchatters in privacy notices. Snapchatters can remove a Snap they submitted to Snap Map or place-tagged in Spotlight at any time via their profile.

Snapchat+ Features relevant to Snap Map

Customizations - Snap Map offers a couple of customization features for Snapchat+ users including the ability to customize your home on the Map and adding your pet on the Map (either via a preselected group of pets or by creating a generative AI version of your pet).

Footsteps - Snap Map also lets SC+ users see how much of the world they've explored. This content is only made available to the user, and we have assessed it to be out of scope of Snapchat's designation.



Gen AI Features being used by Snap Map

Gen AI Pet - Snap Map offers a Gen AI pet feature as described above. Note however that while this feature is integrated into the Snap Map, the content is only available to Friends with whom the user chooses to share their location. It does not disseminate content to the broader public, and we have assessed it to be out of scope of Snapchat's designation. Nevertheless we have outlined our approach to both Gen AI creation and dissemination in the Mitigations section (in particular Section 5.9 (Content Authenticity)).

2.6 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 fes.

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.

Snapchat+ Features relevant to Lenses

Certain Lenses are only available to Snapchat+ subscribers.

Gen AI Features being used by Lenses
Lens Studio features a GenAI Suite which lets developers take advantage of our generative AI technology to create assets (such as text, effects and backgrounds) for Lenses. Note however that Lens studio is only a creation tool. It does not disseminate content to the broader public, and we have assessed it to be out of scope of Snapchat's designation. Nevertheless we have outlined our approach to both Gen AI creation and dissemination in the Mitigations section (in particular Section 5.9 (Content Authenticity)).

2.7 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.

Snapchat+ Features relevant to Advertising

None

Gen AI Features being used by Advertising

Advertisers can take advantage of generative AI tools during ad creation (such as text translations and background image creation). These are in systems directly related to the dissemination of content (advertising) to the public on Snapchat (although they are common tools and minor in nature). We have taken the use of generative AI tools into account in this Report (in particular Section 4.3.1 (Democracy/Elections) and Section 5.9 (Content Authenticity)).

3. DSA Risk Assessment Methodology

There have been no changes to our risk assessment methodology since our 2023 Report. 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 safety[12] and the ICO's DPIA template. It is worth noting that, in the UK, Ofcom's recent consultations on its illegal harms and minors codes of practice have proposed a similar methodology in its draft risk assessment guidance documents.[13]

The risk assessment methodology used by Snap 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):** Negative effects on:
   i. Democratic and electoral processes
   ii. Civic discourse
   iii. Public security

[12] EU general risk assessment methodology (Action 5 of Multi-Annual Action Plan for the surveillance of products in the EU (COM(2013)76)), url.
[13] Consultation on online harms, url and consultation on protecting children from online harms, url.

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)

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, 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:

Relative likelihood of risk occurring on Snapchat	*Description*
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.

In order to assess likelihood, Snap uses a mix of internal information (such as Prevalence Testing data or illegal / violating content reporting data or complaint data, input from our safety advisory board and Snap commissioned research) and external information (such as external research, news reports and government and NGO guidance). Where internal information is required, this was obtained from the relevant teams responsible for maintaining that information.

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:

Harm classification industry wide	*Description*
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

Harm classification industry wide	Severe harm industry wide	Level 3	Level 1	Level 1
	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
		Relative likelihood of risk occurring on Snapchat		

3.5 Snap's Mitigations

As a fifth step, Snap considered the mitigation measures that it has taken to address each of the risks identified in the overall potential risk prioritization assessment. 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

3.6 Conclusions

As a final step, Snap confirmed whether the mitigation measures it has taken were reasonable, proportionate and effective for each risk identified. To determine this, Snap considered if the mitigations it has in place were effective to address the risk, given its overall potential risk prioritization category, by considering available evidence from its Prevalence Testing data or illegal / violating content reporting data or complaint data, input from our safety advisory board

and Snap commissioned research) and external information (such as external research, news reports and government and NGO guidance).

Where there was evidence that the existing measures risks may need some improvement to ensure reasonable, proportionate and effective measures had been taken, Snap identified this in its conclusion and explained what steps it would be taking to achieve the improvement.

3.7 Supporting Documentation

The data and documentation supporting the risk and mitigation assessment report were preserved and will be retained for a minimum of 3 years.

4. DSA Risk Assessment Results

In this Section of the 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.

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

4.1.1 Dissemination of child sexual abuse material	Extremely low Likelihood	Severe harm industry wide	Level 1	Low Risk / Reasonable, proportionate and effective mitigations
4.1.2 Dissemination of illegal hate speech	Extremely Low Likelihood	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
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)	Extremely Low Likelihood	Severe harm industry wide (Drugs)	Level 1 (Drugs)	Low Risk / Reasonable, proportionate and effective mitigations
	Extremely Low Likelihood	Serious harm industry wide (Weapons)	Level 2 (Weapons)	Low Risk / Reasonable, proportionate and effective mitigations
	Extremely Low Likelihood	Significant harm industry wide (Other goods)	Level 3 (Other goods)	Low Risk / Reasonable, proportionate and effective mitigations
4.1.4 Dissemination of terrorist content	Extremely Low Likelihood	Serious harm industry wide	Level 2	Low Risk / Reasonable, proportionate and effective mitigations with monitoring due to a very slight increase in prevalence to ensure this remains very low.
4.1.5 Dissemination of content that infringes on intellectual property rights	Extremely Low Likelihood	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
4.1.6 Dissemination of adult sexual content	Extremely Low Likelihood (Adult sexual crimes)	Serious harm industry wide (Adult sexual crimes)	Level 2 (Adult Sexual Crimes)	Low Risk / Reasonable, proportionate and effective mitigations
	Very Low Likelihood (Other adult sexual content)	Significant harm industry wide (Other adult sexual content)	Level 3 (Other adult sexual content)	Reasonable, proportionate and effective mitigations, which are being monitored to confirm prevalence continues to decline and further measures are not required.
4.1.7 Dissemination of content regarding harassment & bullying	Very Low Likelihood	Serious harm industry wide	Level 2	Low Risk / Reasonable, proportionate and effective mitigations, with ongoing monitoring.

4.1.8 Dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders	Extremely Low Likelihood	Serious harm industry wide	Level 2	Low Risk / Reasonable, proportionate and effective mitigations
4.1.9 Dissemination of content encouraging or engaging in violent or dangerous behavior	Extremely Low Likelihood	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
4.1.10 Dissemination of harmful false misinformation	Extremely Low Likelihood	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
4.1.11 Dissemination of fraud and spam	Low Likelihood (Advertising) Extremely Low Likelihood (Content)	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations, which are being monitored to confirm prevalence continues to decline and further measures are not required.
4.1.12 Dissemination of information related to other illegal activities	Extremely Low Likelihood	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations, which we will keep under review.

4.1.1 Dissemination of child sexual abuse material

Snap has recognised the risk 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 / Discover 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.[14] 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

[14] See Section 6.4 (Prevalence Testing).

(even though all are low in absolute terms). In our 2023 Report, we noted that CSEAI content represented an extremely low percentage of total views of Snaps in Public Stories. We also noted that 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.

In February 2024 by prominent Finnish NGO Protect Children published a study on the use of tech platforms by online child sexual abuse offenders, funded by the Tech Coalition and Safe Online.[15] Snapchat ranks last among the social media platforms used to search, view, and share CSAM (10%), compared to Instagram (29%), Twitter/X (26%), Discord (23%)TikTok (21%), Facebook (20%), Youtube (18&), Reddit (17%).

As at 30 July 2024 and as a result of our continued focus, the percentage of CSEAI violating views has seen a further, substantial fall to an extremely low rate**.** The steps Snap has taken to mitigate this risk have diminished the likelihood that Snapchatters will encounter CSEAI on Snapchat's in-scope services. Although Snap is aware of concerns regarding the dissemination of CSEAI created using generative AI tools in the wider industry, Snap has not identified this as being a material issue on Snapchat's in-scope services.

As a result, Snap continues to place CSEAI into the Extremely low Likelihood.

Severity

For this Report, as in our 2023 Report, Snap still considers all CSEAI to have a risk of the '**severe harm' category** To this end, Snap has assessed information published by governments and other third party sources.

Overall potential risk prioritization

Although the prevalence of CSEAI on Snapchat has continued to decline and is now at extremely low levels (and 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 and the continued growth in the prevalence of this issue online in general, Snap still considers CSEAI to be a **Level 1 risk priority**. There is no change in this assessment from our 2023 Report.

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

[15] Tech Platforms Used by Online Child Sexual Abuse Offenders, February 2024, url.

Snap prohibits any activity that involves sexual exploitation or abuse of a minor, including sharing CSEAI, grooming, or sexual extortion (sextortion). 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 respond with swift and strict consequences 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 our platform, the content is 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 knowledge of any of the following activity, Snap will take enforcement action and report the user's account to NCMEC:

Snap works with 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 harms 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. There are also industry wide initiatives such as the Tech Coalition's Lantern Program which was launched on 10 November 2023.[16]

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.

[16] https://www.technologycoalition.org/newsroom/announcing-lantern.

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 given the risk of severe harm. Our median turnaround time for child sexual exploitation reports in the second half of 2023 was 52 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 CSEAI.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend CSEAI i.e. there is no 'CSEAI' interest category.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage CSEAI-related risk.
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.	Yes, we cooperate with trusted flaggers in relation to CSEAI/child safety.
Codes and Crisis Protocols	Yes, we cooperate with other providers through various groups e.g. EUIF, the

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.	Technology Coalition, WeProtect Global Alliance.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), and how to get help in our Safety Center.
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.	Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting. **Our new parents site provides additional guidance for parents and carers on risks and support.**[17]
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.	**We recognise there is growing concern regarding use of generative AI tools for CSEAI in general online.** **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

Conclusion

Given the severity of the harm industry-wide, Snap still treats CSEAI as a Level 1 risk priority In response to which 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 continued to help us to improve this proactive content moderation. As a result, we've seen the prevalence of CSEAI on Snapchat fall to an extremely low level. In addition, while we are alive to the risk, we have not identified any material issue with Snapchat's in-scope services being used for the dissemination of generative CSEAI materials.

[17] https://parents.snapchat.com.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for the dissemination of CSEAI. There is no change in this conclusion from our 2023 Report.

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 / Discover, 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.

In our 2023 Report, we highlighted that 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.[18] 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).

In this Report, we are pleased to confirm that our prevalence testing has shown that the PVP for Hate Speech has continued to fall dramatically and **is now at an extremely low level.** The steps Snap has taken to mitigate this harm mean that it is very unlikely that Snapchatters will encounter hate speech on Snapchat's in-scope services, and Snap continues to place Hate Speech in the **extremely low likelihood** category**.**

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.

[18] ARCOM, '*Lutte contre la diffusion de contenus haineux en ligne',* July 2023, url.

Overall potential risk prioritization

As Snap continues to qualify hate speech as 'significant' in terms of severity but 'lowest' in likelihood given the lowest relative prevalence on the platform, Snap considers hate speech a **Level 3 risk prioritization.** There is no change in this assessment from our 2023 Report.

Snap's Mitigations

Highlights

Hate speech is strictly prohibited on Snapchat, and we continue to 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 hate speech content.

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 Discover, 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 Discover 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 we become 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 2023 was **46 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 illegal hate speech reports in the second half of 2023 was 46 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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage illegal hate

their activities in particular as regards detection of systemic risk.	speech.
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.	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.
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.	Yes, we cooperate with other providers through various groups in relation to illegal hate speech. Snap remains a signatory of the EU Code of Conduct to counter illegal hate speech online and has worked hard to ensure Snap meets the requirements (including with respect to recent revision of that Code).
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on harms (see also Annex) and how to get help in our Safety Center. We make available robust reporting; and we provide guidance to parents on the web.
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.	Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center, reporting and guidance. **Our new parents site provides additional guidance for parents and carers on risks and support.**[19]
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.	**Yes, Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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

[19] https://parents.snapchat.com.

code of practice (including the recent revision) 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 risks.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for the dissemination of illegal hate speech. There is no change in this conclusion from our 2023 Report.

4.1.3 Dissemination of information related to the sale of prohibited products or services

The dissemination of information related to the sale of prohibited products or services (such as dangerous products, counterfeit products or illegally-traded animals) is a pervasive challenge for digital platforms.On Snapchat, without mitigation, 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 / Discover and information about places to facilitate sales on Snap Map. It might also include ads promoting the sale of illegal goods, e.g. drugs or malicious content/malware.

Likelihood

Our testing shows that the prevalence of information related to the sale of prohibited products or services on Snapchat continues to decrease year on year. According to our analysis, in July 2024, 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 a extremely low rate,see our Prevalence section. This is a further, very significant decrease from that we reported for August 2023 in our 2023 Report. The likelihood of encountering such violating content on Snapchat's in-scope services is now at its lowest level and falls within the extremely low likelihood category.

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 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 dissemination of illegal or otherwise violating content that relates to the use of online platforms for selling other illegal goods or services

Overall potential risk prioritization

Thankfully, as reported in our 2023 Report, this is still not a common issue on Snap. Our prevalence testing revealed that communication around Illegal goods and activities now has only a very small PVP rate (. 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. There is no change in this assessment since our 2023 Report.

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.

Across Snapchat, we offer a number of resources to users to raise awareness on safety topics and protect them. As explained in the Transparency section of the Report, one of the examples of this is our in-app tool, Heads Up. This surfaces educational content from experts to Snapchatters if they try to search for drug-related content. Our expert partners include the Centers for Disease

Control and Prevention (CDC), the Substance Abuse and Mental Health Services Administration (SAMHSA), Community Anti-Drug Coalitions of America (CADCA), Shatterproof, Truth Initiative, and the SAFE Project.

It is possible, despite Snap's terms and policies prohibiting such practices and offering support, 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 2023, we enforced against more than 365,000 pieces of content and more than 240,000 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 continued to double year on year during the same period.

We also use Prevalence Testing to continuously improve our moderation. We continue to use violating drug Snaps discovered from PVP sampling, which we consider a Level 1 potential overall risk prioritization for illegal goods and services, to help train our proactive machine learning detection models. As a result of this and other efforts since our 2023 Report, proactive detection and enforcement volumes for violating drug Snaps has increased.
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. Our median turnaround time for violating content relating to illegal and other regulated goods were all less than 60 mins in the second half of 2023 (32 minutes for regulated goods, 35 mins for weapons and 55 mins for drugs).
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 presentation of advertisements in association with the service they provide.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage information related to the sales of prohibited products and services.

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.	Yes, we cooperate with trusted flaggers in relation to the sale of prohibited products or services, in particular the Danish Safety Technology Authority.
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.	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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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. Our in-app tool, Heads Up, surfaces educational content from experts and we try to flag this resource to Snapchatters if they search for drug-related content.
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.	Yes, we have protective measures to limit Teen contact with strangers, Family Center, reporting, and guidance. **Our new parents site provides additional guidance for parents and carers on risks and support.**[20]
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

Despite the continued very low prevalence, we still 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.

[20] https://parents.snapchat.com.

Snap continues to consider 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 continues to take steps to mitigate these harms, which has further 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 are still looking to achieve further reductions in the likelihood of this risk where possible.

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. There is no change in this conclusion since our 2023 Report.

4.1.4 Dissemination of terrorist content

As noted in our 2023 Report, 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.[21] 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 / Discover 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 – see our Prevalence chapter. This is a slight increase from the prevalence rate we observed in our 2023 Report (although still extremely low). In the second half of 2023, for example, we removed 320 accounts in the EU for violations of our policy prohibiting terrorist and violent extremist content, as recorded in our Transparency Reports. This is also a slight increase compared with 23 for the second half of 2022, which we observed in our 2023 Report. These increases are consistent with the slight increase in terrorism related content we observed following the attacks in Israel on 7 October 2023 and the subsequent conflict in Gaza - Israel.

As reported in our 2023 Report, we have previously 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. More recently, in response to the various events influencing violent extremism and terrorism online through 2023, about 349 removal orders were sent by 6 EU member States' competent

[21] J.F. Bender and J. Kenyon, *Terrorism and the internet: How dangerous is online radicalization?,* Front. Psychol., 13 October 2022, url.

authorities to 13 online platforms under the Regulation on dissemination of terrorism content online;[22] none were directed at Snap.[23]

These data indicate that, while there has been a slight increase in the likelihood of encountering terrorist content on Snapchat, it remains extremely low and it still falls within the **lowest likelihood category.**

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, Snap has decided to deviate from the standard risk framework, and has placed terrorist content within our **Level 2 overall potential risk prioritization category** (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) There is no change in this assessment from our 2023 Report.

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

[22] Regulation (EU) 2021/784.
[23] Violent Extremism and Terrorism Online in 2023: The Year in Review, April 2024, url.

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 Discover, 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. Our median turnaround time for Terrorism & Violent Extremism reports in the second half of 2023 was 45 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 terrorist content and accounts.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend terrorism content i.e. there is no 'terrorism' interest category.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	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 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.
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.	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.

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.	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 the primary multi stakeholder organization: The Global Internet Forum to Counter Terrorism (GIFCT).
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on harms (see also Annex) and how to get help in our Safety Center. We make available robust reporting; and we provide guidance to parents on the web.
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.	Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center, reporting and guidance. **Our new parents site provides additional guidance for parents and carers on risks and support.**[24]
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.	**Yes, Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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 prevalence and enforcement rates for terrorist content have experienced slight increases since our 2023 Report but there continues to be a very low likelihood of users being exposed to this illegal content.

We have concluded therefore that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for dissemination of terrorist content. However, due to the slight increase in prevalence and enforcement rates since our 2023

[24] https://parents.snapchat.com.

Report, consistent with a slight risk in terrorism related activity we observed following the 7 October attacks in Israel, we continue to carefully monitor this risk category to ensure prevalence remains very low.

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 Discover.

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 2023:

- We received 1,159 copyright notices; 49% of those requests led to the removal of some content. This compares with 558 copyright notices - 78% of those requests leading to the removal of some content - in H1 2022.
- We received 224 trademark notices; 28% of those requests led to the removal of some content. This compares with 96 trademark notices - 29% of those requests leading to the removal of some content - in H1 2022.

In H2 of 2023:

- We received 1297 copyright notices; 57% of those requests led to the removal of some content. This compares with 905 copyright notices - 73% of those requests leading to the removal of some content - in H2 2022.
- We received 203 trademark notices; 23% of those requests led to the removal of some content. This compares with 172 trademark notices - 13% of those requests led to the removal of some content - in H2 2022.

This data continues to show slight overall increases in reports of intellectual property issues year on year but a consistently low prevalence in absolute terms. As a result, Snap continues to consider the likelihood of encountering content that infringes intellectual property on Snapchat is within the **extremely low likelihood category.**

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. There is no change in this assessment since our 2023 Report.

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 Adapting the design, features or functioning of their services, including their online interfaces.	Yes, content on Snapchat is typically short in nature, the average Snap is 10 seconds, and reporting tools help with the detection of IP infringing material.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, terms prohibit IP infringements and they are strictly enforced.
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 reactive moderation procedures to expeditiously remove content that infringes intellectual property rights.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, we have a notice procedure to flag and enable us to respond to intellectual property infringements.
Trusted Flaggers	General trusted flagger measures.

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.	
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.	Not applicable. We respond to reports of infringement on an individual basis.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we warn users not to publish content that infringes on intellectual property rights 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.	**Yes, Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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. There is no change in this conclusion since our 2023 Report.

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 content[25]. 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, Discover, promoted on Public Profiles, features as part of our Lenses or as places on the Snap Map and be the subject of advertisements via Snap Ads.

Likelihood

As shown in our Prevalence Testing, prevalence of adult sexual content continues to be on a downward trajectory confirming that ongoing detection and enforcement enhancements have been having the desired impact.

Although adult sexual content is still our most prevalent illegal or violating content compared with the other risks on Snapchat's in-scope services, we are very pleased to have been able to achieve significant further reductions in prevalence since our 2023 Report. The prevalence is significantly lower than the prevalence of adult content on the Internet in general.

Further, as reported in our Transparency Reports, Sexually Explicit Content has seen a steady reduction in the % of the total content enforced by Snap:

Period	% of the Total Content Enforced by Snap
1H 2022	76.6%
2H 2022	67.9%
1H 2023	52.6%
2H 2023	42.1%

As a result of these substantial reductions in both PVP and proportion of enforced content, in this report we are now placing adult sexual content in our **Very Low likelihood** category. This is a

[25] 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.

decrease from our 2023 Report when adult sexual content was placed in the low likelihood category. We continue not to identify any significant volumes of adult sexual crimes and so we still place this in our **extremely low likelihood** category.

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** risk on Snapchat. We consider the dissemination of sexually explicit content or depictions of nudity to be a **Level 3** potential overall risk prioritization. There is no change in this assessment since our 2023 Report.

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 adult 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.

Our Prevalence Testing has continued to have a very significant impact in reducing the extent to which adult sexual content is present on Snapchat to low levels. We have seen further substantial decreases in the prevalence of adult sexual content since our 2023 Report confirming that ongoing detection and enforcement enhancements have been having the desired impact. This has resulted in a further reduction in prevalence.

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.

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 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. Our median turnaround time for adult sexual content reports in the second half of 2023 was 7 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 adult sexual content.
Algorithmic Systems	Yes, our algorithmic systems do not knowingly recommend adult sexual content

Testing and adapting their algorithmic systems, including their recommender systems.	i.e. there is no 'adult sexual content' interest category.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, for example we have specific prevalence testing and transparency reporting which we use to help detect and manage adult sexual content.
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.	Yes, we cooperate with trusted flaggers in relation to Non-Consensual Intimate Image Abuse (NCII), notably Stop Fisha in France.
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.	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).
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on harms (see also Annex) and how to get help in our Safety Center. We make available robust reporting mechanisms; and we provide guidance to parents on the web.
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.	Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center, reporting and guidance. **Our new parents site provides additional guidance for parents and carers on risks and support.**[26]
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	**Yes, Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are**

[26] https://parents.snapchat.com.

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.	**disseminated on Snapchat's inscope services.**

Conclusion

Adult sexual content is our highest prevalence, largest enforced content category issue on Snapchat's in-scope services, and represents a Level 3 overall potential risk. We also treat sexual crimes as a Level 2 risk given the risk of serious harm, despite Snap considering this to fall within the extremely low likelihood category. We have continued to dedicate substantial resources and taken significant steps, including further improvements to our proactive detection mechanisms. This has resulted in further substantial decreases in the prevalence rates for adult sexual content. 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. We continue to work towards further reductions in the prevalence and enforced content percentages for adult sexual content.

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. We will continue to carefully monitor this risk category as we look to achieve further reductions 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

In our 2023 Report, we concluded that we would continue to monitor the prevalence of harassment & bullying content as we had seen a rise in prevalence and reporting. In particular, as we observed a slight increase in PVP for harassment (which includes bullying). At that time harassment was in the very low risk category observed on Snapchat's in-scope services and we wished to determine whether the rise in prevalence and reporting was a temporary issue due to the design of new reporting options, or if additional measures need to be put in place.

We have since made reporting flow enhancements that simplified Snapchatters ability to report harassment & bullying content and have also expanded the scope of reporting. We are pleased

to say that, while harassment remains within the very low category, since our 2023 Report we have observed a significant reduction in PVP.

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. As a result we were investigating the cause of this but at the time of our 2023 Report we believed this was due to “harassment and bullying” being the first reporting option in the reporting menu. In addition, we noted that we received a significant number of reports of “harassment” in bad faith i.e. 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.

The above-discussed trends continued through the second half of 2023, in which we received even more reports related to harassment in the EU: 1,371,668 reports, with only 16% of those resulting in some enforcement action. We believe that the sharp increase in the number of total reports continued to be due to the wider use and rollout of new reporting options, as well as the issues mentioned in the preceding paragraph. Additionally, we believe that the Israel-Hamas conflict, and resulting anti-Israel and anti-Hamas sentiment, may have had an impact as well. We adjusted our reporting menu to place “harassment and bullying” as the second (and not first) option on the reporting menu (second to “I just don’t like it”); however, this change does not seem to have had much of an impact on report volumes.

While we continue to see significant increases in the content reported and enforced as harassment & bullying content, as in our 2023 Report, we would flag that the bulk of true harassment occurs in non-public surfaces **which are out of scope of our DSA risk assessment.** 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 bullying & harassment content is even lower. In 2023, we rejected 1903 Snaps on Spotlight and 3065 Snaps on Discover. We have seen

a significant fall in the prevalence of harassment & bullying content from our testing of the in-scope services of Snapchat. As a result, in respect of Snapchat's online platforms within the scope of our DSA risk assessment only, we consider harassment & bullying to still fall within our **Very Low likelihood category**.

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 involve 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) in general, would 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 categoryHowever, 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. There is no change in this assessment since our 2023 Report.

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 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 **7 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.

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 presentation of advertisements in association with the service they provide.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage harassment & bullying risk. Our Safety Advisor Board also has several anti-bully experts which we call on for independent review and expertise.
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.	Yes, we cooperate with many trusted flaggers in the EU in relation to child safety, including for example E-Enfance in France.
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.	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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
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.	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.

	Our new parents site provides additional guidance for parents and carers on risks and support.[27]
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. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.** This includes applying an AI sparkle icon in specific situations, such as our Bitmoji Backgrounds. We continue to assess whether to include such an icon 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. Where there is a risk of severe harm, we consider bullying and harassment has a Level 1 overall potential risk. In general, we consider the dissemination of content on Snapchat's in-scope services that includes harassment & bullying is a Level 2 potential risk on Snapchat. 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 continued to result in a significant rise in reports combined with a fall in the enforcement rate. We are still investigating the reasons for this but believe these continue to relate to the out of scope services on Snapchat. Our prevalence testing has shown significant further reductions in the PVP rate for harassment and bullying for **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 harassment and bullying content. While we have seen a significant fall in the prevalence of harassment & bullying content from our testing of the in-scope services of Snapchat, we will continue to carefully monitor this risk category as we look to achieve further reductions.

4.1.8 Dissemination of content that glorifies self-harm, including suicide

The risk of young people encountering content that promotes glorifies self-harm, including the promotion of self-injury, suicide, eating disorders, body image dissatisfaction and distorted values

[27] https://parents.snapchat.com.

and attitudes online and on social media in general has been identified in several studies. Without mitigation, this content could conceivably appear on Snapchat's in-scope services.

Likelihood

In our 2023 Report, we concluded that on this in-scope services of Snapchat, the dissemination of self-harm and suicide content fell within our Extremely Low likehood category due to the low rates prevalance and reporting.

Since our 2023 Report:

- Snap has assessed information published by governments and other third party sources relating to the relative likelihood of self-harm content on online platforms, including Snapchat.
- Our Transparency Report continues to show that "self harm & suicide" is an issue that still leads to a moderate volume of content and account enforcements. In the second half of 2023, we received 188,124 reports related to self-harm and suicide (up from 129,785 reports in the second half of 2022), enforcing against approximately 32,841 pieces of content (up from 20,054 in 2H 2022) and 28,207 accounts (up from 18,311 in 2H 2022). However, as we noted in the 2023 Report, 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 very low. **In 2023, we rejected 4248 Snaps on Spotlight and 110 Snaps on Discover.**
- Our prevalence testing continue to show very low prevalence rates for self-harm and suicide content. This continues the trend we highlight in our 2023 Report where we observed a significant decrease to negligible rates in August 2023. Very few 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.

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 risk prioritization overall**. 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. There is no change in this assessment since our 2023 Report.

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. Since our 2023 Report, we have continued to make improvements to our proactive detection tools, including with respect to self harm and suicide content, which has likely contributed to the substantial reduction in prevelance of public Stories on Snapchat's inscope services.

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 **44 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?
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 for content that promotes self-harm.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	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 44 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	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.

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.	
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage self-harm and suicide.
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.	Yes, we cooperate with trusted flaggers in relation to content that promotes self-harm.
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.	Yes, we cooperate with other providers through various industry groups.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
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.	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. **Our new parents site provides additional guidance for parents and**

	carers on risks and support.[28]
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. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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, 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. 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 less than one hour. Our prevalence rates for self-harm content on Snap's in-scope services continue to decline and are at very low levels.

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. There is no change in this conclusion since our 2023 Report.

4.1.9 Dissemination of content relating to violent or dangerous behavior

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

[28] https://parents.snapchat.com.

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).

In our 2023 Report, we observed that our prevalence testing showed that violent or disturbing content accounted for an extremely low percentage of PVP in August 2023. . A very low percentage of all Lenses submitted were found to include violent or dangerous behavior, which were all rejected.

For this Report, it is worth noting that in Q3 2023, our ongoing monitoring identified an uptick in the prevalence of violating content views for the violent and disturbing category. . In response, we launched new proactive detection mechanisms to target violent and disturbing content. Immediately after launch, the team were able to enforce significantly more proactively detected Snaps daily. Consequently, we have reduced the prevalence of this category and we expect this to decrease further
As a result, we are pleased to report that prevalence for violent or disturbing content (as well as for Dangerous Activities) has seen a further substantial fall in PVP to an extremely low category . This is a result of our specific efforts to reduce exposure to illegal and other violating content falling within our 'Violent and disturbing content' category. Our proactive content moderation has successfully evolved to reduce the prevalence of violent and disturbing content.

On Snapchat, our Transparency Report shows that Threats & Violence still account for a relatively modest amount of all enforcement actions we took across all categories: 2.1% of all enforcement actions in the second half of 2023, compared with 2.6% in the second half of 2022. We received a slightly greater number of reports related to Threats & Violence: 836,125 reports in the second half of 2023, compared with 753,467 in the second half of 2022. We saw a slight decrease in confirmed violations, with action against 114,315 of content and 83,743 accounts in the second half of 2023 (compared with action against 167,811 pieces of content and 132,915 accounts in the second half of 2022).

As a result of the low prevalence from our testing, and relatively consistent levels of reporting and enforcement, we continue to place 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. We continue to note that all of the risks we track on Snapchat have a relatively low prevalence compared to the prevalence of these issues elsewhere online and offline.

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, Snap considers content encouraging or engaging in violent or dangerous behavior to fall within 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 continue to consider falls within our **Level 1** overall potential risk (in deviation from our standard risk matrix), to unfortunate or even silly "fails" which we continue to consider falls within our **Level 3 potential overall risk prioritization**. There is no change in this assessment since our 2023 Report.

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 (in particular our Community Guidelines) address the dissemination of content encouraging or engaging in violent or dangerous behavior.

In addition, 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. As well as the content that is prohibited from Snapchat by our Terms above, content that relates to dangerous

challenges that is violating or harmful is also ineligible for promotion or recommendation on the Discover and Spotlight surfaces of Snapchat.

We have explained in Section 5.6 of the Report (Algorithmic Systems) the specific measures we have taken to address risks including prevention of dissemination of content encouraging or engaging in violent or dangerous behaviour (such as dangerous challenges). In particular, we would note that sensitive content distribution is limited on both Spotlight and Discover:

- In Spotlight, we limit the distribution of sensitive content based on the following rules:
    - We do not recommend sensitive content to users under 18 by default
    - We do not recommend sensitive content to new users .
    - For all other users, by default, we ensure the initial video watched in a session is not sensitive and after that we ensure that sensitive content is only shown sparingly.
- In Discover, as in Spotlight, we limit the display of sensitive content for all users. We also do not show sensitive content to users under 18 by default and display of sensitive content can be disabled entirely in the Family Center.

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 **27 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 **27 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.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage content encouraging or engaging in violent or dangerous behavior.
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.	Yes, we cooperate with trusted flaggers who focus on child and digital safety.
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	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.

crisis protocols referred to in Articles 45 and 48 respectively.	
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on harms (including a specific Threats, Violence & Harm explainer) and how to get help in our Safety Center.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[29]
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 **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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. There is no change in this conclusion since our 2023 Report.

[29] https://parents.snapchat.com.

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 still not common on Snapchat. In our 2023 Report we observed that our prevalence testing showed an extremely low prevalence of ‘harmful false information’. We observed a significant further reduction in prevalence.

As explained in our Transparency Reports, False information continues to account for only 0.1% of the total of all content enforced on Snapchat. This figure remained steady throughout both halves of 2022, as well as through the second half of 2023. We track Impersonation separately, and it similarly accounts for a very low percentage of our enforcement actions (0.3% in the second half of 2023, compared with 0.2% in the second half of 2022). Lenses with this type of information are rarely submitted (only 0.0021% of all Lenses included misinformation in our 2023 Report, all of which were rejected).
As a result, we continue to place dissemination of harmful misinformation into our **extremely low likelihood** category relative to other risks.[30]

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. There is no change in this assessment from our 2023 Report.

[30] 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.

Snap's Mitigations

Highlights

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 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, Discover, 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 Discover, 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. 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 **12 minutes** and for an "impersonation" report is **3 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.[31]

During election seasons, such as the recent European elections, 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. Regarding the European elections specifically, we consider that these unfolded in a positive online environment with no major threats overall. This was confirmed by the European Commission and independent observers, who confirmed that they did not observe major online threats. Snap saw a small uptick in reported activity, but did not receive or observe any material incidents or threats. Our moderation and reporting tools worked well and none of the reported pieces of content were verified as misinformation on Snapchat. In the lead up to the elections, Snap attended multiple cross functional stakeholder meetings, including with civil society organizations, regulators, including the European Commission, and other platforms to share information. We believe these stakeholder meetings contributed to the positive outcome, and we look forward to continuing these engagements. Snap sent a push notification to over 50 million users to urge them to vote in the elections, and made AR election lenses available to promote civic engagement. We are proud to have done our part to contribute to the highest observed

[31] 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.

turnout of the last 30 years, with 51.08% of the 357 million eligible citizens participating in the election. We published a blog post detailing our measures and experiences.[32]

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 Adapting their terms and conditions and their enforcement.	Yes, our terms prohibit misinformation. We have a specific Harmful False or Deceptive Information explainer which explains our approach to enforcement. Our Transparency Report shows that the median turnaround time for a "false information" report is **12 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, we use specific proactive and reactive moderation procedures to prevent and remove misinformation. In particular, Discover 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.

[32] https://newsroom.snap.com/snap-eu-election.

Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	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.
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.	Yes, for example we have specific prevalence testing and monitoring moderation and enforcement data which we use to help detect and manage harmful false misinformation.
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.	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.
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.	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. **We have also worked closely with the Commission and other stakeholders during the recent EU elections.**
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental	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

control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.	web. **Our new parents site provides additional guidance for parents and carers on risks and support.**[33]
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.	We recognise the risk that generative AI could be used to generate harmful false misinformation, including deep fakes. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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 **12 minutes** and for "impersonation" is **3 minutes**. We have also observed further significant falls in the prevalence of harmful false information from our prevalence testing.

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. There is no change in this conclusion from our 2023 Report.

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.

[33] https://parents.snapchat.com.

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).

In our 2023 Report:

- We observed from our prevalence testing that Fraud and Spam had a lowprevalence August 2023. We are happy to confirm in this Report that Fraud and Spam continue to be on a downward trajectory confirming that ongoing detection and enforcement enhancements have been having the desired impact resulting in a reduction in prevalence. There was a substantial further decrease in prevalence in July 2024.

- Our Transparency Report shows that "Spam and Fraud" reports continue to lead to a moderate volume of content and account enforcements on our content online platforms, such as Spotlight and Discover. In 2023 we started to see an influx of Spam accounts. Consequently, we significantly increased Fraud and Spam enforcement, which jumped from 657,077 in the second half of 2022 (8.2% of all enforcement actions) to 1,380,341 in the first half of 2023 (22.2% of all enforcement actions). This started to decrease in the second half 2023 where the number of enforcements dropped to 1,002,278 (18.6% of all our enforcement actions). The project was able to successfully complete in December 2023.

In addition, evidence[34] submitted to a UK Home Affairs Committee inquiry[35] into fraud by Alison Thewliss MP (SNP, Glasgow Central) quoted figures from TSB on the prevalence of fraud on different platforms. She stated that 70% of the frauds that TSB was picking up were being perpetrated on Meta—24% on Facebook and 46% on Instagram—**4% on Snapchat** and 23% across other platforms. This is further evidence that indicates that Fraud has low prevalence on Snapchat in general.

Overall, given the further reductions in prevalence of Spam and Fraud content on Spotlight and Discover, we consider the dissemination of fraud and spam in terms of relative likelihood on Snapchat's content platforms to fall within the **Extremely Low likelihood category.** This is a change from our conclusion in our 2023 report where we considered that Spam and Fraud fell within the very low likelihood category.

We observed that fraud was by far the most common reason for advertising on Snapchat to be rejected between 1 July 2022 and 30 June 2023. This trend has continued this year, As a result,

[34] UK House of Commons Home Affairs Committee, Home Affairs Committee. Oral evidence: Fraud, HC 125, Wednesday 22 November 2023, url.
[35] UK House of Commons Home Affairs Committee, Committee Press Release.

we consider dissemination of fraud and spam to be the low likelihood risk for our advertising systems, relative to other risks.

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 risks** 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. There is no change in this assessment since our 2023 Report, save to reflect the fact that Fraud and Spam in relation to content is now in the Lowest Relative Likelihood category.

Snap's Mitigations

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

Snap's Terms address 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 **2 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?
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 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.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	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 **2 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	Yes, we use specific proactive and reactive moderation procedures to prevent and remove misinformation. In particular, Discover 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.

decision-making processes and dedicated resources for content moderation.	
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend fraud or spam, i.e. there is no 'fraud' or 'spam' interest category.
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.	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.
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.	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.
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.	Yes, we cooperate with trusted flaggers, our trusted flaggers may also report fraud spam, but this is not generally the focus of their efforts.
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.	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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[36]
Content Authenticity	Generalcontent authenticity measures. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or**

[36] https://parents.snapchat.com.

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.	**otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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 3potential risk prioritization. We handle significant volumes of enforcement and rejections every month. Our prevalent testing shows this is working, with significant further reductions in the prevalence of fraud and spam on public content surfaces of Snapchat. Ad rejection rates have remained consistent with the overall increase in ads reviewed.

As a result, we have concluded that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures for fraud and spam. We are pleased with the further reductions in the prevalence of fraud and spam on Snapchat's public content surfaces. Snap will continue to monitor this category to ensure the consistently higher proportion of ad rejections for fraud and spam are not an indication that further mitigating measures might be required with regards to ads. This amounts to a slight change of focus from the conclusion in our 2023 Report.

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 continue to 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, as in our 2023 Report, and we are not currently aware of other significant issues.

With the introduction of the Digital Services Act, we introduced a new reporting option to report 'illegal content' in general. We have seen very few reports being made through this reporting option and, when it is used, the quality of the reports are very low (for example, missing key information to be able to identify the content in question and/or the illegal nature of the activity or the report concerned harmless activity) and are usually not actionable. With regards to the very small number of reports that were actionable, almost all of these related to our existing illegal and other violating content categories referred to above. We have not observed any significant new illegal activity categories.

As a result we still believe the dissemination of information related to other illegal activities to fall within the **Extremely Low likelihood category** relative to other risks identified by Snap.

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** i.e. Snap's lowest risk category compared to other risks. 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. There is no change in this assessment since our 2023 Report.

Snap's Mitigations

<u>Highlights</u>
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?
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 for some other illegal activities.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our 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.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and,	Yes, general proactive and reactive moderation procedures to prevent and remove other illegal activities.

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.	
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not knowingly recommend information relating to illegal activity i.e. there is no 'illegal activity' interest category.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	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.
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.	Yes, we cooperate with trusted flaggers who are able to flag other illegal activities.
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.	Yes, we cooperate with other providers through various industry groups on prominent issues facing online platforms.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
Protection of Minors	Yes, we have protective measures to limit Teen contact with strangers; we offer Family Center; we make available robust reporting; and we provide

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.	guidance to parents on the web. **Our new parents site provides additional guidance for parents and carers on risks and support.**[37]
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: Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.

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 compared to other harms and have taken steps to mitigate these risks. We regularly review our risk categories using our Risk Detection and Management processes. We have seen few reports using our new option to report 'other illegal activity' and have seen any new categories emerge as the vast majority are either not actionable or relate to one of our existing categories.

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. Since our 2023 Report, we have monitored DSA enquiries, including to our new reporting option, but have so far not identified any new trends or the need to establish new illegal or harmful content categories.

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

[37] https://parents.snapchat.com.

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")[38]. 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*	*Snap's mitigations*
4.2.1 Right to human dignity	Extremely Low Likelihood	Severe harm industry wide	Level 1	Low Risk / Reasonable, proportionate and effective
4.2.2 Right to freedom of expression	Extremely Low Likelihood	Significant harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective
4.2.3 Right to private life	Extremely Low Likelihood	Serious harm industry wide	Level 2	Low Risk / Reasonable, proportionate and effective
4.2.4 Right to data protection	Low Likelihood	Severe harm industry wide	Level 1	Low Risk / Reasonable, proportionate and effective
4.2.5 Right to non-discrimination and freedom of religion	Extremely Low Likelihood	Serious harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective
4.2.6 Children's rights	Extremely Low Likelihood	Severe harm industry wide	Level 1	Low Risk / Reasonable, proportionate and effective and we are actively participating in efforts to develop an EU wide guidance to assess if further industry

[38] Charter of Fundamental Rights of the European Union, url.

				measures are needed.
4.2.7 Right to consumer protection	Extremely Low Likelihood L	Significant harm industry wide	Level 3	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 mitigation such content could conceivably appear in any of Snapchat's in-scope services displaying user generated content, from videos featured on Spotlight / Discover, to Place Stories on Snap Map. Advertising could, for example, include hate speech or discriminatory elements.
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.

As Snap takes these issues very seriously and has implemented several levers to prevent this content from being distributed on the platform

Likelihood

We have assessed the relative likelihood of Snapchat's inscope services disseminating content that may undermine human dignity, based on Policy Violating Prevalence (PVP) via random sampling and our transparency report data in Section 4.1 of this Report, as follows:

Category	*Relative likelihood of risk occurring on Snapchat*

4.1.1 Dissemination of child sexual abuse material	Extremely Low Likelihood
4.1.2 Dissemination of illegal hate speech	Extremely Low Likelihood
4.1.4 Dissemination of terrorist content	Extremely Low Likelihood
4.1.6 Dissemination of adult sexual crimes	Extremely Low Likelihood
4.1.8 Dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders	Extremely Low Likelihood

We therefore continue to assess that the relative likelihood that the in-scope services of Snapchat would have an actual or foreseeable negative effects of the right to human dignity falls 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 human dignity were to materialise on an online platform, it would fall within our 'severe harm' category.
We have assessed the severity of harm caused by each of these categories of content that significantly undermines the right to human dignity in Section 4.1 of this Report, as follows:

Category	*Relative likelihood of risk occurring on Snapchat*
4.1.1 Dissemination of child sexual abuse material	Severe harm industry wide
4.1.2 Dissemination of illegal hate speech	Significant harm industry wide
4.1.4 Dissemination of terrorist content	Serious harm industry wide
4.1.6 Dissemination of adult sexual crimes	Serious harm industry wide
4.1.8 Dissemination of content that glorifies self-harm, including the promotion of self-injury, suicide or eating disorders	Serious harm industry wide

Our assessment shows a variety of harm ranging from significant to the most severe. We continue to choose to assess the category of actual or foreseeable negative effects of the right to human dignity using the highest severity rating of the categories we have assessed. Given the outsize potential for harm for some of the risks to human dignity, Snap continues to consider this risk to fall within the **severe** risk of harm category.

Overall potential risk prioritization

Although the prevalence of content that negatively impacts users' rights to human dignity has been assessed to be in the lowest likelihood category, we have assessed the severity of this risk to be severe. As a result, overall, we consider the negative effects on the right to human dignity to fall within 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. There is no change in this assessment from our 2023 Report.

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.	Yes, Snapchat's in-scope services have been adapted to include proactive moderation for CSEAI and other illegal content that undermines human dignity. We also have tools within the app where individuals can report this type of activity to our Trust and Safety team. 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.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our Terms prohibit CSEAI and other illegal content that undermines human dignity and they are strictly enforced.
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 CSEAI and other illegal content that undermines human dignity. We have terms in place to prevent Media Partners from publishing illegal or harmful content on Discover. 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.

Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, we have specific prevalence testing and transparency reports for CSEAI, terrorist content, and other illegal content that undermines human dignity.
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.	Yes, we cooperate with trusted flaggers in relation to CSEAI and other illegal content that undermines human dignity.
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.	Yes, we cooperate with other providers through various industry groups e.g. EUIF.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the Annex), as well as 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.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[39]
Content Authenticity	Some content authenticity measures have been taken in respect of content that

[39] https://parents.snapchat.com.

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.	undermines human dignity. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

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 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. There is no change in this conclusion from our 2023 Report.

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 continue to 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 continue to show that in general we receive low incidents of illegal content reports from recipients of Snapchat or authorities where we chose not to take enforcement actionBased on the lack of reporting Snap has received and the overall design of Snapchat (which does not generally provide a platform for political public 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 public 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. There is no change in this assessment from our 2023 Report.

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 public content. Such content is not eligible for promotion on Spotlight and user content on Discover 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 public 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 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 users' rights to freedom of expression and access to accurate information.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, terms prohibit harmful false or deceptive information and they are strictly enforced.
Moderation Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where	Yes, specific proactive and reactive moderation procedures to protect users' rights to freedom of expression and access to accurate information.

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.	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).
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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	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.
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.	No, we do not work with trusted flaggers for users' rights to freedom of expression and access to accurate information.
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.	Yes, our Crisis Protocols handle issues related to users' rights to freedom of expression and access to accurate information. We have recently exercised these protocols successfully during the French riots in June 2023. Note, we will continue to reassess and explore the opportunity to join the EU disinformation code.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.

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.	Yes, we have protective measures to allow Teens to express themself 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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[40]
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.	Yes, **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

Conclusion

Snap continues to consider the overall risk to be within the level 3 risk prioritization 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. There is no change in this conclusion from our 2023 Report.

[40] https://parents.snapchat.com.

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 / Discover and Snap Map. 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.

Likelihood

In our 2023 Report, we explained that:

- Our prevalence testing showed 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**lowest relative likelihood category**.

We have been pleased to observe that this situation has continued:

- Our prevalence testing showed that "invasion of privacy" had seen a further, substantial fall in prevalence. It is now at a very low level.

We therefore continue to assess that the relative likelihood that the in-scope services of Snapchat would have an actual or foreseeable negative effects of the right to private life falls 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. There is no change in this assessment from our 2023 Report.

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?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	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.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, for example, our Community Guidelines prohibit impersonation, our Commercial Content Policy prohibits non-consensual sexual material and our Spotlight Terms require "*you must have any necessary third-party rights including, without limitation, music copyrights and rights of publicity, for all content in your Snaps*".
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 protect the privacy interests of our community. 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".
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not categorize or recommend content that violates users' rights to private life. For example, we have terms, moderation and enforcement to prevent distribution of illegal / violating content. We also do not process sensitive category information.
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.	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.
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.	Yes, we have specific prevalence testing and transparency reports for sexual content and intrusion of privacy. We also monitor privacy-related inquiries as detailed above.

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.	Yes, we cooperate with trusted flaggers in relation to sexual content and Teen safety which may impact users' right to private life.
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.	Yes, we cooperate with other providers through various industry groups e.g CIPL, FPF. Our content moderation policies provide de facto content moderation crisis protocol.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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. Our Privacy Center offers a suite of information on our products, users' choices to safeguard their privacy and how to contact us.
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.	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". Our Family Center includes resources and guidance for Teens and their parents or trusted adults. **Our new parents site provides additional guidance for parents and carers on risks and support.**[41]
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.	Yes, **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

[41] https://parents.snapchat.com.

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. There is no change in this conclusion from our 2023 Report.

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 / Discover 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. In our 2023 Report, we explained that:

- Our prevalence testing showed that "invasion of privacy" made up an Extremely Low PVP in August 2023 (see our Prevalence Testing chapter), Snap also receives low numbers of privacy-related queries

We have been pleased to observe that this situation has continued:

- Our prevalence testing showed that “invasion of privacy” had seen a further, substantial fall in prevalence. It is now at a very low level.

However, despite low numbers of privacy related queries from our users, given the significant volume of personal data being processed by Snap in relation to Snapchat, we continue to assess this risk falls 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 our assessment of the risk of severe harm from negative effects on the right to data protection within the European Union, we have assessed this to be a **Level 1 overall potential risk prioritization**, notwithstanding the low incident of privacy related queries from recipients for Snapchat’s in-scope services. There is no change in this assessment from our 2023 Report.

Snap's Mitigations

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.

Sections 5 and 6 contain many examples of how we designed Snap's inscope services with privacy and safety principles in mind.

As a result of the measures that Snap takes to protect personal data and provide users with actionable tools and transparent information, Snap continues to receive 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?
Snapchat Design and Function Adapting the design, features or functioning of their services, including their online interfaces.	Yes, Snapchat is a platform with strong privacy principles. These principles are reflected into the architecture of our platform. Product changes are subject to privacy by design reviews and we maintain data protection impact assessments. 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 Discover, 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.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our Privacy Center provides a suite of policies, including our Privacy Policy and they are enforced. 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 have any necessary third-party rights including, without limitation, music copyrights and rights of publicity, for all content in your Snaps"* 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. In our Snap Spotlight Submission and Revenue Terms we state "*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*". 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.
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 that violates users' right to data protection. For example, on Discover, Media Partners are proactively moderated and only a small pool of Snapchatters are shown in Discover ("Snap Stars" or "Popular Users".
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not categorize or recommend content that violates users' right to data protection. For example, users can opt out from personalized recommendations based on inferred interest and we do not process sensitive category information.
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.	Yes, our Advertising Systems has a suite of protections including: • No microtargeting • We offer controls to turn off most personalized ads. Users can learn more about their choices here How to Adjust My

	Advertising & Interest Preferences on Snapchat. • 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.
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.	Yes, we consult with experts and our community, and we also monitor and respond to privacy-related inquiries.
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.	No, we don't cooperate with trusted flaggers in relation to data protection violations.
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.	Yes, we cooperate with other providers through various industry groups e.g. CIPL, FPF. We also have a well-established protocol to deal with privacy incidents, as well as a Security Incident Response Policy.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we provide guidance on privacy protection in our Privacy Center. For example, we explain to users How We Rank Content in Discover, How We Rank Content on Spotlight – Snapchat Support and Snapchat Ads Privacy & Transparency.
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.	Yes, we have protective measures to limit disclosure of Teens' data and we avoid nudge techniques to encourage Teens to change their privacy settings and select less privacy-enhancing choices. We offer Family Center; we make available robust reporting; and we provide guidance to parents on the web. **Our new parents site provides additional guidance for parents and carers on risks and**

	support.[42]
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

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. There is no change in this conclusion from our 2023 Report.

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

[42] https://parents.snapchat.com.

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 / Discover 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).

We have assessed the relative likelihood of Snapchat's inscope services disseminating content relating to matters that may undermine the right to non-discrimination and freedom of religion, based on Policy Violating Prevalence (PVP) via random sampling and our transparency report data in Section 4.1 of this Report, as follows:

Category	*Relative likelihood of risk occurring on Snapchat*
4.1.2 Dissemination of illegal hate speech	Extremely Low Likelihood
4.1.4 Dissemination of terrorist content	Extremely Low Likelihood

In addition, although we identified in Section 4.2.4 we placed risks to the right to data protection in the highest relative likelihood category, we have noted overall the number of privacy and data protection related queries we receive are very low. When focused on algorithmic bias specifically, we have not received any material volume of queries, which is not surprising as we noted in Section 4.2.2 (Right to freedom of expression and information) that Snap's inscope services do not generally provide a platform for political public content.

We therefore continue to assess that the relative likelihood that the in-scope services of Snapchat would have an actual or foreseeable negative effects of the right to human dignity 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.**

In addition, we have assessed the severity of harm caused by categories of content that may undermines the right to non-discrimination and freedom of religion in Section 4.1 of this Report, as follows:

Category	*Relative likelihood of risk occurring on Snapchat*
4.1.2 Dissemination of illegal hate speech	Significant harm industry wide
4.1.4 Dissemination of terrorist content	Serious harm industry wide

Overall potential risk prioritization

Snap would consider this risk to the right to non-discrimination and freedom of religion to fall within the **Level 3** category overall. Although we consider the risk to fall within our serious harm category, there is a relatively low prevalence of hate speech, terrorism and bias concerns on our online platform. There is no change in this assessment from our 2023 Report.

Snap's Mitigations

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 Discover, 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.

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

- We survey a subset of our users quarterly to understand whether they find their time spent on our experience entertaining 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 divers 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.

As reported in our 2024 Diversity Annual Report, we know DEI is critical for long-term growth - whether it's Snapchatters demanding products to meet their diverse needs, or the desire to reach new and different markets. It also highlights in particular two new initiatives since our 2023 Report that showcase examples of how empathy can inspire new perspectives and tangible business impact.

- **Snap Out Loud** - Snapchat is a platform that celebrates authenticity. That's why our team created an AR experience to spotlight the different communities who share the LGBTQIA

umbrella. Led by SnapPride, this Snap Show educates Snapchatters about the meaning of LGBTQIA, and celebrates the people who make up the community. Snapchatters were welcomed into seven separate spaces, denoted by the letters of the acronym, to explore each world. 25 million unique users were reached across 11 countries and the lens was shared one million times.

- **8 Mars 8 femmes** - In France, only 10% of statues in public spaces honor female figures. SnapWomen partnered with the Sales and AR Studio team in Paris to launch an AR activation on International Women's Day 2023 across 8 major cities. Called 8 Mars 8 femmes, or 8th of March, 8 women, the activation featured female AR statues next to male ones to celebrate great women in history who were never given appropriate credit for their impact. These AR statues are permanently activated, and honor Josephine Baker, Olympe de Gouges, Manon Tardon, Hubertine Auclert, Simone Veil, Françoise de Graffigny, Élisabeth Vigée Le Brun, and Simone de Beauvoir. The initiative received widespread media coverage in France. Snap's AR Studio team was honored to win the global Drum Award for Marketing for their outstanding creativity.

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 for hateful content or activities supporting terrorism or violent extremism. We also work with civil society organizations to ensure our policies are enforced responsibly. Product Inclusion helps us create equitable experiences by intentionally involving and

	considering marginalized groups at critical moments throughout the product development process.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	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 veteran status, immigration status, socio-economic status, age, weight, or pregnancy status. We strictly enforce these rules.
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 hateful content or activities supporting terrorism or violent extremism. We provide in-app reporting for hateful content or activities supporting terrorism or violent extremism.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not categorize or recommend hateful content or activities supporting terrorism or violent extremism.
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.	Yes, other mitigations listed here also apply to our Advertising Systems. 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).
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.	Yes, we have specific prevalence testing and transparency reporting for hate speech, terrorist and violent extremist content.

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.	Yes, we cooperate with trusted flaggers in relation to illegal hate speech, terrorist and violent extremist content.
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.	Yes, we cooperate with other providers through various industry groups e.g. EU Internet Forum. We have signed onto the EU hate speech Code.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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 in-app reporting for hateful content or activities supporting terrorism or violent extremism.
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.	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. Our Family Center includes resources and guidance for Teens and their parents or trusted adults. **Our new parents site provides additional guidance for parents and carers on risks and support.**[43]
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	General content authenticity measures. **However, Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat’s inscope services.**

[43] https://parents.snapchat.com.

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

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 Discover, we take a proactive approach to moderating any content that may violate these rules. Further, our diversity and inclusion efforts continue to help us create equitable experiences and build inclusive products.

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. There is no change in this conclusion from our 2023 Report.

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 Charter[44] 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.**

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

Likelihood

As explained in Snapchat Community as part of our Introduction to this Report, Snapchat is used by a wide demographic, with 18-24 year olds still 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 still 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 continue to 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 within 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. There is no change in this assessment from our 2023 Report.

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

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 Discover 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	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

processes and dedicated resources for content moderation.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	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.
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.	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.
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.	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. Note, we are actively working to support efforts to agree an EU Age appropriate design code to protect children's rights.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
Protection of Minors Taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping	Yes, we have protective measures to allow Teens to express themself 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

minors signal abuse or obtain support, as appropriate.	reporting; and we provide guidance to parents on the web. **Our new parents site provides additional guidance for parents and carers on risks and support.**[45]
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.	Yes, for example we display an icon in some Lenses that manipulate an image of a Snapchat to make them look younger.

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. There is no change in this conclusion from our 2023 Report.

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

[45] https://parents.snapchat.com.

protection rights poses a serious threat to the rights of EU citizens who may be vulnerable to deception or invasion of privacy.

Likelihood

Snap has implemented safeguards, both through product design and policy enforcement, to effectively diminish the likelihood that consumer protection rights are violated on the platform. We monitor the number and nature of Privacy and Data Protection requests we receive. In our 2023 Report, we explained that:

- Our prevalence testing showed that “invasion of privacy” made up an Extremely Low PVP in August 2023, a decrease from April 2022.
- Snap also receives low numbers of privacy-related queries.

We have been pleased to observe that this situation has continued:

- Our prevalence testing showed that “invasion of privacy” had seen a further, substantial fall. It is now at a very low level.
- Overall we have seen a slight increase in line with expectations and we continue to receive low numbers of privacy-related queries.

We monitor the number of community support requests we receive relating to the European Union. Note, these figures (and the privacy figures above) concern the requests which we review manually and excludes automated responses. We have observed a steady decrease in EU consumer support requests.

Finally, we also monitor for significant changes in our ad review processes.

Given these safeguards we have in place, and the low, decreasing level of consumer queries and consistent levels of ad rejections for our main fraud category, Snap considers that the risk to the right of consumer protection falls within the **Extremely Low Likelihood**.

Severity

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.**

As a result Snap treats risk of harm from a negative impact on the right to consumer protection as **significant** due to the potential harm it can cause to users.

Overall potential risk prioritization

Given we have assessed the potential for negative impacts to consumer protection rights to fall within the lowest likelihood category and to have a risk of significant harm, we consider this risk to fall within our **Level 3 potential risk prioritization category**, given the Extremely Low

Likelihood and significant harm categorization. There is no change in this assessment from our 2023 Report.

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.	Yes, Snapchat's in-scope services have been adapted to include proactive moderation for ads that violate our policies, or false or deceptive content. 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. We also require a specific minimum audience of Snapchatters to prevent advertisers from manipulating small audiences.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, our advertising policies safeguard consumer protection and they are strictly enforced.
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 ads that violate our policies, false or deceptive content. 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.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not categorize or recommend false or deceptive content.
Advertising Systems Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting	Yes, our Advertising Systems are set up to safeguard consumer protection. For example, we automatically place an "Ad" marker on all paid ads.

the presentation of advertisements in association with the service they provide.	
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.	Yes, we have specific prevalence testing and transparency report false information, impersonation, spam and other regulated goods.
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.	Yes, we cooperate with trusted flaggers in relation to regulated goods.
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.	Yes, we cooperate with other providers through various industry groups e.g. CIPL, FPF.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we have an ad marker on all ads and provide transparency on our privacy practices including ads on our Privacy Center.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[46]
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	**Yes, Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.**

[46] https://parents.snapchat.com.

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

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. There is no change in this conclusion from our 2023 Report.

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 continue to 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 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 democratic and electoral processes, civic discourse and 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*	*Conclusions*
4.3.1 Negative effect on Democratic and Electoral Processes	Extremely Low Likelihood	Severe harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
4.3.2 Negative effect on Civil Discourse	Extremely Low Likelihood	Severe harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
4.3.3 Negative effect on Public Security	Extremely Low Likelihood	Serious harm industry wide	Level 3	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

As outlined in Section 1, a significant proportion of European citizens use Snapchat. As at 1 August 2024, we have 92.4 million average monthly active recipients of our Snapchat app in the European Union (EU), and significant recipient numbers in individual Member States. There is the potential for public content on Snapchat to reach a sizable audience within the European Union (particularly within the 18-24 age bracket which accounts for the biggest share of our registered accounts).

However, Snapchat's platform architecture, combined with its commitment to responsible policy enforcement across our content surfaces, establishes unique safeguards against risks to democracy. 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),[47] 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. We participated in the Commission's stress test and multi-stakeholder roundtable dialogues ahead of the European Elections in 2024 and were able to successfully navigate the test exercises. As we reported in our election blog post on 24 June 2024, Snap saw a small uptick in reported activity, but did not receive or observe any material incidents or threats. Our moderation and reporting tools worked well, and none of the reported pieces of content were verified as misinformation on Snapchat.

As highlighted in Section 4.1.10 (Dissemination of harmful false information), Snap's own reporting metrics confirm the limited occurrence of content harmful to democracy:

- Our prevalence testing has consistently shown a very low prevalence of 'harmful false information'. Our testing in July 2024 showed a further significant reduction in prevalence.
- Our most relevant transparency reporting category on this topic is "harmful false information" which our policies define as including content that "undermines the integrity of civic processes." False information continues to account for only 0.1% of the total of all content enforced on Snapchat. This figure remained steady throughout both halves of 2022, as well as through the second half of 2023. We track "Impersonation" enforcements separately, and this similarly accounts for a very low percentage of our enforcement actions (0.3% in the second half of 2023, compared with 0.2% in the second half of 2022).

Snap's product design and policy practices continue to substantially reduce the likelihood of negative impacts on democracy. Our ongoing quantitative and qualitative analysis shows that the risk of potential negative impact on democratic and electoral processes on Snapchat falls into our **Extremely Low likelihood category.**

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.**

Accounting for the real-world examples illustrating the potential disruptive effects that digital platforms can have on democratic and electoral processes, we understand that a negative effect on democratic and electoral processes has a risk of **severe harm** if not properly mitigated.

[47] Election Integrity Partnership (2020), url.

Overall potential risk prioritization

Taking into account the real-world examples illustrating the potential disruptive effects on democracy, there is risk of a severe harm risk if not mitigated, As a result, we assess this risk to fall within our **Level 3 overall potential risk prioritization** category. There is no change in this assessment from our 2023 Report.

Snap's Mitigations

Highlights

Snapchat's architecture and 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. However, the steps Snap has taken to mitigate threats to democracy mean that likelihood is substantially diminished.

As we highlighted in our 2023 Report, Snap has for some time taken a multifaceted approach to mitigating negative impacts to democracy, including policy enforcement, product design, and expert engagement. This approach aligns with Guideline for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes,[48] as follows:

Internal Processes

As we have explained in this Report, and highlighted above, our assessment of risk prioritization with regards to negative effect on democratic and electoral processes and Snap's measures to mitigate the risk, are guided by information on elements such as the presence and activity of political actors on the service and the number of Snapchatters in the EU and evidence regarding the use of tactics, techniques and procedures for information manipulation.

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).

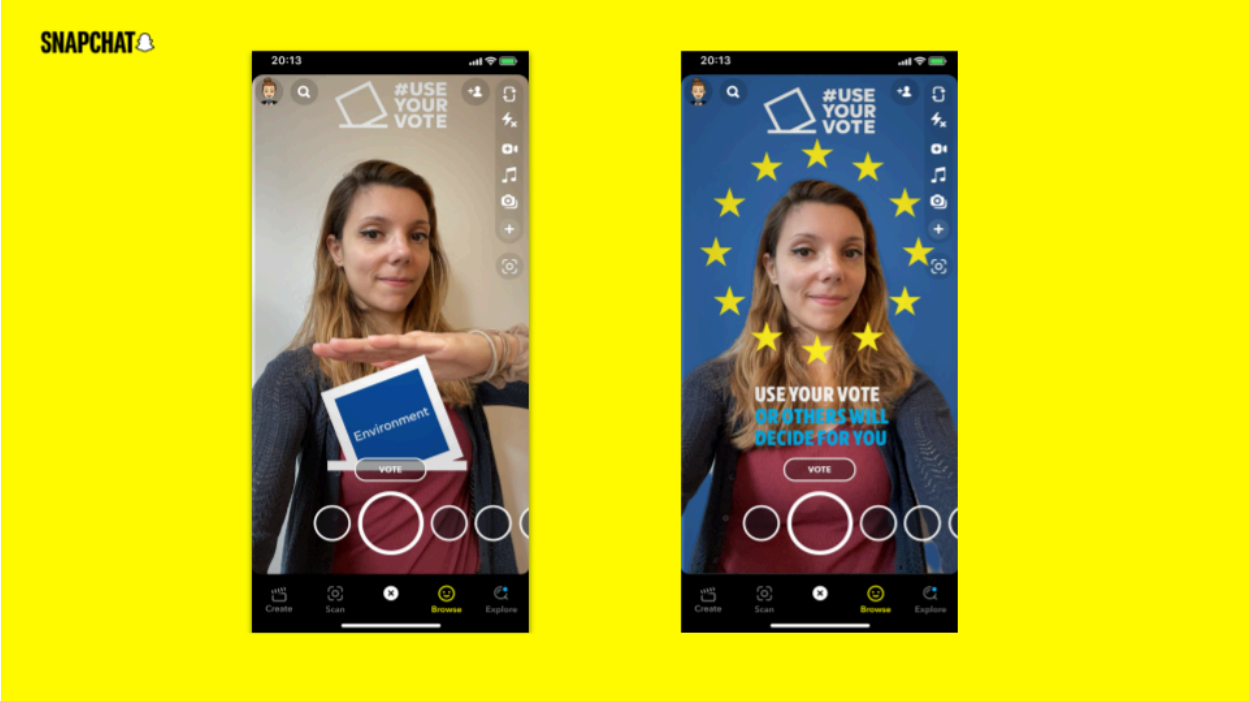
Access to official information on the electoral process and Media literacy initiatives

We have regularly partnered with governments around the world to inform Snapchatters about elections and invite them to go vote. We believe that civic engagement is one of the most powerful forms of self-expression and have previously worked with election authorities in France, Netherlands, and Sweden to raise awareness of elections and encourage participation. A recent

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

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

The recent European elections saw even more first-time voters eligible to participate – following the decision by Belgium and Germany to join Austria, Malta and Greece in lowering the voting age to 16. Ahead of this year's EU elections, we teamed up with the European Parliament on a special AR elections Lens that encourages people to get out and vote. During the election, we shared this Lens with all EU Snapchatters along with a message to remind them to vote and a link to the Parliament's election website.



Snapchat partnered with the European Parliament and European Commission to promote their 'Use your vote' information campaign on elections, including a dedicated Lens, and their awareness campaign on the risks of disinformation and deceptive content. Snap sent a push notification to over 50 million users to urge them to vote in the elections. Although out of scope of this Report, we note that we also further instructed My AI to avoid engaging on political topics. We are proud to have done our part to contribute to the highest observed turnout of the last 30 years, with 51.08% of the 357 million eligible citizens participating in the election.

Measures to provide users with more contextual information

One key way in which we mitigate the risk is 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.[49] 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).

Snap's policies expressly prohibit content that undermines the integrity of elections and civic processes. Drawing from expert research from the Election Integrity Partnership,[50] 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.

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

As technologies have evolved, we have updated our policies to cover all content formats – whether created by a human or generated by artificial intelligence (including deep fakes). In preparation for the recent EU elections, we also:

- Signed up to the AI Elections Accord, alongside other technology firms, where we pledged to work collaboratively on tools to detect and limit the spread of AI generated content which aims to deceive voters.
- Introduced contextual symbols to help our community understand when they are interacting with Snap generated AI content.

Snap does not allow Lenses that encourage a particular political perspective. In line with this approach, politically related Lenses are disabled in Discover. Snap also rejects Lenses that

[49] 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.

[50] Election Integrity Partnership, '*Evaluating Platform Election-Related Speech Policies,* October 2020, url.

perpetuate false information to elections (e.g. the wrong date). AR moderators are given strict guidance during elections to escalate misinformation.

As a result, when it comes to the inscope content services of Snapchat, rather than taking measures to provide users with more context around disinformation and Foreign Information Manipulation and Interference (FIMI) content through labels and other indications, Snap's approach is to take steps to avoid recommending such content to a public audience in the first place (see below) and to remove such user generated content promptly when it is detected or reported. To the extent political content is distributed on in-scope services (i.e. political ads), Snap has safeguards in place, which are detailed in the political advertising section below.

Recommender Systems

Content that is approved for broader audiences must comply with both our Community Guidelines and our Content Guidelines for Recommendation Eligibility. All Spotlight and non-professional user generated Discover content goes through both automoderation, and often human review, against these guidelines before it is eligible for recommendation to a wide audience. As an additional safeguard, we monitor content that is achieving large-scale reach (and ensure a human reviews it) as a sort of "virality circuit breaker" and a means of checking that our pre-moderation systems are working effectively. Any content that is reported will be reviewed against the guidelines again for compliance.

Political Advertising

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

[51] For more details url.

In preparation for the EU elections we also partnered with Logically Facts, a leading fact checking organisation and signatory of the EU Disinformation Code of Practice, to help fact check political ad statements across the EU.

We do not require ads to label when advertisement includes generative AI content nor do require advertisers to disclose to us the tools they used to edit or create their ad creative. Instead, our approach is to subject all of our ads to a review process, and political ads are also subject to fact checking. Deceptive ads are rejected, irrespective of whether they use AI, photoshop, or other digital editing tools. Ads that are not deceptive, and otherwise comply with our Ad Policies, are approved to run (and if they are a political ad, they must include a "paid for by" disclaimer and are catalogued in Snap's political ads library).

Influencers

Our commercial content policy requires all organic content posted by influencers to be marked appropriately. Commercial content that relates to the following is not permitted:

- Election-related content about candidates or parties for public office, ballot measures or referendums, political action committees, and content that urges people to vote or register to vote.
- Advocacy or issue content concerning issues or organisations that are the subject of debate on a local, national or global level, or of public importance. Examples include: content about abortion, immigration, the environment, education, discrimination and guns.

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 make clear that Snap restricts the paid promotion of political messaging to traditional ad formats. This is in order to be responsible to our community and to maintain transparency.

Demonetisation of disinformation content

The policies and other mitigations highlighted in this Section ensure that the placement of advertising does not provide financial incentives for the dissemination of disinformation and FIMI with regards to electoral processes and hateful, (violent) extremist or radicalising content that can influence individuals in their electoral choices.

Integrity of services

As explained in this Section, we have appropriate procedures to ensure the timely and effective detection and disruption of manipulation of the service when this has been identified by them as a relevant systemic risk, taking into account the best available evidence. We explicitly prevent the use of "any robot, spider, crawler, scraper or other automated means or interface to access" Snapchat; use of Snapchat "in any manner that could interfere with, disrupt, negatively affect or inhibit other users from fully enjoying" Snapchat and any "attempt to circumvent any content-filtering techniques we employ" on Snapchat.

When we determine that a user has violated our Terms, we may remove the offending content, terminate the relevant account, and/or notify law enforcement. We may also briefly limit the visibility of content suspected of being illegal or otherwise violating our terms if needed to enable time for human moderators to review and provide confirmation (known as "temporary soft removal"). See Section 5.5 for more information.

Third party security and research

The severity of these risks is reflected in the resourcing Snapchat has committed to partnerships and collaborations with leading researchers and civil society organizations who are analyzing threats to democratic information environments, including the Atlantic Council's Digital Forensics Research Lab, the Center for a New American Security, the Stanford Cyber Policy Center and University of Washington, and the Poynter Institute (which is also secretariat for the International Fact-Checking Network). Also reflecting its serious approach to this risk, Snapchat has agreed on voluntary rules for the 2021 Dutch elections in a Code of Conduct, which governs transparency commitments regarding online political advertisements during election campaigns (see below for more detail).[52]

Snap is also subject to audit under the DSA which includes a review of Snap's compliance with its risk assessment and mitigation obligations.

Fundamental rights

As set out in Section 4.2 of this Report, in line with the requirements of the DSA, when assessing its risks and mitigations, Snap has paid due regard to:

- the protection of **fundamental rights** enshrined in the Charter of Fundamental Rights of the European Union, in particular the right to freedom of expression and of information; and
- the impact of measures to tackle illegal content such as public incitement to violence and hatred to the extent that such illegal content may inhibit or silence voices in the democratic debate, in particular those representing vulnerable groups or minorities.

As explained above, our platform does not, for example, widely distribute an unvetted feed of algorithmically curated political information. Under our Content Guidelines for Recommendation Eligibility, Political content is also not eligible for promotion in Spotlight, limiting the ability of any user to share political content with strangers on Snapchat, unless it's from trusted news partners and creators, and pre-moderate that surface to ensure that other such political content is not distributed.

Mitigation measures linked to generative AI

[52] The Dutch Code of Conduct Transparency Online Political Advertisements, url.

Snap maintains robust policies––applicable to both the dissemination and the creation of generative AI content––that function to mitigate risk and advance safety.

Creation
On-platform features for creating generative AI content are not part of Snap's inscope services and are out of scope of this Report (save for certain commonplace ad creation tools). Nevertheless, outside of its DSA obligations, we note that Snap has developed several internal policies relating to generative AI. In particular,

(1) Content and Product policies: We have developed a suite of policies that disallow the generation of harmful content (including deceptive political content). Our policy and moderation teams work in partnership with engineering and data science colleagues to ensure that our AI products are responsibly trained on these policy parameters.
(2) Acceptable Use: We have similarly developed Acceptable Use Policies that prohibit the use of our AI tools to attempt to generate violative content at the prompt-level.

These aligned very closely with the rules for content dissemination, which are explained below.

We have also introduced contextual symbols to help our community understand when they are interacting with Snap generated AI content. We have created a generative AI support page to explain our approach to Snapchatters and other stakeholders.

Dissemination
In the context of dissemination of content on Snapchat's online platform, in scope of the DSA, 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.

Our Community Guidelines and Terms of Service set out the rules on what content is allowed on Snapchat. They are focused on preventing harm to Snapchatters and the broader community from content and behaviour, whether or not caused by generative AI or any other form of IT tools (such as Photoshop). These rules apply to all content formats across our platform, including content that is AI-generated. While the rules are agnostic to content format or creative tools, the Community Guidelines specifically note: "We implement safeguards designed to help keep generative AI content in line with our Community Guidelines, and we expect Snapchatters to use AI responsibly. We reserve the right to take appropriate enforcement action against accounts that use AI to violate our Community Guidelines, up to and including the possible termination of an account."

Our rules and internal enforcement guidance include clear provisions related to content risks for civic discourse and electoral processes. In particular, our Community Guidelines 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).

As technologies have evolved, we have updated our policies to cover all content formats – whether created by a human or generated by artificial intelligence. Our Community Guidelines rules on false information refer to a more detailed Explainer that prohibits content that undermines the integrity of civic processes, or deep fake content or other media that is manipulated for false or misleading purposes. The Community Guidelines further explain that these prohibitions extend to the following types of harmful content:

- *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 rumours 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.

Sharing such content will violate Snap's Community Guidelines irrespective of whether it is AI-generated or user-generated, or whether it is generated on Snapchat or on another platform.

Snap enforces these Community Guidelines fairly and consistently, using internal policies and guidelines, and applies outcomes that are commensurate with the severity of risk. Accounts that we determine are used to perpetrate serious, high-severity harms will immediately be disabled. For other violations of our Community Guidelines, Snap generally applies a three-part enforcement process:

- Step one: the violating content is removed.
- Step two: the Snapchatter receives a notification, indicating that they have violated our Community Guidelines, that their content has been removed, and that repeated violations will result in additional enforcement actions, including their account being disabled.
- Step three: our team records a strike against the Snapchatter's account.

A strike creates a record of violations by a particular Snapchatter. Every strike is accompanied by a notice to the Snapchatter; if a Snapchatter accrues too many strikes over a defined period of time, their account will be disabled.

This strike system ensures that Snap applies its policies consistently, and in a way that provides warning and education to users who violate our Community Guidelines. The primary goal of our policies is to ensure that everyone can enjoy using Snapchat in ways that reflect our values and mission; we have developed this enforcement framework to help support that goal at scale.

Snap has a suite of internal policies and guidelines to help our content review and trust and safety teams apply the Community Guidelines to user generated content disseminated via our online platforms (such as Spotlight and Discover). They provide more granular information for our content review teams.
In preparation for the recent EU elections, we also signed up to the AI Elections Accord, alongside other technology firms, where we pledged to work collaboratively on tools to detect and limit the spread of AI generated content which aims to deceive voters. However, as noted above, Snap's product design and policy practices outlined above have been demonstrated to be effective in mitigating the risks of deceptive political content, including content generated using AI tools, from achieving meaningful scale on Snapchat and substantially reducing the likelihood of negative impacts on democracy. As noted above, all of our ads are subject to review, and political ads are also subject to fact checkingDeceptive ads are rejected, irrespective of whether they use AI, photoshop, or other digital editing tools.

We continue to detect and monitor risks as outlined in Section 6 of the Risk Report (including working with our Safety Advisory Board on the intersection of safety and generative AI technology) and adapt our mitigations accordingly.

Cooperation with national authorities, independent experts and civil society organisations
Snap has closely followed the negotiations on the EU AI Act and plans to continue to actively engage and assess collaboration opportunities on the upcoming AI Act, as well as on the drawing of the related codes of practice for providers of general-purpose AI models and those regarding the detection and labelling of artificially generated or manipulated content.

More broadly, tackling risks stemming from generative AI requires (among others) broad industry-wide technical solutions which have not been clearly identified so far. This is why Snap is actively engaging with its peers and industry experts in different fora to share best practices and advance the technical debate. These partnerships, industry collaborations and efforts include:

- *OpenAI*: Although My AI is out of scope of this Report, the fact that My AI is powered by OpenAI's ChatGPT, has led to a good working partnership with OpenAI. This allows the companies to share best practices, including with respect to content moderation.
- *Tech Coalition / Working Groups on Generative AI*: Snap is a member of the Tech Coalition's Working Group on Generative AI Content, and a member of the GenAI Briefing Subgroup. The Working Group on Generative AI Content meets regularly to facilitate dialogue and information- and idea-sharing around mitigating content-level generative AI risks. The GenAI Briefing Subgroup meets periodically to plan expert briefings for Tech Coalition members on topics related to Generative AI risks; such briefings have included representatives from government, law enforcement, civil society, and the research community.

- *Tech Accord to Combat Deceptive Use of AI in 2024 Elections*: Snap was an initial signatory to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. This compact seeks to set expectations for how signatories will manage the risks arising from deceptive AI election content created through their publicly accessible, large-scale platforms or open foundational models, or distributed on their large-scale social or publishing platforms in line with their own policies and practices as relevant to the commitments in the accord. The Accord was announced at the Munich Security Conference in February 2024.
- *ITI AI Futures Initiative*: Through its membership in the Information Technology Industry Council (ITI), Snap has participated alongside other private sector actors in the AI Futures Initiative. Led by technical and policy experts spanning the tech ecosystem, the Initiative is a forum through which participants are developing action-oriented recommendations for AI policy and working to address emerging questions around AI. Deliverables to date have included the issuance of Global AI Policy Recommendations to help guide governments around the world as to develop responsible regulatory approaches to AI-related issues.
- *HackerOne - Red-Teaming Collaboration*: Snap partnered with HackerOne on red teaming exercises to test the strict safeguards Snap has in place around AI. Together with HackerOne, we made significant developments in the methodology for AI safety red teaming that has led to a more effective approach to surfacing previously unknown problems. We refer to the HackerOne blog for more details: https://www.hackerone.com/ai/safety-vs-security
- As an active member of the EU Internet Forum, Snap will support the upcoming dedicated working group on generative AI matters.
- We are also members of the Centre for Information Policy Leadership (CIPL) and the Future of Privacy Forum (FPF) which work with industry stakeholders (like Snap), NGOs and government agencies in each region to advance a broad array of information topics. CIPL has been a leader in AI matters for many years through its dedicated AI Project and specific Brazilian AI Project. Most recently, in Europe, CIPL has responded to the UK Information Commissioner's Office (ICO)'s consultations on Generative AI, and led various forums on Accountable Governance of AI and AI Regulation in Brussels and the UK. Similary, FPF is working on AI Governance and other responsible Gen AI initiatives.

Further, we actively engaged in the Commission's public consultation on its proposed DSA Election guidelines, and similar consultations and queries raised by national DSCs. As shown above, we have worked to update our risk assessment to take into account the recommendations in those guidelines.

Specific Mitigations

In addition to the detailed highlights above, 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 *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.
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 Discover. 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 Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not promote political content on Spotlight.
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.	Yes, 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 prevent advertisers from manipulating small audiences with micro targeted campaigns, particularly for political ads, by requiring a specific minimum audience of Snapchatters.
Risk Detection and Management	Yes, we review and monitor compliance with our internal terms, policies and procedures.

Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.	
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.	Our Trusted Flaggers are not typically focussed on this risk, but we welcome their input on this matter.
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.	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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we proactively encourage our users to go to vote through interactive campaigns.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[53]
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	We recognise the risk that generative AI could be used to generate harmful false misinformation, including deep fakes. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on**

[53] https://parents.snapchat.com.

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.	**Snapchat's inscope services.** We also label political advertisements, and maintain a political ads library.

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 taken into account the Commission's recommendations set out in the Guideline for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes when carrying out our assessment.

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. There is no change in this conclusion from our 2023 Report.

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:

- In our 2023 Report, the prevalence of harmful false information was measured at an extremely low percentage, and represented just 0.1% of total content enforced in the second half of 2022. Our prevalence testing showed that illegal hate speech made up an extremely low percentage of Policy Violating Prevalence (PVP) on Snap in August of 2023.
- We have subsequently observed a further substantial decrease in the prevalence of content falling within these categories.

Consequently, we continue to conclude that this risk still falls 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 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 prioritization

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 potential risk prioritization category**. There is no change in this assessment from our 2023 Report.

Snap's Mitigations

<u>Highlights</u>
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).[54]

Snap establishes additional safeguards against risks to civic discourse on our surfaces that help distribute content algorithmically. All Spotlight and Discover 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.[55]

[54] The ADL database is available at: url.
[55]The 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 *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.
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 Adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and,	Yes, 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

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.	Community Guidelines and our Content Guidelines for Recommendation Eligibility. Any content that is reported will be reviewed against these guidelines again for compliance.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not create echo chambers and ensure users are subject to different types of content and viewpoints.
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.	Yes, 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 prevent advertisers from manipulating small audiences with micro targeted campaigns, particularly for political ads, by requiring a specific minimum audience.
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.	Yes, we review compliance with our terms and processes.
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.	Our Trusted Flaggers are not typically focussed on this risk, but we welcome their input on this matter.
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.	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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Yes, we proactively encourage our users to go to vote through interactive campaigns.

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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[56]
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.	We recognise the risk that generative AI could be used to generate harmful false misinformation, including deep fakes. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.** We also label political advertisements, and maintain a political ads library.

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.

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

[56] https://parents.snapchat.com.

micro targeted advertising. We specifically monitor the dissemination of harmful false information (Section 4.1.10) and dissemination of illegal hate speech (Section 4.1.2), which are considered primarily sources of negative effects on civil discourse. Both have been observed to have very low prevalence on Snapchat's inscope services such as Discover and Spotlight.

As a result, we have concluded that Snapchat's in-scope services have reasonable, proportionate and effective mitigation measures against negative effects on civil discourse. There is no change in this conclusion from our 2023 Report.

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).

To assess the likelihood of Snapchat's inscope services having a negative effect on public security, we have reviewed the sources of data relating to the following illegal and otherwise violating content categories considered to have a particular impact on undermining public security:

Category	*Relative likelihood of risk occurring on Snapchat*
4.1.2 Dissemination of illegal hate speech	Extremely Low Likelihood
4.1.4 Dissemination of terrorist content	Extremely Low likelihood
4.19 Dissemination of content encouraging or engaging in violent or dangerous behavior	Extremely Low Likelihood
4.1.10 Dissemination of harmful false misinformation	Extremely Low Likelihood

The low prevalence rate of these harms supports our continued assessment that the volume of content presenting risks to public security is quite low on Snapchat, 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**.

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.**

To assess the likelihood of Snapchat's inscope services having a negative effect on public security, we have reviewed the sources of data relating to the following illegal and otherwise violating content categories considered to have a particular impact on undermining public security:

Category	*Relative likelihood of risk occurring on Snapchat*
4.1.2 Dissemination of illegal hate speech	Significant harm industry wide
4.1.4 Dissemination of terrorist content	Serious harm industry wide
4.19 Dissemination of content encouraging or engaging in violent or dangerous behavior	Significant harm industry wide
4.1.10 Dissemination of harmful false misinformation	Significant harm industry wide

As these range from significant to serious, and given the context outlined above showing serious consequences of a negative effect on public security, we assess that this category would fall within a **'serious' 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**.** There is no change in this assessment from our 2023 Report.

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 (see Section 5.12). 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 and Public Stories 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?
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 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, Discover, 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.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, terms prohibit content that may have a negative impact on public security, including harmful, dangerous, and inciteful content, and they are strictly enforced.
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 that may have a negative impact on public security, including harmful, dangerous, and inciteful content. Content on Spotlight undergoes 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. All Discover UGC content is moderated by humans, and we proactively remove content that doesn't meet our policies before being broadly distributed. 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.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
Risk Detection and Management	Yes, we have specific prevalence testing and transparency reporting for content that may have a

Reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk.	negative impact on public security.
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.	Yes, we cooperate with trusted flaggers, our trusted flaggers 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.
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.	Yes, we cooperate with other providers through various industry groups.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[57]
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.	We recognise the risk that generative AI could be used to generate harmful false misinformation, including deep fakes. **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.** We also label political advertisements, and maintain a political ads library.

[57] https://parents.snapchat.com.

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. There is no change in this conclusion from our 2023 Report.

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*	*Conclusion*
4.4.1 Negative Effects on Public Health	Extremely Low Likelihood	Severe harm industry wide	Level 3	Low Risk / Reasonable, proportionate and effective mitigations
4.4.2 Negative Effects on	Extremely Low Likelihood	Serious harm industry wide	Level 2	Low Risk / Reasonable,

gender-based violence				proportionate and effective mitigations
4.4.3 Negative effects on Children	Varies	Severe harm industry wide	Level 1	Low Risk / Reasonable, proportionate and effective mitigations and Snap is participating in efforts to develop an EU wide guidance 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	Low Risk / Reasonable, proportionate and effective mitigations

4.4.1 Negative Effects on Public Health

We recognize that without adequate mitigations, digital content platforms like Snapchat could contribute to 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.
While we believe these risks to be probable in the absence of mitigations, we assess that Snap's mitigations appreciably reduce the likelihood of encountering these harms on 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).
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:

- In our 2023 Report, we observed that we had a low prevalence of health misinformation and other harmful false information to be extremely low and self-harm content is extremely low based prevalence testing.
- We are pleased to have subsequently observed a further fall in provenance of this content, and the PVP rates are now 0.001 % for health misinformation and other harmful false information and 0.014% for self-harm and suicide.

For the potential negative effects on physical and mental well-being, see the relevant section.

Therefore, we continue to assess there is an **extremely low likelihood** of negative effects on public health arising from Snapchat's in-scope services.

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. There is no change in this assessment from our 2023 Report.

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 Adapting their advertising systems and adopting targeted measures aimed at limiting or adjusting the presentation of advertisements in association with the service they provide.	Yes, other mitigations listed here also apply to our Advertising Systems.
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.	Yes, we have specific prevalence testing and transparency reporting for harmful false information and self-harm content.
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.	Yes, we cooperate with trusted flaggers in relation to illegal content that harms public health.
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.	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.
Transparency	Yes, we provide guidance on our terms, harms, moderation and enforcement practices (see the

Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	Annex), as well as how to and how to get help in our Safety Center. **This includes specific new resources for sextortion in an effort to support those in distress.**
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[58]
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.	Yes, **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.** This would include harmful false information, such as deep fakes.

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 observed a further substantial fall in prevalence rates for harmful false information and self-harm and suicide content, and these are now at extremely low levels, 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 tools to support our community via online and in-app support tools.

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. There is no change in this conclusion from our 2023 Report.

[58] https://parents.snapchat.com.

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, harmful false information, and violence and we have observed very low levels of prevalence of this type of content on Snap's inscope services:

- In our 2023 Report, we were 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 harmful false information; and all forms of harassment (including NCII, sexual harassment, and sextortion).
- We are pleased to have subsequently observed a significant fall in prevalence across all of these content types.

From this, we continue to conclude that content promoting gender based violence falls within our **Extremely Low Likelihoodcategory** 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. There is no change in this assessment from our 2023 Report.

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 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 help all stakeholders understand these problems across the online community. As part of our Year Two Digital Well-Being study, we conducted a deeper drive into teens' and young adults' exposure to "sextortion" across platforms and services. The target countries were Australia, France, Germany, India, the UK, and the U.S, which includes three of the largest European countries, two of which are in the EU). We have continued this research during 2024 ("Year Three"), and also investigated teens' and young adults' attitudes and sentiments around reporting problematic content to platforms and services, authorities and others. More information on this research can be found in Section 6.6.

We believe our approach to these challenges reflects our commitment to responsibly mitigating harms that may negatively impact gender-based violence. We received 67,152 reports of Hate Speech content in the EU in the second half of 2023, which led to enforcement action being taken against 8,894 unique content items and 15,007 accounts. This suggests users are able and willing to report content and accounts as needed on Snapchat (note these numbers relate to the whole of Snapchat, including private spaces that are out of scope of the DSA). We are encouraged by evidence that our approach has contributed to a low prevalence of CSEAI and hate speech content on the inscope services of 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?
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 for illegal hate speech and violence, which includes gender-based violence.

	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. For Lenses submitted to Snapchat, we reject harmful Lenses to reduce the likelihood that they are distributed on Snapchat.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes, terms prohibit gender-based violence and they are strictly enforced.
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 violent content, and illegal hate speech, as further detailed in previous sections of this report.
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	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.
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.	Yes, other mitigations listed here also apply to our Advertising Systems.
Risk Detection and Management Reinforcing the internal processes, resources, testing, documentation, or supervision of any of	Yes, as outlined in the previous sections on hate speech and violent content.

their activities in particular as regards detection of systemic risk.	
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.	Yes, we cooperate with trusted flaggers in relation to illegal hate speech.
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.	Yes, we cooperate with other providers through various industry groups e.g. EUIF.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
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.	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. **Our new parents site provides additional guidance for parents and carers on risks and support.**[59]
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.	Yes **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.** This includes hate speech and violent content.

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

[59] https://parents.snapchat.com.

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. There is no change in this conclusion from our 2023 Report.

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.

Several studies have considered the likelihood of underage use of online platforms. For example, in 2022, Ofcom in the UK found that 60% of children aged 8 to 12 use social media with their own profiles.[60] There were similar reports in Belgium.[61] Another example found the percentage of underage users on Snapchat to be relatively low (fewer than 4% of 0-11 year olds in the US in 2024) compared with other online platforms such as YouTube (28.6%), Netflix (17.2%) and Disney+ (15.6%).[62] We take a risk based approach to age assurance at present that aligns with industry practice. We support further proportionate, reasonable and effective industry wide measures supported by device OS / app store account level age assurance and are proactively working with the Commission and others to try to achieve an EU wide approach. We have highlighted our

[60] Children's Online User Ages - Quantitative Research Study, Ofcom, updated July 2022, url.
[61] Réseaux sociaux, règles d'utilisation, intelligence artificielle : comment a évolué l'utilisation des écrans chez les jeunes ?, RTBF, May 31, 2024.
[62] Youth and Social Media, How US Kids and Teens Use Platforms From TikTok to Snapchat to YouTube, EMarketer, March 2024.

measures regarding minors under 13 in the specific mitigation section below (and provided more detailed information in Section 5.8).

All 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. It is primarily in the Extremely Low Likelihood but rises to Very Low Likelihood for dissemination of harassment and bullying content (given the second highest prevalence of violating content in this category relative to other risks) and adult sexual content (given the highest relative prevalence of violating content in this category relative to other risks) and for negative effects on data protection rights (as the high volume of personal data we process gives a high risk of some harm if not handled appropriately).

Severity

As with likelihood above, the 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 suffer a risk of greater harm from the issues we have identified and we have chosen to place the severity of harm arising from an issue that negatively affects children in our **'severe harm' 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**. There is no change in this assessment from our 2023 Report.

Snap's Mitigations

Highlights
The risk of negative effects on children falls within our highest risk prioritization level. To protect Teens on Snapchat, at a high level, we focus on three core things: 1) mitigating unwanted contact; 2) scanning for, detecting and blocking/removing content that violates our Terms (including our Community Guidelines) or the law; and 3) working with law enforcement to help bring criminals to justice. 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 measures are set out in Section 5.8 of this Report.

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.

A more detailed run though of our mitigations to protect Teens on Snapchat is set out in Section 5.8 (Protection of Minors). 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?
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 safety by design features and content moderation for Teens. For example: ● 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 Discover 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.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	Yes. For example, our Terms require that all content is appropriate for 13+, we require all users on our platform to be over the age of 13, and we strictly enforce our terms. If we discover that a user is under the age of 13 we will remove their account.
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,	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.

as well as adapting any relevant decision-making processes and dedicated resources for content moderation.	
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems take user age into account to provide age appropriate recommendations.
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.	Yes, other mitigations listed here also apply to our Advertising Systems. 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.
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.	Yes, we have specific prevalence testing and transparency reporting for violations, including for example in relation to CSEAI.
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.	Yes, we cooperate with numerous trusted flaggers for child safety who are able to flag other CSEAI or other illegal and violating activities involving Teens.
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.	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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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. 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. We also provide Family Center as a resource to Teens and their parents or trusted adults.

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.	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. Our Family Center - Parental Control For Teens | Snapchat Safety provides Teens and their parents or trusted adults a suite of resources and guidance. **Our new parents site provides additional guidance for parents and carers on risks and support.**[63]
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.	Yes, **Snap has taken steps to mitigate the risk that (i) its generative AI tools are used for creating illegal or otherwise violating content and (ii) illegal or otherwise violating content created using generative AI tools on any online platform are disseminated on Snapchat's inscope services.** We display an icon in some Lenses that manipulate an image of a Snapchat to make them look younger.

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 continue to actively participate in cross-stakeholder efforts to develop an EU

[63] https://parents.snapchat.com.

wide AADC and/or guidance on high level of privacy, safety and security under Art 28 of the DSA 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.

On Snapchat, data related to relevant policy enforcements suggests a low prevalence of content associated with harm to physical well-being on Snapchat. For example:

- In our 2023 Report, 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.
- We have subsequently observed a significant decrease in the prevalence of all of these categories of content.

We also monitor community support requests, which are a good indication of the well-being of our Community. We continue to receive very few user complaints from EU users related to dangerous categories that could be considered to have a serious negative effect on physical and mental well-being

We have assessed the amount of average time that minors spend on the features of Snapchat and it is worth noting that the majority of the time spent (i) relates to time spent on surfaces that are not in scope of the DSA and this Report; and (ii) is focussed on communication with close friends and family (see the table below which shows that Teens spend the majority of their time using Chat and Camera features)).

We have also assessed the average time spent, as well as the time of day, that Teens use Snapchat and note that the majority of time is spent in the evening and that average time the average times are not excessive. As noted in the mitigations section below, many devices now include well being settings such as turning on bedtime mode by default and providing screen time controls for parents / responsible adults.

Over 90% of our community says they feel happy, connected, and comfortable while using Snapchat.[64] Research from University of Chicago's NORC[65] stows that 2 in 3 say messaging with family and close friends makes them extremely or very happy. On the other hand, a majority of teens and young adults feel overwhelmed at the way traditional social media makes them feel pressured to post content that will get lots of likes and comments, or will make them look good to others. Perhaps most importantly, according to the NORC data, respondents who use Snapchat report higher satisfaction with the quality of friendships and relationships with family than non-Snapchatters.

As we were finalizing this Report, we were made aware of a new study from the Netherlands (currently in pre-print).[66] According to the researchers, it is the first quantitative study that compares online platforms on three key factors, impact on well-being, self-esteem and friendship closeness. The researchers conducted an extensive 100-day daily diary study among 479 adolescents (14-17 years). With respect to Snapchat specifically, the research concludes that Snapchat is the only platform that positively impacts well-being and Snapchat also has a strong positive effect on friendships and no net negative effect on self-esteem. It notes in particular:

- Snapchat positively affected friendship closeness and well-being but had no significant impact on self-esteem. Using WhatsApp had a notably strong effect on friendship closeness but no significant effect on well-being and self-esteem.
- The majority of adolescents (60%) experienced unity in negative effects of social media, suggesting that social media use is a contributor to mental health issues. Moreover, 13.6% of adolescents experienced duality in effects, indicating that social media use simultaneously harms and benefits different dimensions of their mental health.
- The positive and null effects associated with Snapchat and WhatsApp indicate that we should avoid a blanket condemnation of all social media platforms.

[64] 2022 Alter Agents study commissioned by Snap Inc. url.
[65] https://www.norc.org/about/who-we-are.html.
[66] 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)

- Snapchat scores a 41.4% positive effect on well-being, 23,7% on self-esteem and 71,5% on friendship closeness, as shown in the figures and table below:

Figure 2a. TikTok

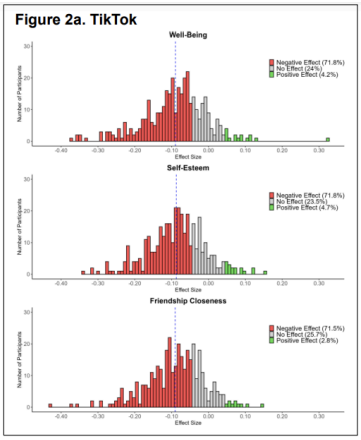


Figure 2b. Instagram

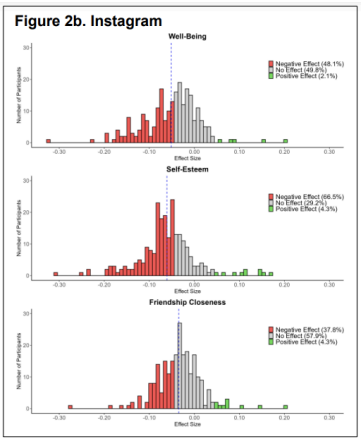


Figure 2c. YouTube

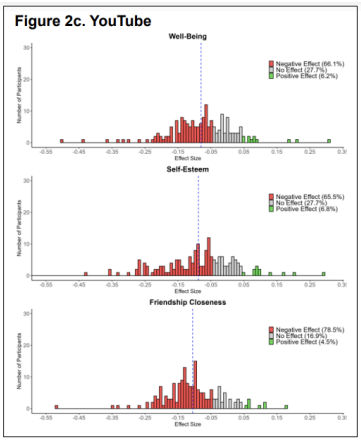


Figure 2d. Snapchat

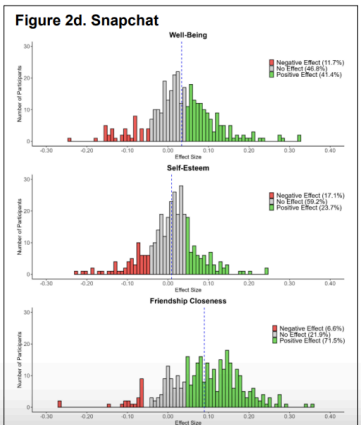
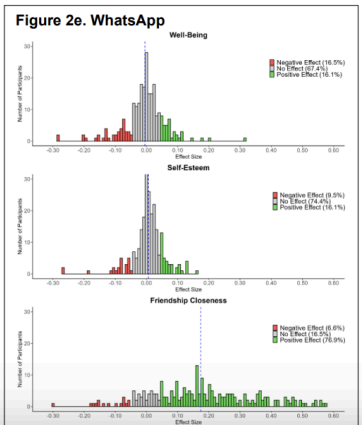


Figure 2e. WhatsApp



Note. Figure 2a-2e shows the distribution of effect sizes for TikTok (n = 358), Snapchat (n= 333), WhatsApp (n = 242), Instagram (n = 244), and YouTube (n = 177) respectively. The x-axes display standardized effect sizes, the y-axes display the number of participants. Vertical dashed blue lines represent the average person-specific effect (see also Table 3).

Table 3

Overall Within-Person Effects of the Top Five Social Media Platforms on Mental Health

	Well-Being	Self-Esteem	Friendship Closeness
TikTok	-.09***	-.08***	-.09***
Instagram	-.05***	-.06***	-.04**
YouTube	-.08***	-.09***	-.11***
Snapchat	.03***	.01	.09***
WhatsApp	-.00	.01	.17***

Note. Cells marked in red indicate significantly negative overall within-person effects, cells marked in green indicate significantly positive overall within-person effects.

*< .05. **< .01. ***< .001.

Accordingly, while Snap assesses these risks across digital platforms to be high in the absence of safeguards and mitigations, we are encouraged by research and data indicating that our approach to mitigating these risks is effective at reducing the likelihood of such negative impacts on physical and emotional wellbeing on Snapchat. We continue to assess there to be an **Extremely low likelihood** of encountering this content on Snapchat.

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**. There is no change in this assessment from our 2023 Report.

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 views:

- **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 the recommendation of content from younger users to older users on Spotlight. 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.

In addition, as we have explained in the mitigations section of this Report, in particular Section 5.8 (Protection of Minors), teenagers, parents and other responsible adults are able to set time limits for their teenagers, amongst other controls, via the device operating system's family tools (e.g. Google Family Link, Apple device parental controls and Family Sharing controls and Microsoft Family Safety). Mobile devices now also commonly provide default settings for late night usage, such as bedtime modes that turn off device and app notifications and turn the screen the black and white to encourage sleep.

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.
- We have repeated this research in 2024 and will be publishing the results in September. Further information can be found in Section 6.6 (DWBI Initiative).

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 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. 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. Our user generated content feature, Spotlight, has both creator and viewer protections in place.
Terms and Enforcement Adapting their terms and conditions and their enforcement.	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.
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	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.

decision-making processes and dedicated resources for content moderation.	
Algorithmic Systems Testing and adapting their algorithmic systems, including their recommender systems.	Yes, our algorithmic systems do not categorize or recommend content that our Community Guidelines prohibit. Our Content Guidelines for Recommendation Eligibility - Snap Inc. further describe how sensitive and disturbing content is demoted for distribution on Spotlight and Discover. For example, glorification of violence is not suggested content to users on Spotlight or Discover and any discussion on self-harm, including eating disorders is demoted to users based on their age, location, or personal preferences.
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.	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 side.
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.	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.
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.	Yes, we cooperate with trusted flaggers in relation to illegal hate speech and child safety.
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.	Yes, we cooperate with other providers through various industry groups e.g. EUIF. 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.
Transparency Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information.	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.
Protection of Minors	Yes, we have protective measures to ensure age appropriate content and our Family Center offers

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.	resources and guidance. **Our new parents site provides additional guidance for parents and carers on risks and support.**[67] Device operating systems commonly provide settings for Teens and their parents/responsible adults to manage screentime and late night use.
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. We are displaying an icon in some Lenses that manipulate an image of a Snapchat to make them look younger.

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. There is no change in this conclusion from our 2023 Report.

[67] https://parents.snapchat.com.

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

5.1.1 Adaptations and Mitigations

As a result of our privacy and safety by design approach (described in Section 6.3 of this Report), Snapchat was designed from the outset with core features and functionalities that mitigate the risks described in Snap's Risk Assessment Report.

As we have made these key foundational design decisions from day one. We hear from Snapchatters about the benefits of these choices all the time, as well as consulting with expert and teen stakeholders (such as those forming part of our Safety Advisory Board and Teen Council) 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 directly 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.[68] 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 and Discover

Simplified Snapchat

As noted in the What's New and Introduction sections of this Report, we are considering simplifying the Snapchat application from 5 to 3 tabs. Rather than having the current separate Spotlight and Discover tabs, this envisages a single unified tab for more public content to the right of the Camera. We plan to run some tests later this year to assess how it performs. This is primarily a cosmetic change, and it should not impact the mitigations including with respect to the design and function of what is currently Spotlight and Discover. For now we have continued to describe these mitigations separately for Spotlight and Discover below.

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, that 'spotlights' the best content on Snapchat. 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.
    - Creators can choose to approve comments on their Spotlight Stories prior to publication

[68] See also our More Snapchat campaign https://www.moresnapchat.com.

    - 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”.

Discover

Discover is dedicated to Creator Stories, which includes Media Partner content and some user generated content from popular users (“Creator Content”). The Creator Content that appears on Discover includes 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
    - Viewer comments are not typically available on Discover content. Where comments are enabled, they are subject to auto-moderation. Creators and other users can report comments, which leads to human review. They can also block commenters which will prevent them from ever seeing the blocked users’ comments again on any content.
    - We do not show “views” on Discover Stories. This 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
    - Content on Discover 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 Stories and Spotlights on their profile. Snapchatters can Follow a Public Profile from Spotlight, Snap Map, Discover, or by using the Follow button on a Public Profile. Unlike friend requests to non-Public Profile owners, Public

Profile owners will not receive a notification for new followers. We have made the 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. 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.
- 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 first 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 the 2024 Risk Assessment we 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

We filter Stories posted from users with new accounts so they do not feature on Snap Map. Further, content posted to Public Stories will only show on Snap Map with a clear location 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.** 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.

- **Permission and prompts.** 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: (1) whether they are sharing background location or only while using location; and (2) who among their friends can see their location at any time right from the settings gear in the Map.

- **All Friends Confirmation.** We check if users really want to share with all Friends the first time they choose the "My Friends" option with an in-app pop-up dialogue. This extra reminder appears once per account when the user selects or switches to All Friends sharing (if user selects "not now", the reminder will still be shown the next time the user attempts to switch to All Friends):

- **Location Sharing Change Confirmations.** When users change their location access settings (e.g. from “While Using the App” to “Always”, we display in-app alerts to confirm the change (e.g. "Your location now updates in real-time"). These are shown on the first time you come back to the app every time you change your access settings.

- **Ghost Mode Exit Confirmation.** Similar to the location change confirmations above, we also display a "You're on the Map!" reminder when the user exits Ghost Mode and is visible to any Friends they are sharing location with.

- **Device OS icon and reminders**. 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.

- **Location Sharing Reminders:** We also 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.

- **Auto-expiration of Last Active Location**. Where users have selected to share location only while using the app, we auto-remove users' location from the Map if they have not opened the app after 24 hours.

- **Creator protections**
    - Currently there is no comments on Snap Map content
    - 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 Discover, 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.

5.1.2 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 Discover. This is explained in Sections 5.2 (Terms) and 5.3 (Transparency).

Content Moderation
We moderate content on Snapchat in a number of ways to mitigate the risks of users being exposed to harmful and illegal content. This is explained in Sections 5.4 (Content Moderation) and 5.5 (Enforcement).

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 distributed 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 Discover or Spotlight.
This is explained in Sections 5.6 (Algorithmic Systems) and 5.7 (Advertising Systems).

5.1.3 Online Interface Design Process

Snap implemented a process and governance around online interface design.

5.1.4 Online Design Principles

Snap also established Online Design Principles which prohibit the use of dark patterns & misleading nudge techniques.

5.1.5 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

5.2.1 Introduction

This document outlines Snap's protocol for communicating its platform Terms and Conditions to users, in compliance with the requirements of Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (the "Digital Services Act" or "DSA"), in particular with regard to Articles 14 and 27.

Snap publishes Terms and Conditions with concise summaries in clear, easily understandable, unambiguous language, in a publicly available, easily accessible, and machine-readable format. Snap's Terms and Conditions include detail on: how use of the service may be restricted; content moderation policies and procedures; information on the use of algorithms, the parameters and criteria behind recommender system functioning, and how to adjust them; instances when user access and/or content may be restricted, suspended or terminated; and instructions on the internal complaint handling system. This information is primarily provided in Snap's Terms of Service, which are translated into all official EU member state languages.

In addition to the Terms of Service, Snap also publishes Community Guidelines, Privacy Policies, Product Specific Terms and Advertising Policies. Snap's Community Guidelines elaborate on restrictions of the use of Snapchat related to: Sexual Content, Harassment & Bullying, Threats, Violence & Harm, Harmful False or Deceptive Information, Illegal or Regulated Activities, and Hateful Content, Terrorism, and Violent Extremism. Snap's Terms of Service prohibit any use of Snapchat to conduct illegal activities.

5.2.2 Terms and Conditions

Snap publishes a series of policies that make up our Terms and Conditions:

- Terms of Service
- Community Guidelines
- Privacy Policy
- Advertising Policies

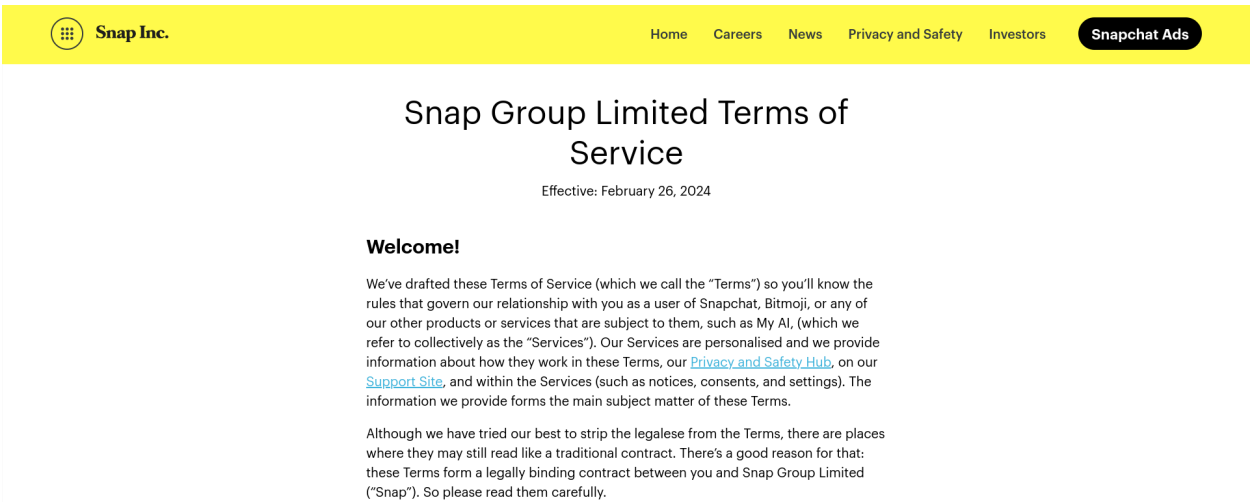
Each of our Terms and how they mitigate each of the DSA risk categories is explained below.

Terms of Service

Snap publishes its Terms of Service on the website. It is easily accessible via search engine and machine readable.

The Terms of Services include information on:

- Restrictions imposed on use of services (these are elaborated upon in Snap's Community Guidelines, Privacy Policy, and Advertising Policy)
- Policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review
- Rules of procedure of Snap's internal complaint handling system and available remedies and redress mechanisms.
- Main parameters used in Snap's recommender systems.



Snap Inc.

Home Careers News Privacy and Safety Investors Snapchat Ads

Snap Group Limited Terms of Service

Effective: February 26, 2024

Welcome!

We've drafted these Terms of Service (which we call the "Terms") so you'll know the rules that govern our relationship with you as a user of Snapchat, Bitmoji, or any of our other products or services that are subject to them, such as My AI, (which we refer to collectively as the "Services"). Our Services are personalised and we provide information about how they work in these Terms, our Privacy and Safety Hub, on our Support Site, and within the Services (such as notices, consents, and settings). The information we provide forms the main subject matter of these Terms.

Although we have tried our best to strip the legalese from the Terms, there are places where they may still read like a traditional contract. There's a good reason for that: these Terms form a legally binding contract between you and Snap Group Limited ("Snap"). So please read them carefully.

Restrictions Imposed

The following restrictions are included in Snap's Terms of Service (as at August 2024):

Who Can Use the Services

Our Services are not directed to children under the age of 13, and you must confirm that you are 13 years or older to create an account and use the Services

Respecting the Services and Snap's Rights

You must also respect Snap's rights and adhere to the Snapchat Brand Guidelines, Bitmoji Brand Guidelines, and any other guidelines, support pages, or FAQs published by Snap or our affiliates.

Respecting Others' Rights

You therefore may not use the Services, or enable anyone else to use the Services, in a manner that violates or infringes someone else's rights of publicity, privacy, copyright, trademark, or other intellectual property right.

Safety

By using the Services, you agree that you will at all times comply with these Terms, including our Community Guidelines and any other policies Snap makes available in order to maintain the safety of the Services.

Content Moderation

The following Content Moderation information is included in Snap's Terms of Service.

Much of the content on our Services is produced by users, publishers, and other third parties. Whether that content is posted publicly or sent privately, the content is the sole responsibility of the user or entity that submitted it. Although Snap reserves the right to review, moderate, or remove all content that appears on the Services, we do not review all of it. So we cannot — and

do not — guarantee that other users or the content they provide through the Services will comply with our Terms, Community Guidelines or our other terms, policies or guidelines. You can read more about Snap's approach to content moderation on our Support Site.

Users can report content produced by others or others' accounts for violation of our Terms, Community Guidelines or other guidelines and policies. More information about how to report content and accounts is available on our Support Site.

We may restrict, terminate, or temporarily suspend your access to the Services if you fail to comply with these Terms, our Community Guidelines or the law, for reasons outside of our control, or for any other reason. That means that we may terminate these Terms, stop providing you with all or any part of the Services, or impose new or additional limits on your ability to use our Services. For example, we may deactivate your account due to prolonged inactivity, and we may reclaim your username at any time for any reason. And while we'll try to give you reasonable notice beforehand, we can't guarantee that notice will be possible in all circumstances.

Before we restrict, terminate or suspend your access to the Services, we will take into account all relevant facts and circumstances apparent from the information available to us, depending on the underlying reason for taking that action. For example, if you violate our Community Guidelines we consider the severity, frequency, and impact of the violations as well as the intention behind the violation. This will inform our decision whether to restrict, terminate or suspend your access to the Services and, in the event of suspension, how long we suspend your access. You can find out more about how we assess and take action against misuse of our Services on our Support Site.

Internal Complaint Handling

The following information on Snap's Internal Complaint Handling is included in Snap's Terms of Service.

Where we restrict, terminate or suspend your access to the Services for violation of our Community Guidelines, we will notify you and provide an opportunity for you to appeal.

We hope you'll understand any decisions we make about content or user accounts, but if you have any complaints or concerns, you can use the submission form available here or use available in-app options. If you use this process, your complaint must be submitted within six months of the relevant decision.

Upon receiving a complaint, we will:

- *ensure the complaint is reviewed in a timely, non-discriminatory, diligent and non-arbitrary manner;*
- *reverse our decision if we determine our initial assessment was incorrect; and*

- *inform you of our decision and of any possibilities for redress promptly.*

Recommender Systems
The following information on Snap's Recommender Systems is included in Snap's Terms of Service.

Our Services provide a personalized experience to make them more relevant and engaging for you. We will recommend content, advertising and other information to you based on what we know and infer about your and others' interests from use of our Services. It is necessary for us to handle your personal information for this purpose, as we explain in our Privacy Policy. You can find more information on personalized recommendations on our Support Site.

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. The Community Guidelines are easily accessible via Search Engine and in Snap's Transparency Center and are machine readable and easily understandable.

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 2024) 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 and 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 and 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, or Violent Extremism

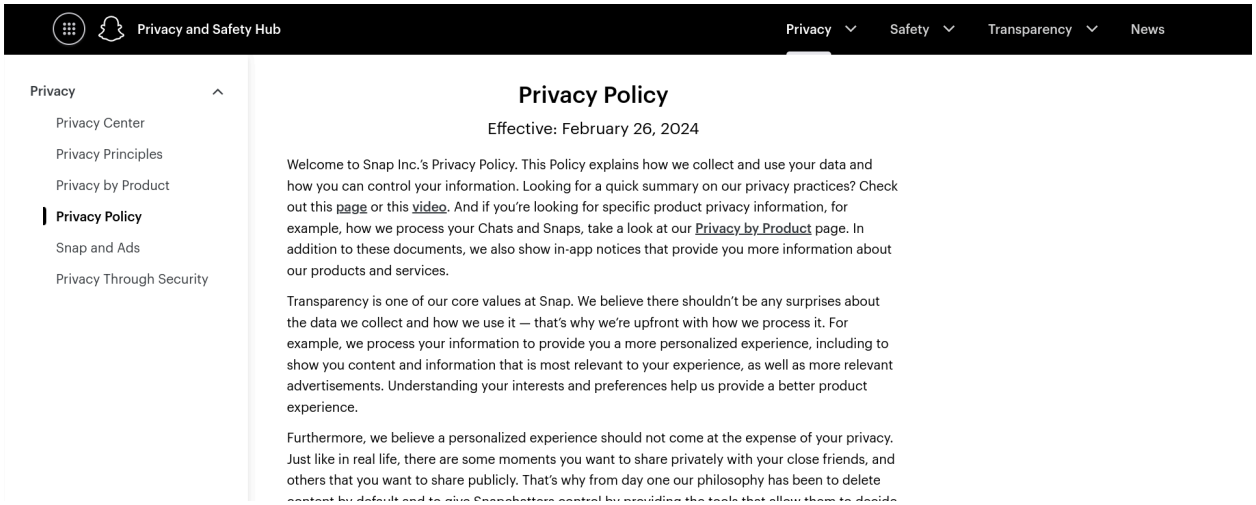
- *Terrorist organizations, violent extremists, and hate groups are prohibited from using our platform. We have no tolerance for content that advocates or advances terrorism or violent extremism.*
- *Hate speech or content that demeans, defames, or promotes discrimination or violence on the basis of 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.

Privacy Policy

Snap also publishes a Privacy Policy, which is easily accessible via Search Engine as well as within our Privacy Center and in the footer of Snap websites as well as in Apple/Android stores. It is machine readable and easily understandable.



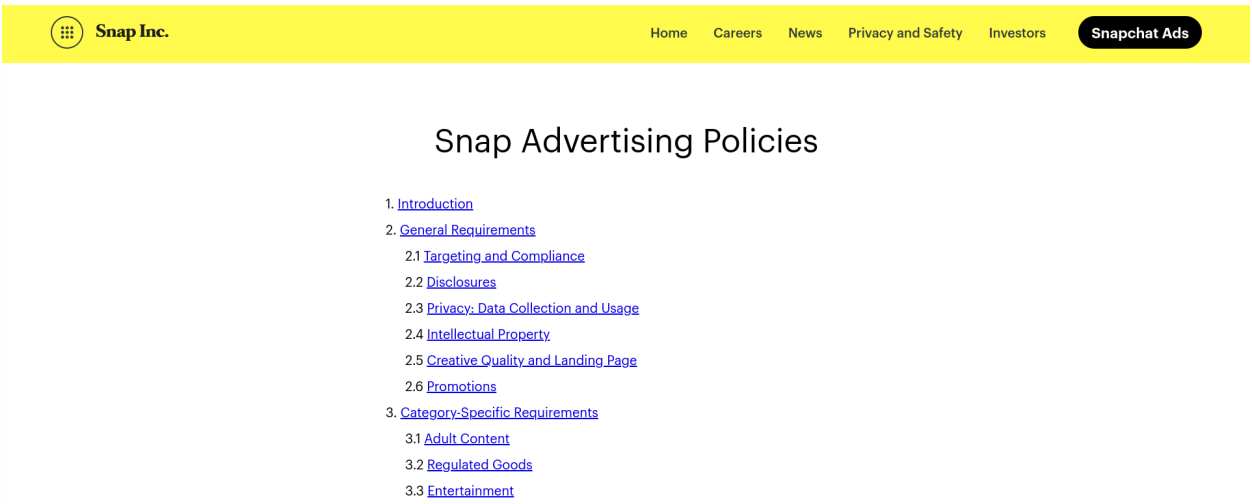
Our Privacy Policy provides a detailed description of our privacy practices, including an explanation of how Snap collects and uses personal data and how individuals can control their information. We recognize that policies and terms can be overwhelming documents. That is why our long standing approach 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. In addition to these documents, we also show in-app notices that provide more information about our products and services.

Advertising Policies

Snap also publishes Advertising Policies, which outline the terms and conditions for use of Snap advertising services. These are easily accessible via Search Engine, machine readable, and easily understandable.



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 update advertisements will run.

5.2.3 Product 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).

Discover

We have specific publishing agreements with our premium partners that post content on Discover, such as media organizations and Snap Stars, that require them to abide by our Terms (including our Community Guidelines).

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).

5.2.4 Other Aspects

Oversight and Administration

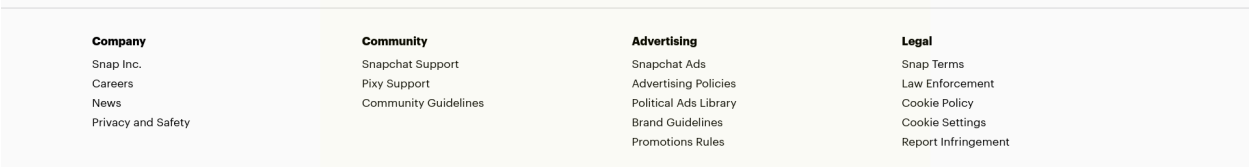
Change Management
Snap's Management body (including Legal and Public Policy team stakeholders) review and update various Snapchat Terms and Conditions (including Community Guidelines, Terms of Service, Privacy Policies, and Advertising Policies) for additional information that may result in impact on our risk and mitigation assessments, and to ensure that the Terms/Guidelines accurately reflect the contractual relationship and/or other obligations between users of the respective Services and Snap and adhere to applicable legal requirements.

Snap's Legal team has a formal process in place to make and track changes to the Terms and Conditions and to communicate key changes to stakeholders in a timely manner.

Snap's Terms are regularly reviewed (and updated as needed) by Legal and Policy team stakeholders to ensure that they accurately reflect the contractual relationship and/or other obligations between users of the respective Services and Snap and adhere to applicable legal requirements. Once the document is finalized and approved, it is then localized in all supported languages, including all official languages of the European Union as explicitly required by the DSA. To the extent the changes to the terms, policies, or guidelines are determined to be material, Snap will provide users with reasonable advance notice.

Snap provides an in-app pop-up to notify recipients of the service of material changes to the Terms and Conditions.

Accessing Terms and Conditions



Accessibility

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 site. This is where Snap publishes formal transparency reports.

The rationale behind the integration of these three centers 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 (various pages, including the Privacy Policy), advertising (Snap and Ads), and specific products (Privacy by Product).

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.

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.

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.

Languages

Our Terms of Service have been translated into all official languages of the European Union as explicitly required by the Digital Services Act.

Readability

As outlined in Snap's 2023 Risk Assessment and Mitigations Report, our European Snapchatter community consists of a diverse range of ages and genders. Snapchat services are not primarily directed at or used by minors. While Snapchat does have a young demographic, only a small

percentage of European Union users fall within the 13-17 age category. The largest age category is 18-24.

Snap's Terms have been designed to be a concise summary in clear, easily understandable, unambiguous language, in EU member state languages, in a publicly available, easily accessible, and machine-readable format., including summaries and explainers. This helps all users to understand what activity is prohibited on Snapchat and the consequences, which reduces the likelihood they will engage in illegal or violating activity.

In addition, our Privacy and Safety Hub and Support pages as explained in the Transparency part of our Report have also been designed to be user friendly and easily understandable. For example, we created our Privacy and Safety Hub, with pages such as our Privacy by Product page, to give Snapchatters a high-level summary of our privacy and safety practices across each of our products and features. We also created a video to visualize our privacy practices, and use icons and other best practices as recommended by privacy and safety experts and the recognised Age Appropriate Design Codes. This helps all users to understand how Snapchat works, what options they may have, how we moderate and enforce our terms and how they can get support. This reduces both the likelihood of illegal or violating activity and the severity of harm in the event they are exposed to illegal or violating activity despite our limitations. Teens reading our Privacy Center can understand how their data is being processed by Snap and find more information about relevant privacy settings which reduces the likelihood and severity of negative effects on Teens' data protection rights.

5.2.5 Conclusion

Snap provides terms and conditions for the recipients of its services, which incorporate the content and meet the language requirements of the DSA.

As explained in Section 4, we have concluded that Snap's terms and conditions, 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.

5.3 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.

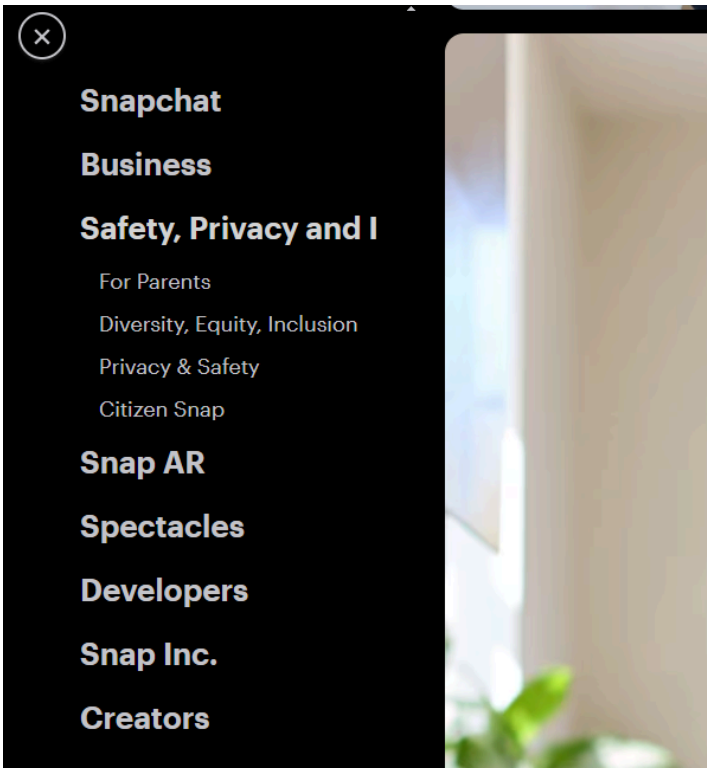
Information provided to users can be divided into three categories:

- Information we provide on our website;

- Information provided in app stores; and
- Information we provide in our application.

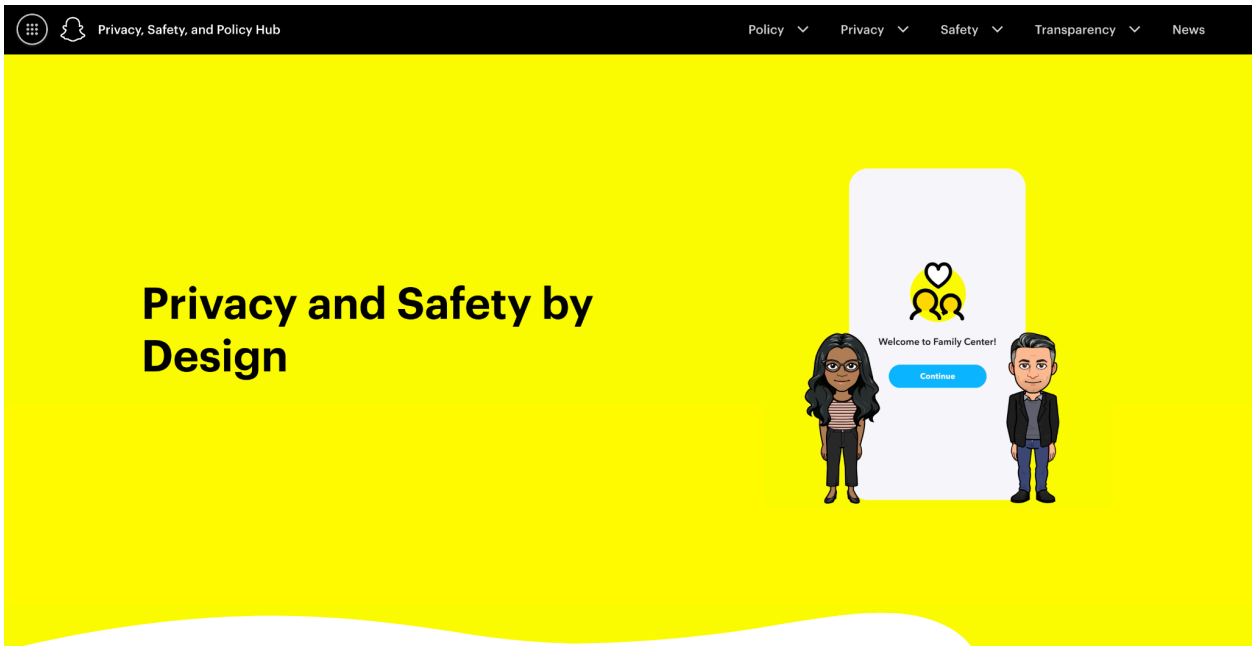
5.3.1 Information we provide on our website

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, Safety, and Policy Hub and our Support Center.



Privacy, Safety, and Policy Hub

Snap's Privacy and Safety Hub was launched in 2022 and combines our Privacy Center, Safety Center and Transparency Center all under one umbrella, and in 2024 we added a dedicated Policy Hub to this center. 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 Policy, 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.

Policy Center

We want Snapchat to be a safe and positive experience for everyone who uses our platform or products. For this reason, we created rules and policies that explain the rights and responsibilities of all members of our community. Our Policy Center provides a central place for our Community Guidelines, Advertising Policies, Content Guidelines, and Commercial Content Policy. We have also reformatted our policies in a way that is easier and more intuitive for users to navigate, replacing our previously long, text-heavy pages with shorter, more digestible segments with clear headings and organization by subject matter.

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 Center we state the following:

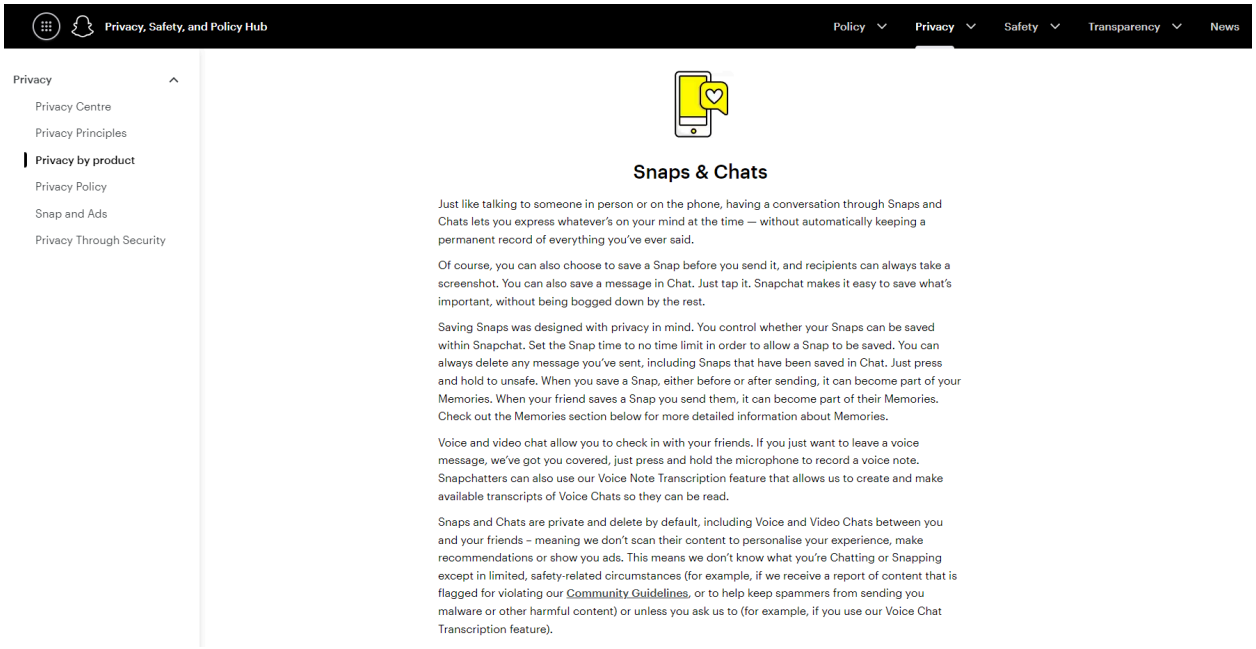
> *"Privacy policies tend to be pretty long – and pretty confusing. That's why we did our best to make our Privacy Policy brief, clear, and easy-to-read!*

You should read our entire Privacy Policy, but when you only have a few minutes or want to remember something later on, you can always take a look at this summary – so you can learn or recall some of the basics in just a few minutes."

Since our 2023 Report, we have made additional updates to our Privacy Policy in February 2024. This update included in particular the following changes:

- The addition of summaries to each section of the Privacy Policy to highlight key takeaways.
- Provided more detail on how Snapchatters can control their information, who they can share content with, and who can contact them.
- Added additional transparency on how we collect and use Snapchatter information and content, with examples of what data we use for purposes like personalization, advertising, and building machine learning models.

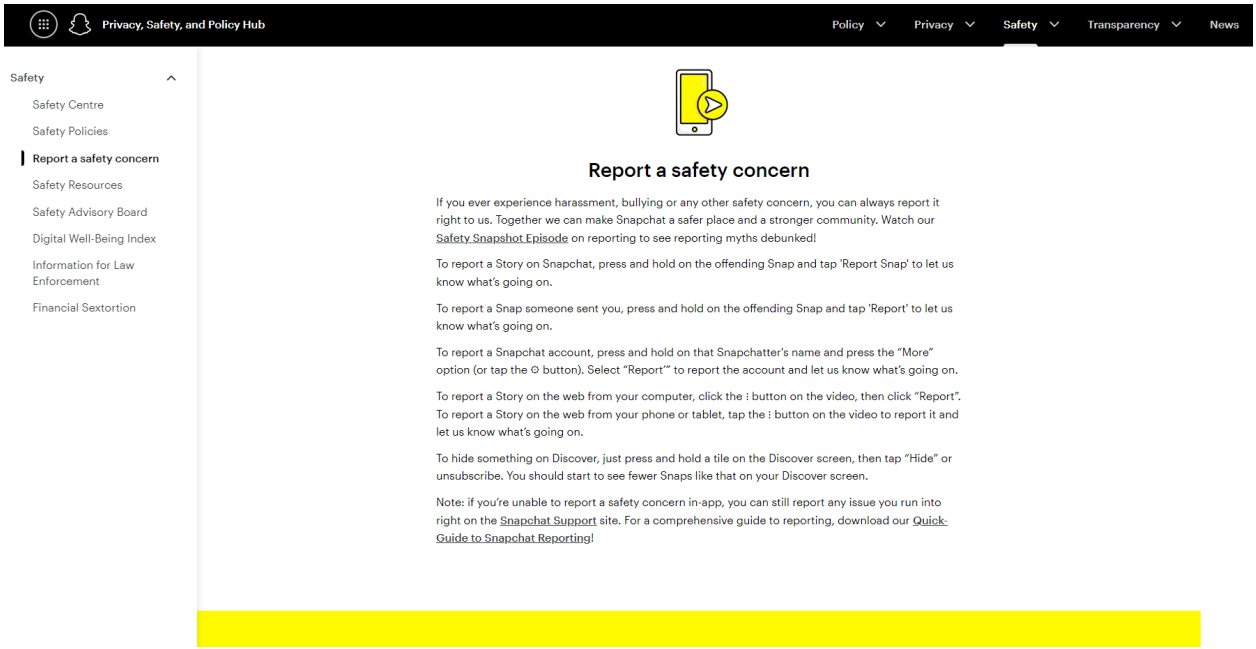
Privacy by Product gives users concise and easily understandable information about our products. For example, this webpage provides an overview of our approach to Snaps & Chats, as well as hyperlinks to more detailed information on specific aspects. Similarly, there's a section on Spotlight, Lenses,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



Safety Center

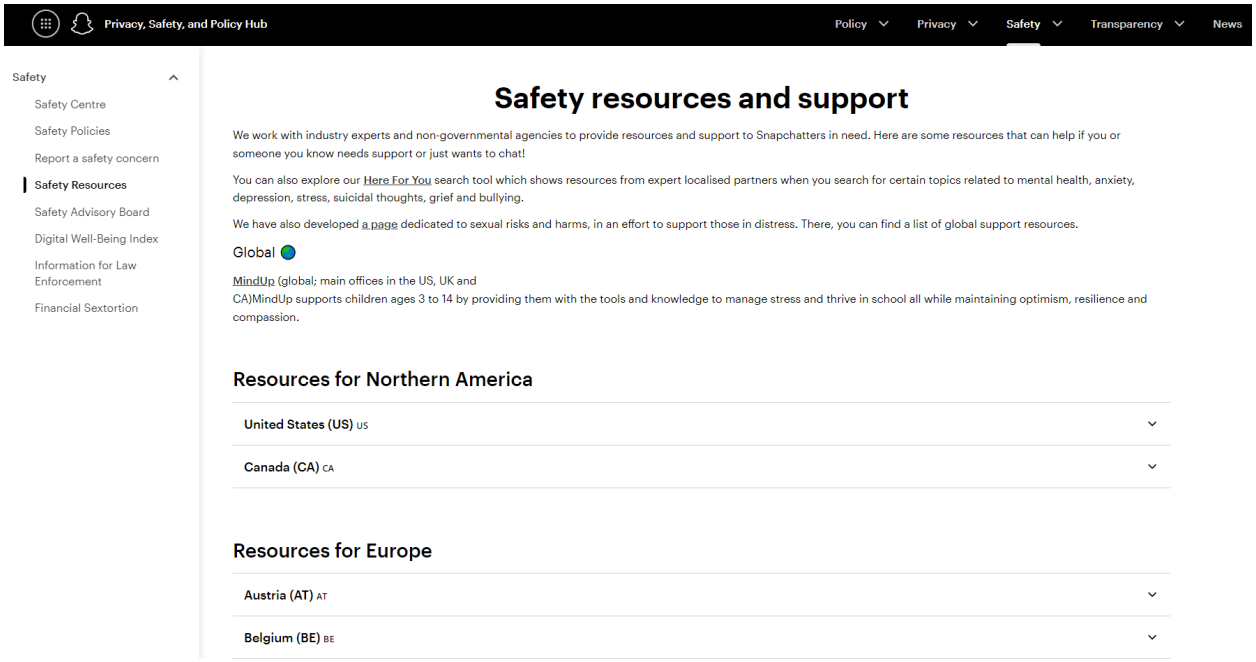
Our Safety Center provides an overview of our Safety resources, including tips on how to report content, the acknowledgment that safety is a shared responsibility, as well information on our Trusted Flagger Program, Safety Advisory Board, Digital Well-Being Index and more. Again, the goal here is to provide easy to navigate and process information. Since our 2023 Report, we have included direct links to our existing pages for the Safety Advisory Board, Digital Well-Being Index and Information for Law Enforcement, as well as a new support page on Financial Sextortion, in the navigation bar of the Safety Center. In the near future, we'll also add a page for the Council for Digital Well-Being.

We have dedicated a page on our Safety Center to reporting. 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. The dedicated page 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. Since our 2023 Report, we have also included additional resources including:

- a page dedicated to explaining Financial Sextortion; and
- a page dedicated to sexual risks and harms, in an effort to support those in distress.



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) 🇫🇷

E-Enfance
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.

Suicide Écoute
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.

SOS Suicide Phénix
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.

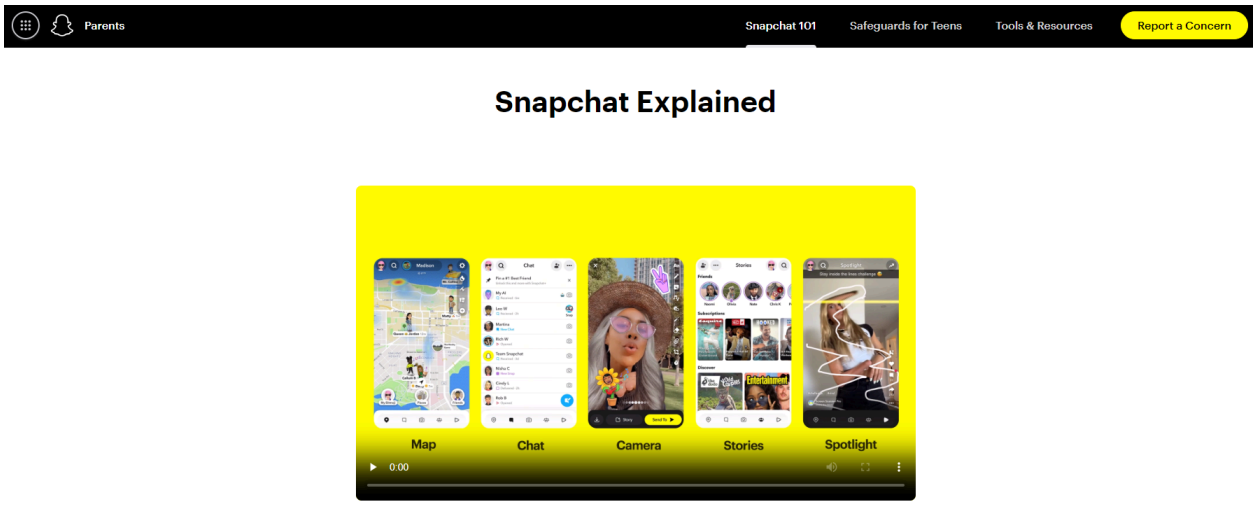
Germany (DE) 🇩🇪

Greece (GR) 🇬🇷

Parents

In September 2023, we launched a dedicated microsite: parents.snapchat.com to provide even more information for parents. 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, the dedicated microsite contains: an updated Snapchat 101, a specific page on Safeguarding for Teens, a Tools & Resources section. This has replaced the 'Parents and Educators' section of the Safety Centre.

The Snapchat 101 page incorporates our previous 'Parent's Guide to Snapchat' but lays it out in an accessible manner (including a short video).



The Safeguards for Teens page summarises the key protections for teens

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.

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.

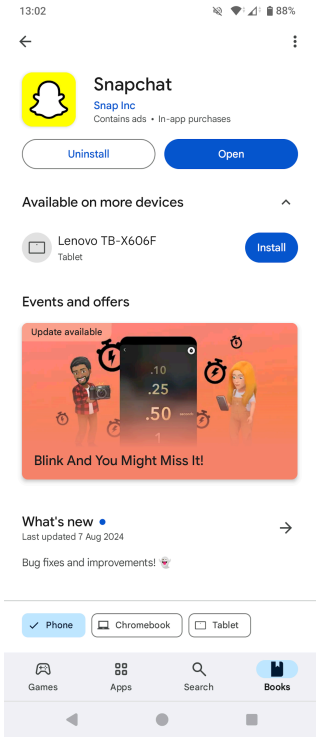
Since 2015, we have also been publishing Transparency Reports twice a year, to provide insight into Snap's safety efforts and the nature and volume of content reported on our platform. We are committed to continuing to make these reports more comprehensive and informative to the many stakeholders who 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.

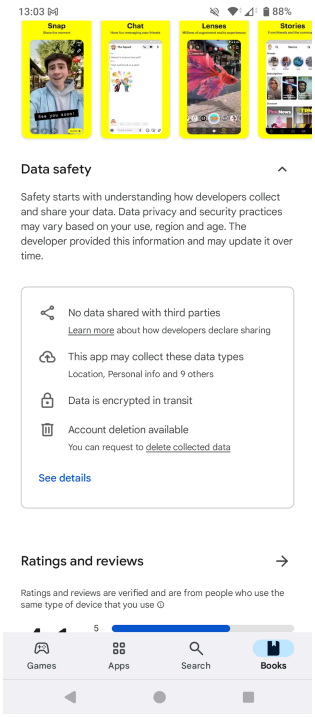
News Page

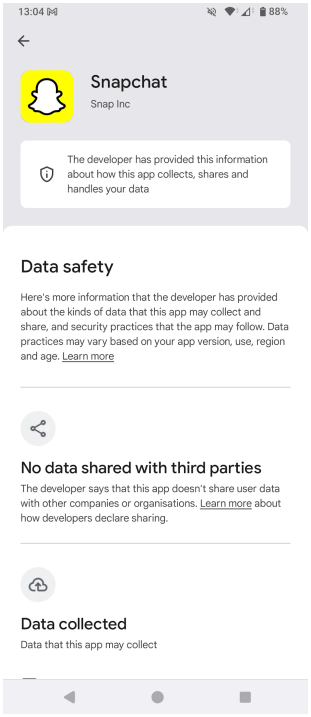
Snap also frequently publishes Privacy and Safety related information on the Hub's News webpage. The purpose of these news articles is to inform the general public about recent developments on issues relevant to privacy, safety and transparency on Snapchat. For example, a recent article introducing Snap's Inaugural Council for Digital Well-Being, or an article informing the public on Snap's approach to keeping its community safe during the 2024 Paris Olympics.

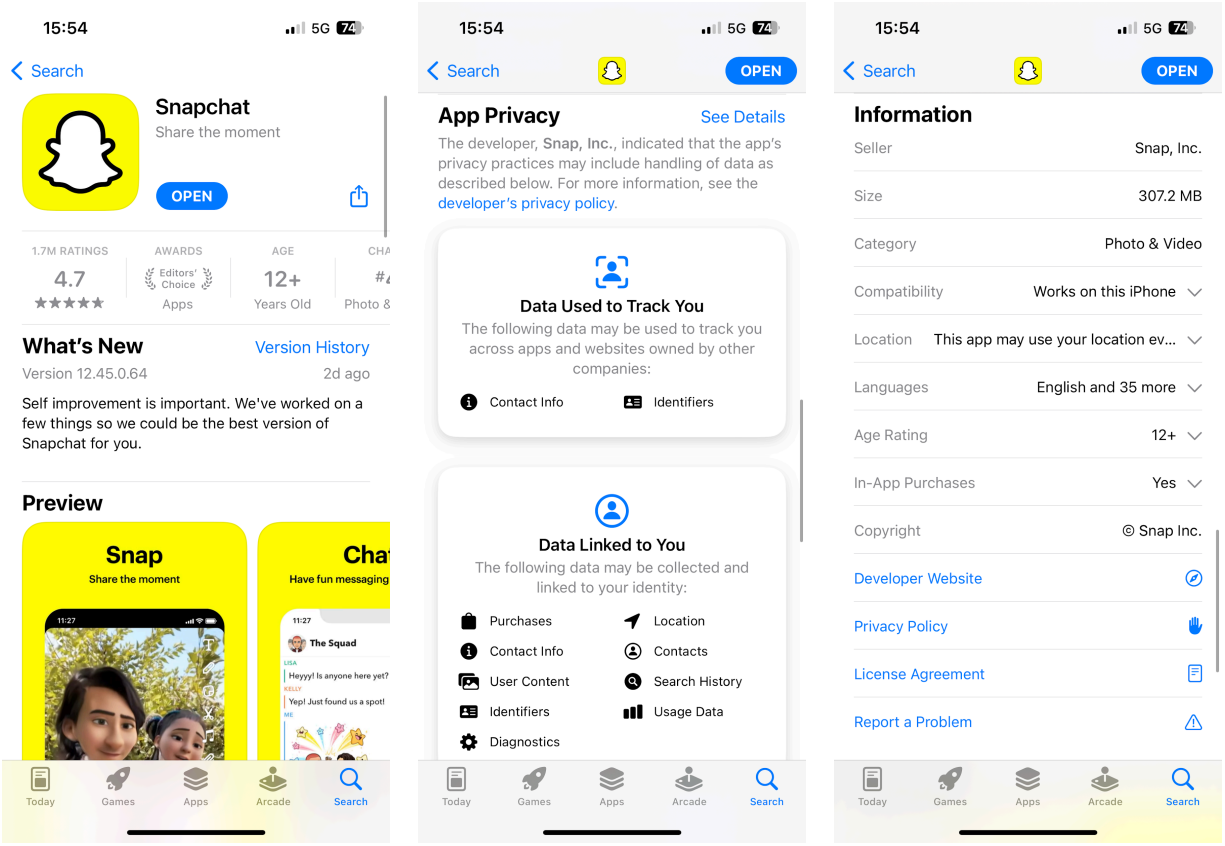
5.3.2 Information provided in app stores

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.









5.3.3 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 steps are optional. 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 Child[69] and UK Age-Appropriate Design Code[70] 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.[71] 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.

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.

---

[69] UN OHCHR, *Convention on the Rights of the Child,* url.
[70] UK Information Commissioner's Office, Introduction to the Children's code, url.
[71] See for example the report on age verification issued by the Australian eSafety Commissioner, url.

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. Below we provide some examples of JITs that create feature-specific awareness.

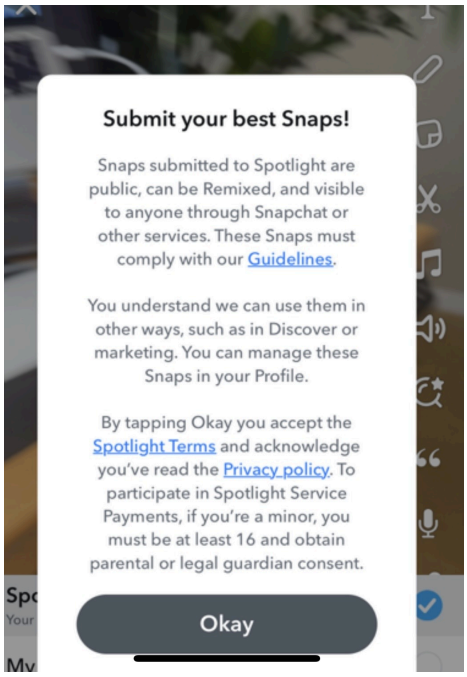
Snap Map
Snap Map is off by default and user location data is off by default. Users choose who can see their location.

Since our 2023 Report, we have simplified location sharing on Snap Map. Location sharing still requires a two step permission to enable location sharing:

1. Users will choose location permissions in device settings (e.g. for iOS these are “Never”, “Ask Next Time or When I Share”, “While Using the App”, “Always”).
2. Users must also select sharing with “My friends”, “My Friends, Except” or “Only these Friends”.

However, users no longer have to navigate two separate flows for “live sharing” and “last active sharing”. The selected Friends will see “live location" if the user chose “Always” in device settings or “last active” if the user chose “While Using the App” in device settings. Users can still decide to turn on Ghost Mode at any time when they want to go off the grid.

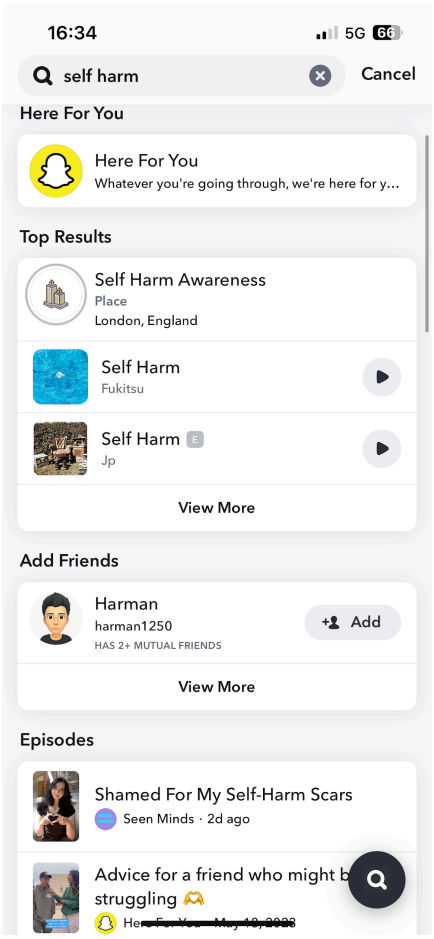
Spotlight
Before a user submits a Snap to Spotlight they are presented with a JIT informing them that Spotlight submissions are public. This is to create awareness that Spotlight is different from My Story submissions, which can be shared with friends only, unless the user actively chooses to share them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

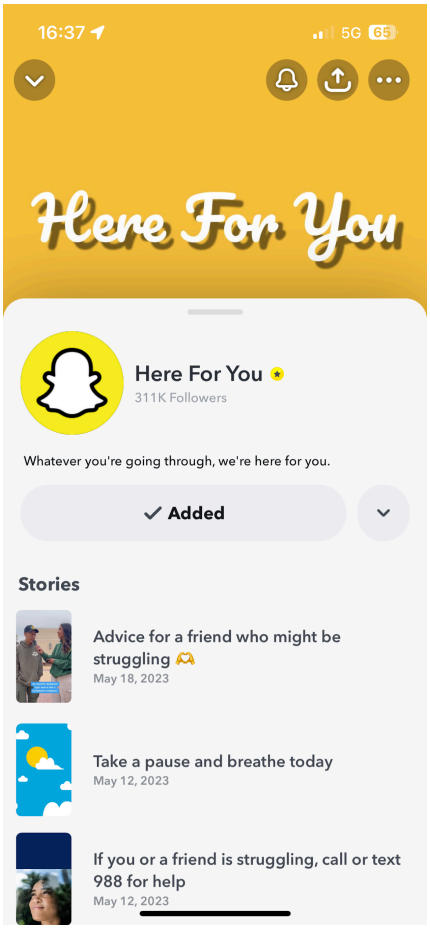


Thematic awareness and notices

Across Snapchat, we offer a number of resources to users to raise awareness on safety topics and protect them. For example:

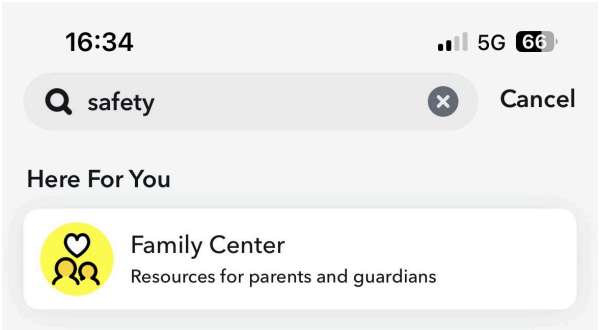
1. **Here for you**
   If a user types in "selfharm" or related terms in our Search functionality, we try to prominently show them relevant 'Here For You' resources among the search results.

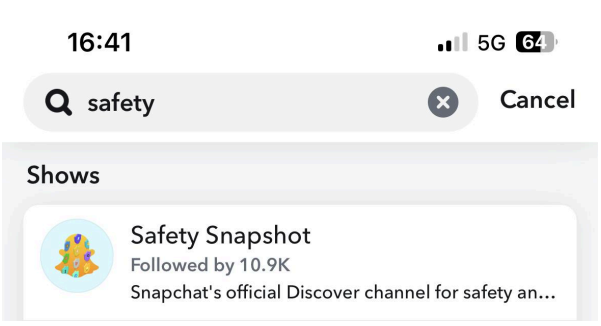




2. **Safety**

   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.

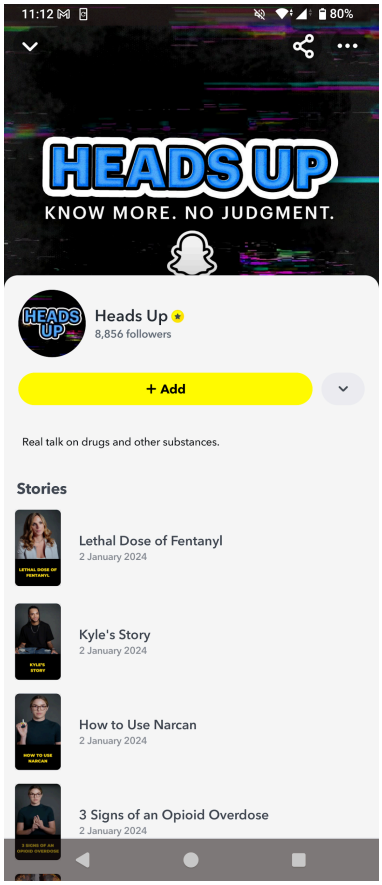




3. **Heads up**

   If a user types in terms related to illicit drugs in our Search functionality, we try to prominently show them our ‘Heads Up’ resources among the search results. Heads up is

our in-app tool that surfaces educational content from experts to Snapchatters if they try to search for drug-related content. Our expert partners include the Centers for Disease Control and Prevention (CDC), the Substance Abuse and Mental Health Services Administration (SAMHSA), Community Anti-Drug Coalitions of America (CADCA), Shatterproof, Truth Initiative, and the SAFE Project.



We also run campaigns on Snapchat to raise awareness about certain themes. For example, on Global Data Privacy Day 2024, we informed the general public about our new Privacy Policy, announced the updated parents guide to Snapchat, and launched a dedicated page on Privacy through Security, as relaunched our interactive Lenses with tips on how to stay safe online.

Similarly, on Safer Internet Day 2024, we raised awareness around parents' options to participate and monitor their child's online activities through Snapchat's Family Center. We publish updates on efforts and 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.

5.3.4 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.

5.3.5 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 Content Moderation

5.4.1 Approach

Across Snapchat, we're committed to advancing safety while respecting the privacy and freedom of expression 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.

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.

Snapchat Design and Function

As a reminder, we have also designed Snapchat with privacy and 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, "Discover" content on the Stories tab, Lenses, and Advertisements.
2. "Private content" is distributed to friends/followers. Private content includes Private Stories, Chat, Groups, Accounts.

More information about our mitigations relating to Snapchat's design and function can be found in Section 5.1 (Snapchat Design and Function).

Community Guidelines and Terms of Service

When considering our Content Moderation approach, it is also important to bear in mind that 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. More information about our mitigations relating to Terms and user awareness can be found in Section 5.2 (Terms) and Section 5.3 (Transparency).

Content Moderation

Our content moderation processes assess each piece of content against the above Terms, policies and guidelines to determine if that content is compliant. Main policy violation categories include: Sexual Content, Harassment & Bullying, Threats, Harm & Violence, Harmful, False & Deceptive information, Illegal or Regulated Activities, and Hateful Content, Terrorism, and Violent Extremism.

We assess content with proactive moderation using a combination of automated tools and human review to moderate content recommended for broad distribution on Snapchat, such as content in Spotlight, Discover, Lenses and Advertisements and non broadcast content such as stories. Technology review and our Machine Learning models are built based on our policies, so policies are applied consistently across both automoderation and human review. 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.

More information about our mitigations relating to content moderation can be found in the following paragraphs of this Section 5.4.

Enforcement

We provide in-app and web-based reporting tools that enable EU users to report content and accounts they think violate our Terms, which expressly prohibit the dissemination of harmful and illegal content on Snapchat. We also have mechanisms enabling non-users in the EU to report content on Snapchat they believe is illegal. We respond to user reports quickly, and we use feedback to improve the content experience for all Snapchatters. User reports are classified by the reporting reason for moderator review. More information about our mitigations relating to enforcement can be found in Section 5.5 (Enforcement).

5.4.2 Content Moderation

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 violates 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.

Proactive Moderation (Content Reviews)

We use a combination of automated tools and human review to proactively moderate broadcast content across Snapchat, i.e., content recommended for broad distribution on Snapchat, such as content in Spotlight, Discover, Lenses and Advertisements and non broadcast content such as stories.

5.4.3 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

5.5.1 Introduction

We strive to continuously update and improve our enforcement mechanisms to protect Snapchatters and our broader communities. 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 4 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.

5.5.2 Protections against Misuse (Art. 23)

Snap has implemented a Strike Roadmap to perform enforcement actions against users who repeatedly post illegal content or violate the Terms and Conditions. The enforcement action is determined based on the severity of the violation in a given timeframe.
Snap's Trust and Safety team has a process to review and update, if necessary, the Strike Roadmap.

5.5.3 Transparency for Measures of Protection

Snap publishes information within its Terms and Conditions information regarding its policies and enforcement mechanisms related to misuse of the platform, including the following:

- Snap's Community Guidelines related to Illegal Content
- Snap Illegal Content Explainer
- Snapchat Moderation, Enforcement, and Appeals

5.5.4 Notification of Criminal Offenses (Art. 18)

Proactive referrals to law enforcement and governmental agencies

As discussed above, we have various tools within the app where individuals may report illegal or harmful activity to our Trust & Safety team. The Trust & Safety team then will investigate the report and if needed, take action, which can range from deleting the content and warning the user to locking the violating account.

In the event the report concerns any threat to a person's life or safety, Snap will alert the appropriate authorities.

Snap's Content Moderators are trained to preserve relevant information, including user identifying information, and content, regarding the involvement of law enforcement.

Law enforcement takedown requests (Articles 9 and 10)

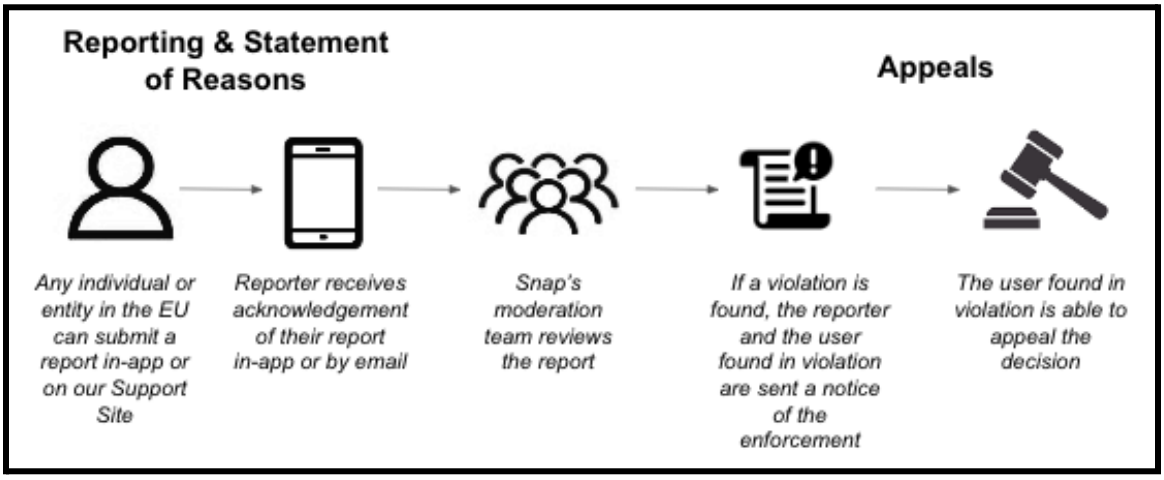
EU Law Enforcement ("LE") may submit orders for takedown of content or accounts (Article 9) via email to a dedicated email address.

Similar to the process for takedown requests, EU LE can submit specialized requests to provide information (Article 10) via email to a dedicated email address.

5.5.5 Complaint Handling System (Art. 20)

Snap provides users to access the internal complaint handling system where they are able to appeal content removals or account takedowns dependent on the relevant policy and type of content violation. Snap's internal complaint handling system is configured to record the submission of complaints for a 6 month period.

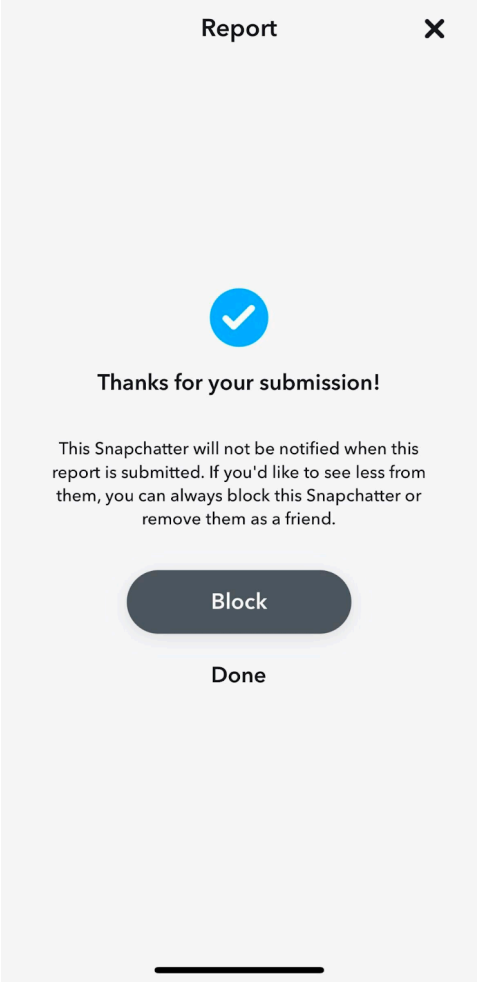
Statement of Reasons (Art. 17)



We provide further detail regarding each step of this reactive moderation process below.

Notice to Reporter

After a reporter reports a piece of content or an account for violating our Terms or causing harm, we will immediately send them a notice confirming receipt of their report and assuring them that we are investigating it.



We endeavor to resolve all reports as quickly as possible while ensuring that we do a thorough review and achieve the correct result. Some reports can be resolved much more quickly than others, which may be more nuanced and require escalation and/or input of other teams. However, on an average basis, our Trust & Safety team resolves reports regarding particular pieces of content in two hours, and reports regarding chat and account-level concerns in 24 hours. Snap publishes information about the Median Turnaround Time to process appeals in its Transparency Report.

5.5.6 Effectiveness of Enforcement

We regularly meet with NGOs and other stakeholders to discuss our measures and generally receive positive feedback. For example, in a meeting in January 2024 with Child Focus, a Belgian Safer Internet Centre operating hotline, helpline and awareness centre, Snap was praised for its responsiveness. They stressed that our system works well and we act fast when it comes to

taking down illegal content and escalating and that they are pleased with their collaboration with us.

Since our 2023 Report, there are now no illegal or other violating content categories in our low likelihood category and we have observed a further decrease in our overall PVP compared to last year. See Section 6.4 (Prevalence Testing) for further details.

5.5.7 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 continued drop in Policy Violating Prevalence (PVP) on Snapchat overall since our 2023 Report. In other words, as we get better at detecting and enforcing against an increased number of violations, the frequency of violations found on Snapchat continues to decrease overall.

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.

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

5.6.1 Introduction

This section describes specific mitigation measures that Snap has put in place with regards to Snapchat's algorithmic systems for the in-scope services on Snapchat to address the risks identified in its risk assessment pursuant to Article 34(1), DSA.

In line with Article 34(1), our risk assessment is proportionate to the risks identified taking into account their severity and probability, and the design of our recommender systems and other relevant algorithmic systems. While Snapchat uses several algorithmic systems across all of Snapchat's in-scope services: Spotlight, Discover, Map, Lenses, Public Profiles, Advertising, the risks identified in our risk assessment focused on the following algorithmic systems:

1. Content Recommendation Systems in Spotlight and Discover
2. Advertising Systems
3. Content Moderation Systems

While we do have mitigation measures in place for other algorithmic systems, such as algorithmic systems relating to our Lenses and Maps products (as explained, for example, in Section 5.4 (Content Moderation)), we have therefore focused our efforts on the above content recommendation and advertising systems when considering specific mitigation measures pursuant to Art 35(1); other algorithmic systems are not considered in this Section 5.6.

The specific mitigations put in place for our Content Moderation Systems and Advertising Systems are covered in Sections 5.4 and 5.7 respectively. This Section 5.6 is therefore primarily focused on the specific mitigations put in place for our Content Recommendation Systems in Spotlight and Discover (which we refer to in this Section as the "content recommender systems").

As noted in the What's New and Introduction sections of this Report, we are considering simplifying the Snapchat application from 5 to 3 tabs. Rather than having the current separate Spotlight and Discover tabs, this envisages a single unified tab for more public content to the right of the Camera. This may also result in a unification of our recommendation systems across Spotlight and Discover. We plan to run some tests later this year to assess how it performs. This is primarily a cosmetic change, and it should not impact the mitigations including with respect to the design and function of these recommendation systems. For now, we have continued to highlight below a few differences that exist between the content recommender systems for Spotlight and Discover.

5.6.2 Content Recommendation Systems

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. Users find new content on Discover and Spotlight primarily through our algorithmic personalization/recommendation service.

Algorithmic content recommendation systems, like the one Snapchat uses, present a number of risks. For example, they may 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. We have designed our systems and processes to mitigate these risks. This includes the use of appropriate descriptive terminology, 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 as described in the testing and adaptation section of this document.

How do our Content 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.

5.6.3 Oversight and Administration

Algorithmic System Review

Snap conducts a comprehensive review of its Algorithmic Systems. The purpose of this review is to centrally catalog algorithmic systems that are significant to the functioning of Snapchat products as well as to safeguarding user safety and fundamental rights. This process is used to confirm understanding and documentation of significant algorithmic systems and review alignment of algorithmic systems with Snap's standards of care for them.

5.6.4 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 of this Report.

Summary

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 more detail:

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 Discover has a menu that allows users to report content. All reported user-generated content in Spotlight, Discover 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.

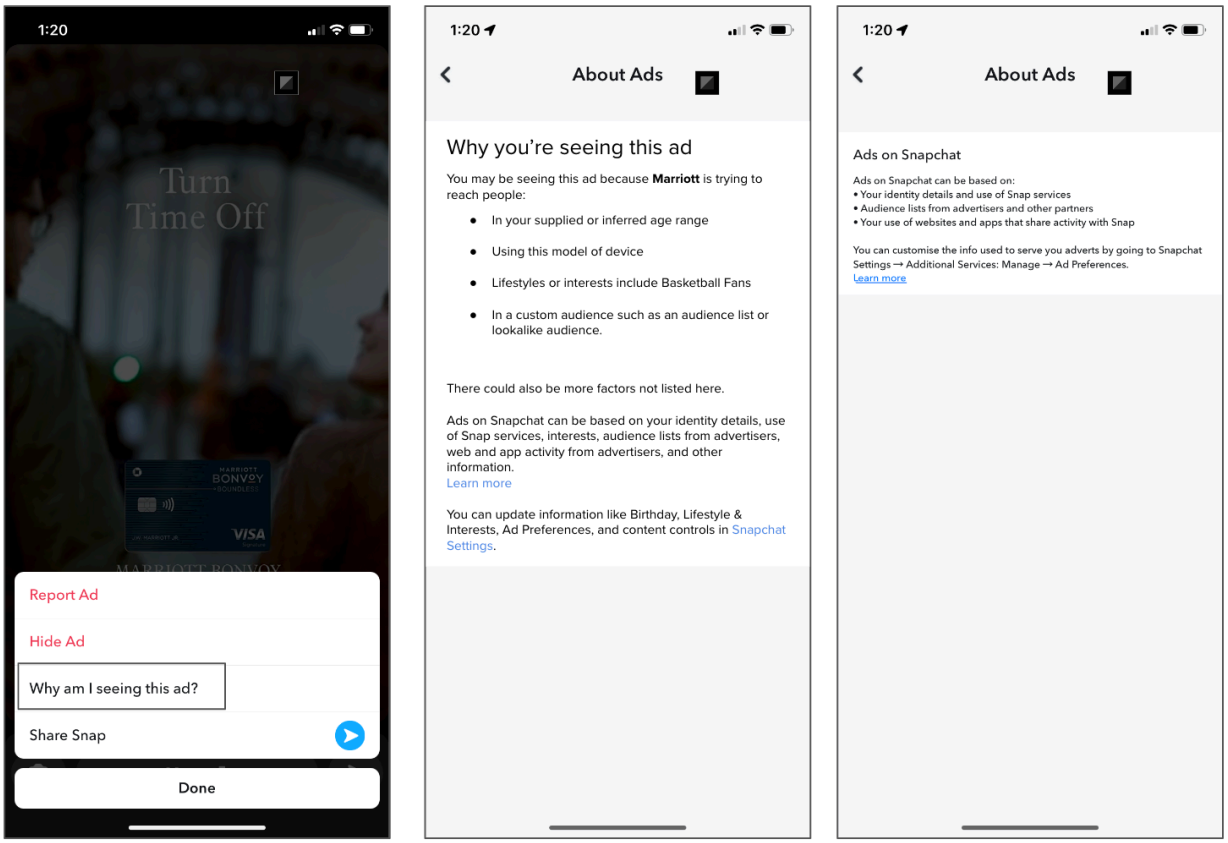
The effectiveness of these measures is tested through prevalence testing and by reviewing privacy and other consumer queries raised to our community support teams, our Data Protection Officer and our DSA Compliance Team.

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, Discover 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.



2. 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.

Intrusive personalized recommendations

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 Discover 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

Discover 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 Discover and Spotlight and then tapping 'Enable' in the 'Show More Personalized Content?' screen or by going to Settings in 'European Union Controls'.

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 content. 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 spread of illegal or false content & crisis exposure

There is a risk of rapid and widespread illegal or false content on Spotlight and Discover, 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 Discover, only approved Creators can have their Stories distributed in Discover. Those that are approved have their Stories and the 'tile' art moderated. We also monitor reporting and hide rates on both Discover and Spotlight. In recent major events / crisis situations, such as the Israel - Gaza conflict and the EU elections, these mitigations were shown to be very effective as we did not suffer significant illegal or false content on the in-scope services of Snapchat.

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 Discover, 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 Discover 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 try 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.

Erroneously excluding 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 moderated. 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 Discover and Spotlight. In addition, as explained in the Enforcement section of this Report, we have additional moderation transparency messages (statements of reason) and a more comprehensive appeals flow for moderated creators and content as part of our efforts to comply with the DSA.

Viewers could be watching but not enjoying content

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).

5.6.5 Change Management

From a high level, Change Management over algorithmic systems at Snap is governed by the previously described Privacy and Safety by Design Review process. Material updates to algorithmic systems and material changes to model pipelines, input data, and third party user data are documented and reviewed.

5.6.6 Monitoring and Quality Assurance

Performance Monitoring

Snap monitors deployed algorithmic systems for anomalies and issues and establishes alerts to notify Engineering teams when potential issues arise. These alerts look for relevant spikes or anomalies in statistics.

Quality Assurance

Snap monitors algorithmic systems related to content moderation for quality and precision on a continuous basis. Monitoring may include:

- User Reports
- User Hides
- Content removal and user appeals
- Policy Violative Prevalence (PVP)
- Content Rejection

Snap uses dashboards to visualize content moderation statistics and allows various users to establish their own alerts based on spikes in content moderation activity. Snap Engineers may also investigate spikes in user reported content or automatically detected violative content to identify correlation between model deployment to feed back into broader Engineering teams.

5.6.7 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

5.7.1 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.

5.7.2 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 aggregated demographics, including age, as well as locations, interests and device overviews of their targeted audience. User-level data is not directly available to advertisers through these dashboards.

Some of Snap's advertising tools allow advertisers to benefit from Snap's use of 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.

5.7.3 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 adverts[72]. 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 Snapchatter 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. Advertisers are also conscious about safety on Snapchat. With this in mind, in March 2024 IAS, we announced a partnership with a leading global media measurement and optimization platform[73], to:

[72] N. Fourberg e.a., on '*Online advertising: the impact of targeted advertising on advertisers, market access and consumer choice*', 2021, url.
[73] https://forbusiness.snapchat.com/blog/snap-partners-with-integral-ad-science-brand-safety.

- Conduct a measurement sample study on the advertiser suitability of our public content, specifically Spotlight and Creator Stories. In the study results, IAS found that both Spotlight and Creator content on Snapchat is 99% brand safe.[74]
- Jointly develop a new brand safety reporting solution that would give advertisers transparency into the percentage of safe and suitable content their ads are appearing against. The new solution launched in June 2024.[75]

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 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.

5.7.4 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 more detail:

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:

---

[74] "Brand Safety" is based on the GARM standard, only considering content classified at the "Floor" risk-levels . "Creator content" is image and video user-generated content posted to Public Stories; IAS sampled a wide variety of creators in the U.S. to ensure a representative sample. Spotlight content is user-generated video content that appears in the Spotlight tab on Snapchat; IAS audited content from US, CA, GB, UK, NZ, AU markets. IAS audited both Spotlight and Creator content from Oct 30, 2023 - Jan 2, 2024.

[75] https://forbusiness.snapchat.com/blog/snap-ias-solution-ga.

- 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 content. We do not store content for excessive periods solely for monetisation purposes.
- Advertisers can only use our data for ads targeting 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, UK, Norway and Switzerland), 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. All users can manage their advertising interest categories in settings and view them via our Download My Data tool (DMD). None of these SLCs are aimed at 13-17s specifically and the user-level targeting 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. We have a global team that supports ad moderation across 15+ languages and is composed of both full time employees (FTEs) and contractors. Ad Review team members are responsible for reviewing ad submissions to ensure ads abide by Snap's creative policies and technical requirements. Ad Review team members use Snap's Advertising Policies to assess compliance. Ads must comply with Snap's Community Guidelines and Advertising Policies in order to be approved. Grey area ads are discussed with Snap's Legal and Policy teams. Depending on the seniority, members of the Ad Review team also collaborate with the Sales team to create a consistent review experience for our Snapchat partners.

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.

We aim to ensure that all advertising review is maintained within a 24 hour SLA window from when the advertisement is created by the advertiser. More information on Snap's ad review process, including the timeliness of review, are located on Snapchat's For Business website.

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 or 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.

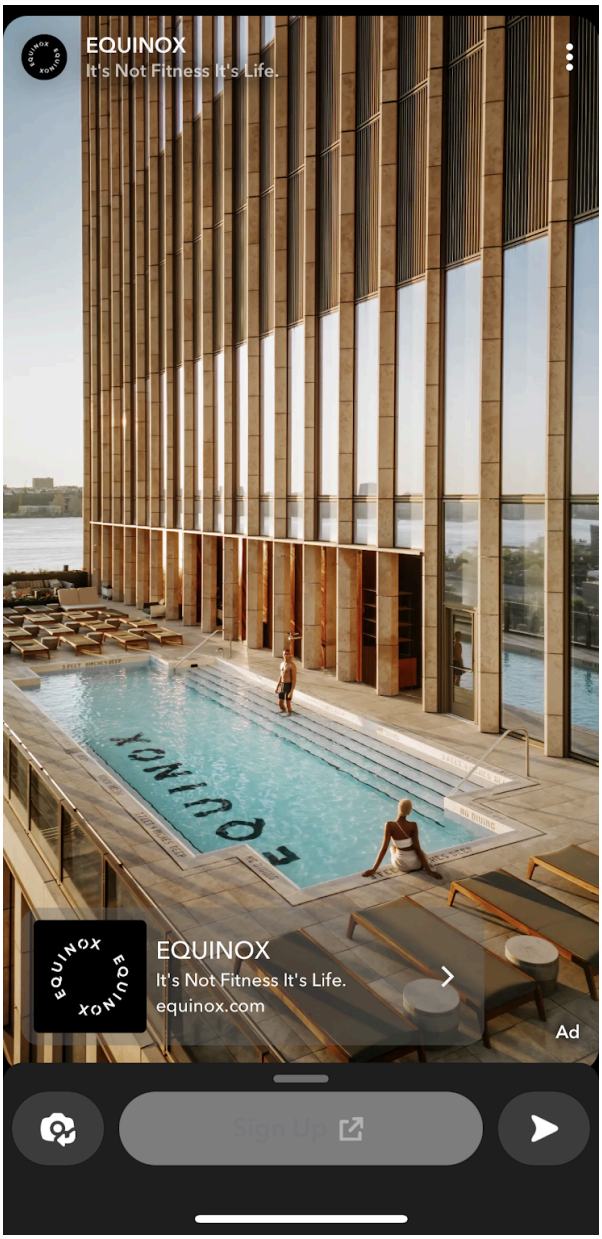
Ad Markers

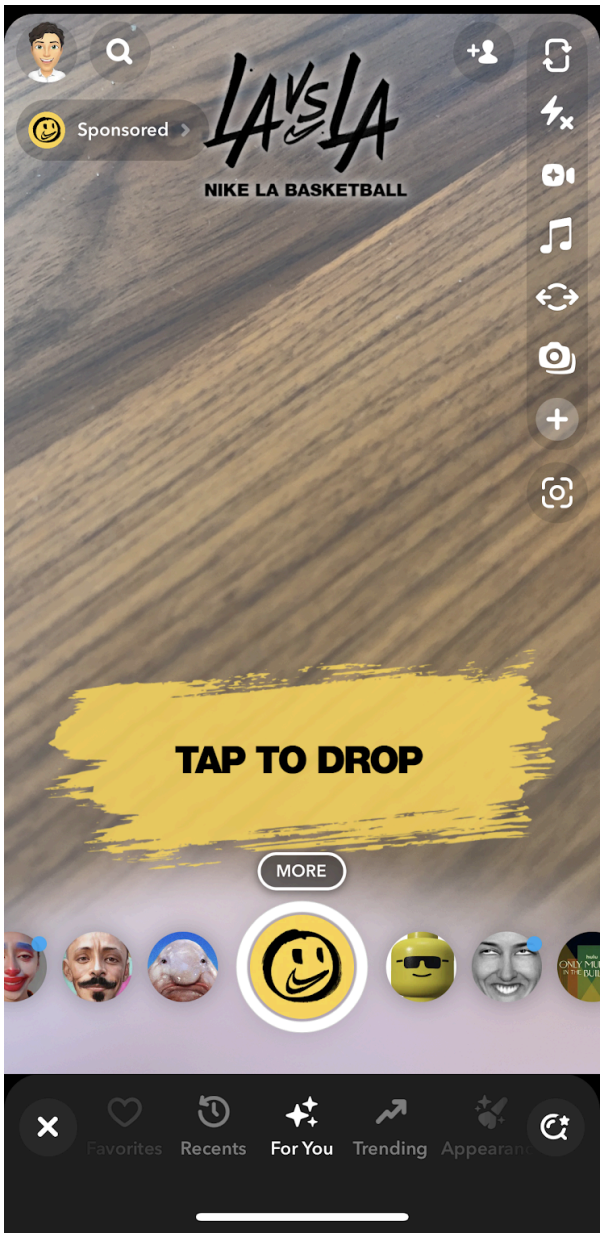
If users are not aware when content is an ad or sponsored or other commercial content, there is a risk that without additional mitigations that this may lead to confusion, deception and exploitation.

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.

Ad marker example

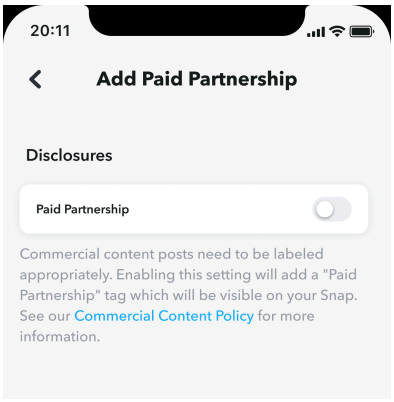
Sponsored Lens example

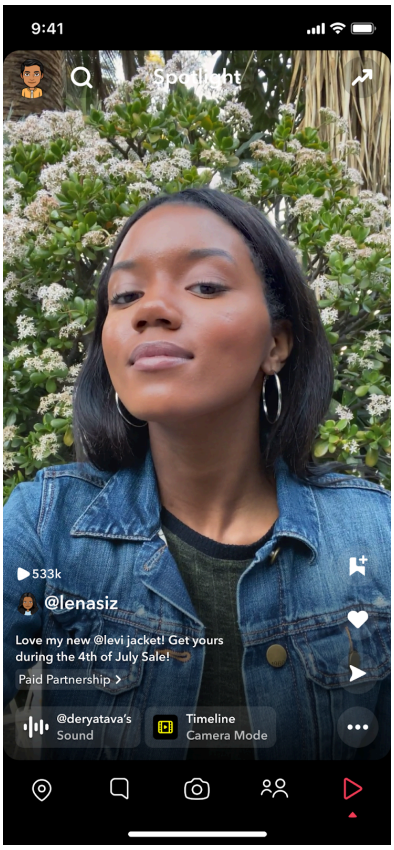


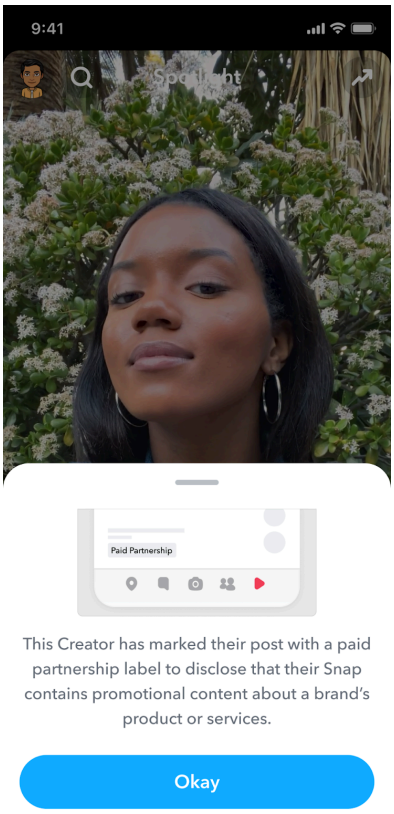


See below for examples of the “Paid Partnership” tag tool that influencers and users may use when they post commercial content to help them comply with this policy and their legal obligations.

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 and UK, Norway and Switzerland. 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.

Ads Gallery

There is 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 an 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) and includes an API interface as well. This allows anyone to check who has paid for an advert and, if different, on whose behalf is the advertisement being published. In the Commercial Content section of the Ads Gallery, we also include links to all live organic content that has been marked with the “Paid Partnership” tag.

Information included for each ad is shown in the screenshots below. When a user clicks on the “See Details” link they are taken to the Ad Details modal on the right. Per DSA guidelines, data includes:

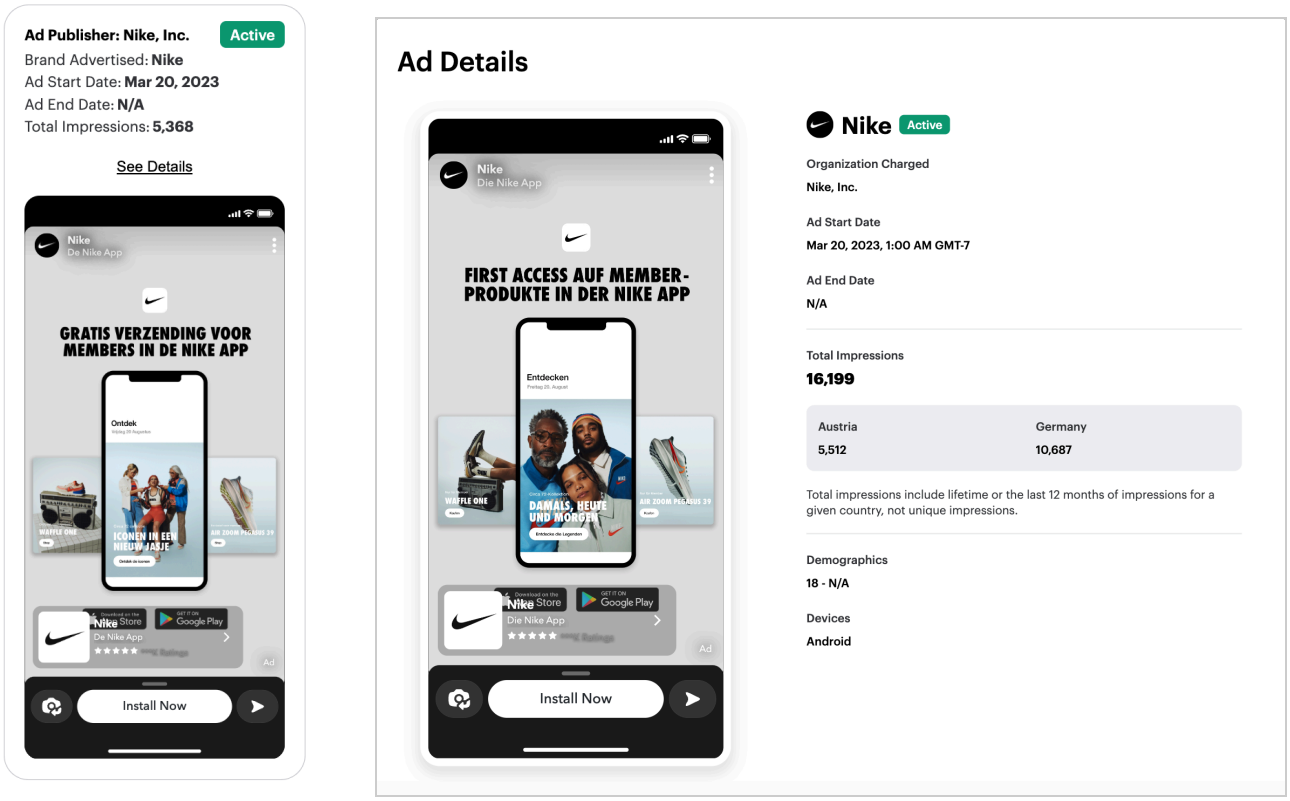
Main Ad Modal

- Ad Publisher - the natural or legal person who paid for the advertisement
- Brand Advertised - the natural or legal person on whose behalf the advertisement is presented
- Ad Start Date and Ad End Date - the period during which the advertisement was presented
- Ad Creative - the content of the advertisement, including the name of the product, service or brand and the subject matter of the advertisement
- Total Impressions - the total number of recipients the service reached

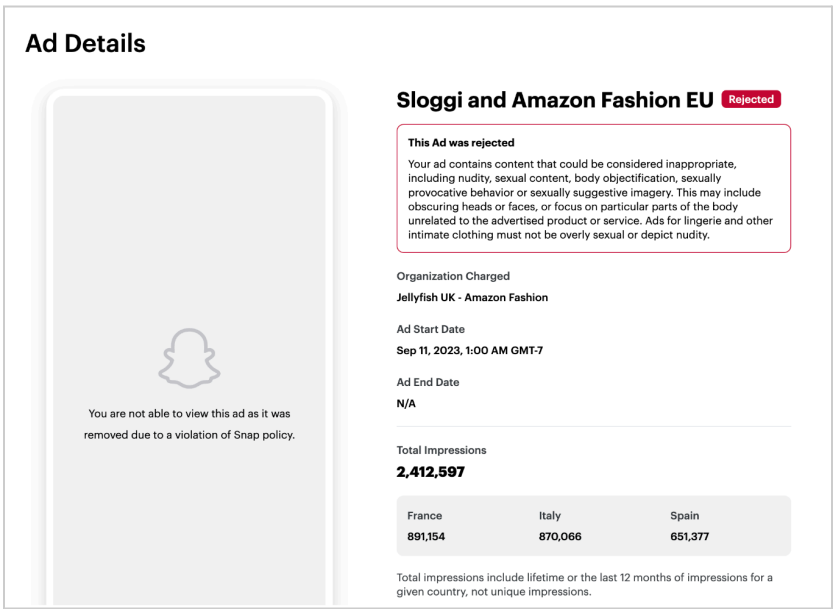
Ad Details Modal

- Impressions by Member State - aggregate numbers for the recipients reached by country (if 0 recipients were reached, the ad will not appear).
- Targeted devices and demographics - whether the advertisement was intended to be presented specifically to one or more particular groups of recipients, specifically, devices and demographics; These options do not support exclusion targeting.

Ads Gallery - Ads Data



Per the Advertising Review section above, we use a combination of automated and human review to prevent ads that violate our policies or the law from appearing on Snapchat. Ads that were delivered and subsequently taken down are marked as Rejected in the Ads Gallery.



The Snap Ads Gallery is maintained by the Ads API and Ads Manager teams. The ads library

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.

5.7.5 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 reject 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

5.8.1 Introduction

Snap's utmost priority is the safety and wellbeing of our users, especially young people, aiming to ensure they have positive experiences on our service. Privacy, safety, and security are key values of the company and sit at the core of our value proposition to our users. Since Snapchat's inception, we have embraced a privacy and safety by design approach and recognise that our platform architecture and product choices play a major role in the protection of our users, including minors. We put significant thought and consideration to ensure our values are reflected in the architecture of our platform, and in the design and implementation of our products, policies, and enforcement actions.

We take the protection of Teens seriously on Snapchat. We've designed Snapchat to protect their privacy, safety, and security. Our key tenets include acting in the best interests of Teens, offering strict default settings for all users, and 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 help guide teens in their responsible use of our platform,

attaching a heightened safety interest to Teens using our products, and establishing processes 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.

5.8.2 Overview and Approach

Age Appropriate Design Code

Snap has adopted the Age Appropriate Design code (AADC) guidance for developers to incorporate appropriate and proportionate measures to safeguard the privacy, safety, and security of users aged 13-17 within platform products and features.

This design code includes 15 core standards:

1. **Best interests of the child**: The best interests of the child should be a primary consideration when we design and develop products.
2. **Data protection impact assessments**: Product must be covered when appropriate by minor data protection impact assessment.
3. **Age appropriate application**: Requires a risk-based approach to recognising the age of individual users to ensure we effectively apply the AADC standards.
4. **Transparency**: The privacy information we provide to Snapchatters, such as our privacy center, support pages and in-app notices, must be concise, prominent and in clear language suited to the age of the child.
5. **Detrimental use of data**: We should not use children's personal data in ways that have been shown to be detrimental to their wellbeing, or that go against industry codes of practice, other regulatory provisions or government advice.
6. **Policies and community standards**: Uphold our own published terms, policies and community standards (including but not limited to privacy policies, age restriction, behavior rules and content policies).
7. **Default settings**: Settings must be 'high privacy' by default (unless wecan demonstrate a compelling reason for a different default setting, taking account of the best interests of the child).
8. **Data minimisation**: Collect and retain only the minimum amount of personal data we need to provide the elements of our service in which a child is actively and knowingly engaged. Give children separate choices over which elements they wish to activate.
9. **Data sharing**: Do not disclose children's data unless we can demonstrate a compelling reason to do so, taking account of the best interests of the child..
10. **Geolocation**: Switch geolocation options off by default.
11. **Parental controls**: If providing parental controls, give the child age appropriate information about this. If our online service allows a parent or carer to monitor their child's

online activity or track their location, provide an obvious sign to the child when they are being monitored.

12. **Profiling**: Switch options which use profiling 'off' by default unless there are appropriate measures in place to protect the child from any harmful effects.
13. **Nudge techniques**: Be mindful of and avoid using nudge techniques to lead or encourage children to provide unnecessary personal data or weaken or turn off their privacy protections.
14. **Connected toys and devices**: If providing a connected toy or device ensure we include effective tools to enable conformance to the Code.
15. **Online tools**: Provide prominent and accessible tools to help children exercise their data protection rights and report concerns.

We continue to actively support the efforts of the Commission and others to establish an EU-wide Age Appropriate Design Code and/or guidance on the application of Article 28 by online platforms, and consider whether further mitigation measures may be reasonable, proportionate and effective for online platforms, 'gateways' and other online services. We have recently submitted our feedback to the Commission on its initial proposals for Article 28 guidance and hope to continue to work closely with the Commission and other stakeholders on its adoption. This is also discussed further in Section 5.12 (Codes).

Privacy, Safety, and Security of Minors on Snapchat

Privacy, safety and security are key values of the company and at the core of our value proposition to our users. Snap has dedicated extensive resources to ensuring protections to safeguard the rights of Teens on the platform, greatly reducing the likelihood of rights infringement. At the highest level Snap follows Age Appropriate Design Codes (or similar) established by the United Kingdom, France, California, etc, as well as our own key tenets described earlier in this playbook. Snap's approach to privacy and safety by design means that we generally design for our youngest users first and work upwards from there. This means our first layer of protection for minors are the mitigations that are designed for Teens but apply to all to advance safety across our community.

Snap has put in place a range of mitigation measures to protect the privacy, safety, and security of minors on Snapchat. 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.

In addition to these safeguards for Teens, we consistently enforce our policies disallowing users under the age of 13 from creating or maintaining an account. Persons who are younger than 13 will be blocked from creating an account at the time of registration; accounts that are discovered to be operated by persons under the age of 13 are removed from the platform at the time that Snapchat discovers such violations.

Advertisements for Minors

Snapchat automatically disables advertisements based on profiling for users aged 13-17 in the EU. Additionally, 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.

Snap has implemented additional safeguards and protections for minors related to advertisements, as described in details in the product-specific subsections below.

Identifying Minors

As the creators of a central communications tool for young people, we take seriously our responsibility to protect teens on our platform. 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 approaches. We look forward to continuing these productive conversations to achieve methods that work for everyone.

Snap currently takes a risk-based approach to provide an age-appropriate experience across Snapchat, consistent with best practices such as the UK Age Appropriate Design Code (AADC). As explained in our response to the Commission's RFI on minors in December 2023:

Registration and access to Snapchat

In order to download the Snapchat app, users first need to create an account with either Apple or Google to access their app stores (Apple App Store and Google Play Store). Both the Apple App Store and Google Play Store have age restrictions, they require users to create an account before they can access the stores, and the age restriction for those accounts is 13+ and in some cases 14, 15 or even 16+ (see Apple and Google age restriction terms).

Both Apple and Google rely on declared age to determine if a user is 13+. If a user provides an age under 13 account creation is persistently blocked unless parental approval is provided. Both Apple and Google offer state of the art and easy to use parent tools (see Apple and Google family link terms). This means that in order to download an app - for example Snapchat - from the Apple or Google Play Store a user needs to declare to be 13+ or parental approval has been provided.

Although Snap has asked for access to Apple and Google's parent tools and age signal to ensure consistency, increase visibility of our guides and settings, reduce the burden on end users and ensure a level playing field with Apple and Google's own apps, Snap does not currently have such access.

As a result, Snap independently asks the user to confirm their age as an additional age assurance measure, despite age already being provided by the same user as part of the Apple and Google account registration flow, as follows:

> 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. In our Terms of Service, Privacy Policy, and other documentation, we make clear that Snapchat is intended for users 13 years old or older. Users must affirmatively add their birthdate when registering for an account, and we deny users declaring they are under the age of 13 the ability to create accounts.
> c. If we determine, or are otherwise made aware through an in-app report from a user, parent, or law enforcement, that an account belongs to someone younger than 13, we take immediate action to prioritize and respond to the information. Our trained internal team will review and terminate the account, including immediately deleting the data associated with the account.

In respect of a)

- We do not use inferred age techniques to prevent individuals under the age of 13 from registering or accessing the app. Reliable age inference is not feasible without data based on user activity once registered and engaging in the app. We do not have such activity level data for any new user at registration, nor do we have this data at any time for users under the age of 13 (since all Snapchatters are declared to be 13 or older).

- As explained below, we have stronger age assurance in place to protect minors from certain content and features of Snapchat targeted at more mature audiences which allow us to protect against potential U18 use of adult accounts despite not having absolute knowledge of U13 use. Our approach is stronger as it allows U18 protections to be applied if there are changes after registration, for example, in potential situations where a parent or other adult user may register an 18+ account using their own information but then provide account access and usage to an U18 bypassing any static age assurance applied on registration.

- Further, as shown in Section 1 (Snapchat Community) the vast majority of Snapchat users access the app to use our messaging services to communicate with friends - not too dissimilar from traditional SMS or other messaging services, which typically do not have any age gates at all. Such interpersonal communication services fall outside the scope of the DSA.

In respect of c) and our trained internal team:

- The team is trained to prioritize these tickets. When our privacy operations team receives a ticket they act upon this promptly, often within a couple of hours.

- If our team is made aware of an account belonging to someone potentially under 13 through external sources (eg. through in-app reporting, Law Enforcement requests), a ticket is created and routed to our human Trust & Safety team who is also trained to prioritize these types of tickets. The team of moderators responsible for reviewing these reports consists of several dozen FTEs.[76] We do not track the specific response time for this type of ticket, these metrics are tracked across all reporting types.

- We provide moderators with training sessions on policies, processes, tooling, current events and cultural norms to be effective at their work. Our moderators are trained through small group training classes and also review a multitude of scenarios while shadowing high-performing peer moderators. Through practice and instruction, they apply our policies and enforcement measures in a manner that protects our Snapchat community. This training is conducted over a multi-week period, in which the moderator is educated on Snap's policies, tools, and escalations procedures. After the training, moderators must pass a certification exam before being permitted to process content.

- In addition to the general moderation training (see below), these team members receive specific training and guidance from our (privacy) legal teams.

In respect of c) and the use of keywords to detect under age users:

- In general, Snap does not scan keywords during account creation or from account information to detect whether a user might be underage. As noted in Section 1 (Snapchat Community), more than 80% of time spent is on private surfaces. Snapchat is primarily intended and used as a communications tool by our users with their close friends and family, and fundamentally different from traditional social media where the majority of the content is public. We apply a privacy-first approach to user communications, such as Chats, and those are not subject to scanning for purposes of learning a user's age or profile. We assessed that doing so would be contrary to the fundamental privacy rights of individuals, as well as privacy laws, including the GDPR and ePrivacy Directive.

- However, since our 2023 Report and the RFI, we have begun testing an additional mechanism to detect under age registrations and access to Snapchat. This involves scanning text fields within user public profiles for certain key phrases.
  We assessed the scanning of public profiles to be less intrusive as this content is public, and on balance would not infringe users' fundamental privacy rights. Where clear statements are made that the user is under age, the corresponding account is further

[76] For a more detailed breakdown of our human moderators please see Section 5.4 (Content Moderation).

investigated and appropriate enforcement action taken (for example, to remove the account). We are continuing to test this new mechanism.

Access to certain content / features

We rely on a combination of declared age and inferred age techniques for stronger age assurance to limit under 18 access to certain content and features targeted at more mature audiences.

2. A combination of declared- and inferred-age techniques for stronger age assurance to protect minors. We not only rely on a users' self-reported age, but have techniques to infer a users' actual age, which considers a combination of various influential signals to limit under-18 access to certain content and features targeted at more mature audiences, including:
    a. Discover (Publisher Partners): At the time of onboarding, Snap provides its Commercial Content Policy to Publisher partners. Publishers must adhere to Snap's Commercial Content Policy. This Policy requires Publishers to ensure their content is appropriate for a 13+ audience. If it's not, our policy requires publishers to age-gate their content. Publisher partners are also given more detailed guidelines on how to comply with the policy (including with respect to age-gating). In addition to this policy, as explained in Section 5.4 (Content Moderation) our global moderation team reviews Publisher Tiles for compliance to ensure content shown in "Stories/Discover" meets our Content Guidelines for Recommendation Eligibility (our standard for "appropriate content"). Our Partnerships team also performs periodic checks of Publishers and Content Creators to ensure that the content they are posting is compliant with these guidelines. Publisher partners receive notification (either emailed or in-app, depending on the creator type) if their content is found in violation of our guidelines. Publisher partners who violate frequently and/or severely are further penalized, after editorial review. These penalties can involve a period of restricted visibility, a suspension during which all publishing is banned, or a permanent channel ban. In rare instances, we've ended relationships with entire organizations. In addition to penalties, Snap regularly communicates updates and clarifications to our guidelines in emailed newsletters to partners, and holds periodic training seminars for publishing partners.
    b. Spotlight and Discover (UGC):
        i. Content that is prohibited by our Community Guidelines is prohibited everywhere on Snapchat and our Content Guidelines for Recommendation Eligibility specifies additional categories of content that, while permitted on Snapchat, will not be eligible for recommendation to a wider audience on Spotlight and Discover. These mitigation measures apply to all Snapchatters.

ii. In addition, Spotlight and Discover have been designed not to distribute sexually suggestive content to 13 - 17 Snapchat accounts and only recommends sexually suggestive content to a 18+ Snapchat account if that content has been created by a creator that the account has subscribed to or favorites suggestive content. This uses our combined declared and inferred age techniques.

iii. Regarding the machine learning classifiers, we use in-house classifiers that were trained to scan and identify sexually suggestive content using state-of-the-art computer vision models. When user-generated content on Discover and Spotlight is scanned by our machine learning classifiers, content that is scored above our threshold and considered "suggestive" is then removed from the content recommended to teens. Snap assessed (and continues to assess) the effectiveness of our machine learning classifiers via (i) quality testing and product/engineering review before deployment and (ii) ongoing review against in-house human labelling of publicly available content. We do not deploy new machine learning classifiers until they achieve at least 80% precision. Note that we also conduct routine quality checks of our human review where our precision is 95%+.

iv. Our systems are designed in a way that a Snapchatter (who is over 18 years old) who has subscribed to a content creator that has posted suggestive content will not see more than one sexually suggestive content video out of seven in their Spotlight feed. If a user hides a video labelled as sexually suggestive or a sexually suggestive creator (here is an example of a piece of Spotlight content that was marked as suggestive), we stop showing that type of content to that user. If a user hides a creator, we stop showing them that creator.

c. Lenses: We age-gate certain Lenses (e.g., related to alcohol, gambling, NFTs, etc.). For example, a Vivino Lens will only be shown to 18+ users in the EU, mitigating the risk of such content being shown to users who are under the legal drinking age.

d. 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.

Our age inference model is used as an integral part of our age assurance method to limit under 18 access to content and features targeted and suited to more mature audiences. The inferred age model on Snap uses a variety of influential signals, rather than only relying on the age that users provide when signing up to the platform. This helps, for example, prevent regulated ads

from being served to those users who have declared themselves to be over the age restriction but we have modelled as likely to be under the appropriate age for such regulated content.

For example, if a user's self-declared age is 20 years old, yet the signals derived from the user's activity within the app and the ages of their friends strongly indicate that they are likely under 18 years old, we may internally "override" their supplied age and flag the user as a minor, and therefore filter regulated ad content (e.g. alcohol) from being displayed to them.

The inferred age model does not utilise any third party age assurance providers. To minimise disclosure of personal data, the model is processed within Snap.
As flagged above, we do not use inferred-age techniques to prevent individuals under the age of 13 from registering or accessing the app. Reliable age inference is not feasible without data based on user activity once registered and engaging in the app. We do not have such activity-level data for any new user at registration, nor do we have this data at any time for users under the age of 13 (since all Snapchatters are declared to be 13 or older). The use case for our age inference model has always been geared toward mitigating the risks for users under 18, for example to ensure users under 18 do not see regulated ads, such as ads for alcohol. As Snap isn't capable of determining that a user is under 13 from our inferred age model, we have never applied it for the purpose of preventing users from registering an account. In line with our COPPA obligations, if Snap receives actual knowledge that a user is under 13 (for example, from a verified parent's request), Snap promptly deletes the account and associated data.

Oversight

We have a dedicated working group overseeing our age-assurance efforts. This cross-functional group consists of 10+ FTEs with representatives from the product, privacy legal, product legal, policy, and trust and safety teams. These team members bring to the table extensive experience and knowledge in the areas of operations, policy, global privacy laws and regulations, privacy-focused product decision making and online safety.

Different iterations of this group meet in working groups that are focused on exploring, discussing and assessing possible ideas and concepts and solutions related to age assurance through risk assessments, discussions with our product team, and collaboration with external legislators, regulators, vendors, experts, NGOs and other stakeholders.

For example:

- This group regularly meets to discuss Snap's age and parental assurance methods, and the legal and regulatory requirements in this space.
- This team has assessed current industry practice with regards to age and parental assurance, including in particular the mechanisms of: (i) device operating systems Google Family Link, Apple device parental controls and Family Sharing controls and Microsoft Family Safety and (ii) other online services such as Whatsapp, Tiktok, Instagram and Youtube.

- The team regularly engages in stakeholder meetings, such as those organised by the Centre for Information Policy Leadership (CIPL) Privacy in Europe.
- This group also advocates for a holistic approach to age and parental assurance. An example of this would be our work with the UK Government and House of Lords during the passage of the Online Safety legislation in Parliament to successfully achieve a requirement for Ofcom (the UK's communication regulator responsible for regulating the UK's Online Safety Act) to consult holistically on child safety and age assurance, including considering the role that infrastructure such as app stores / device operating system accounts have in providing privacy friendly, secure, effective and efficient solutions.
- This team has also met with a number of industry leading third party age assurance providers to assess the technical, legal, financial and user impacts of those services being integrated into Snap.
- This team has reviewed research from external stakeholders, on the positive and negative impacts of age and parental assurance, as well as consulting with users of Snapchat (including our own families and friends).
- Representatives from this group presented to Snap's Safety Advisory Board[77] on Snap's approach to age assurance and Family Center, including potential options and challenges with age assurances. On age assurance in particular, this group of experts advised us that:
    - This was an industry wide issue that needed broad stakeholder discussion and need for service independent solutions (i.e. considering devices, app stores etc) that best met the needs of children and parents/responsible adults.
    - They also felt our 13+ age limit was inhibiting our ability to recognise and keep potential younger users safe and advised us to consider lowering the age so Snapchat's U18 experience was available to those that wish to use and less incentive for children and parents/responsible adults to try to access via 18+ accounts / bypass age assurance methods. They recognised the challenges that COPPA presents in this regard.
    - They felt it was better to have a safe platform for all than to rely on excessive age / parental assurance. This would limit the need for age / parental assurance to the smaller number of mature areas.

Ongoing evaluation

Snap continues to evaluate its approach, and consider possible concepts and approaches with industry peers and third-party age-assurance vendors, to ensure we keep pace with industry practice. We are also supporting legislators and NGOs in the UK, France, and elsewhere in the EU 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 approaches for the online ecosystem.

[77] The Safety Advisory Board is explained in Section 6.7 (Snap Advisory Groups).

Snap earned praise in recent years by the Children's Advertising Review Unit (CARU) for exceeding minimal protections to keep underage users off Snapchat, and for providing numerous safeguards for our users once they are on our platform. Despite this, we continue to reassess our age assurance efforts, including engaging with industry partners, regulators, and third party age assurance vendors, to ensure we keep pace with developments in the space. We actively participate as thought leaders in industry roundtable forums, as well as with policy makers in the US and Europe to understand the evolving landscape of age assurance, and its critical importance coupled with its inherent challenges. We've also held multiple exploratory and deeper dive sessions with leading vendors across the age estimation and identity verification marketplace in recent months, as we consider third party technologies including biometrics, ID scan, and financial transaction methods (among others), to enhance our current approach.

In terms of parameters, Snap consistently focuses on the potential impact to key areas when evaluating the effectiveness of age assurance measures. Such factors include the tradeoff between safety of minors and compromising user privacy/data security, the accuracy and reliability of age estimation technology (particularly for younger or ethnic minority populations), the fairness of methods that may disadvantage users without official government IDs or bank accounts, and the harm to industry competitiveness from the exorbitant cost of adopting third party technology at scale.

It is also worth noting that many organizations have concerns about the effects of introducing age assurance and support device OS account and app store based solutions for age assurance to support online platforms' own mitigation measures, for example ICMEC in the US. Prominent European children's NGOs have expressed similar support when meeting with Snapchat.

We remain focused on thoughtful enhancements to our risk-based, age and parental assurance approach that include balancing the need for safety, accuracy, fairness, and user privacy among other important factors and taking a holistic view of the online ecosystem used by children and their parent(s) (or other responsible adult(s)). However, currently, taking account the results of our risks assessment set out in Section 4 and current industry wide practices, we continue to conclude that our approach is proportionate, reasonable and effective.

Commission Submission

We have recently submitted our feedback to the Commission on its proposals for guidance on Article 28 for online platforms. 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 in particular the role of online platforms that are 'gateways' (such as device operating systems, app stores and web browsers) through which

parents and Teens engage with such platforms. For example, it would make logical sense to further enhance age assurance at the App Store and device operating system levels to ensure a robust upstream solution[78] and support the measures taken by third party services (like Snapchat).

Critically, Snap also consistently focuses on the potential impact to key areas when evaluating the effectiveness of age-assurance measures. Such factors include the tradeoff between the safety of minors and compromising user privacy/data security, the accuracy and reliability of age estimation technology (particularly for younger or ethnic minority populations), the fairness of methods that may disadvantage vulnerable segments including users without official government IDs or bank accounts, and the harm to industry competitiveness from the exorbitant cost of adopting third-party technology at scale.

Transparency to Minors

We continue to make efforts to provide users with information regarding our services in a way that is clear and comprehensible across age groups. We verify readability of our key terms and conditions and privacy notices using automated readability tools. The vast majority of these documents are shown to be understandable for our users. Our main terms and conditions is, necessarily, a formal legal document and contains longer provisions and more complex language which our automated readability tools indicate may be more difficult for our younger users to understand. To improve readability in particular for our younger users, we arranged the terms and conditions into sections that provide a sensible flow, including appropriate and succinct section headers, and added short summaries at the bottom of each section. We have tested the readability of these short summaries, and confirmed they are understandable for our users, including Teens. In addition, we have provided short explainers for our Community Guidelines to facilitate user understanding of this important document and know what they should and should not be doing on Snapchat.

Our Privacy Center was designed for our youngest Snapchatters and was intentionally developed to be easy to read and understood by all members of our community. We created our privacy and safety hub, with pages such as our Privacy by Product page, to give Snapchatters a high-level summary of our privacy and safety practices across each of our products and features. We also created a video to visualize our privacy practices, and use icons and other best practices as recommended by privacy and safety experts and the recognised Age Appropriate Design Codes.

As outlined in the Introduction, our European Snapchatter community consists of a diverse range of ages and genders. Snapchat services are not primarily directed at or used by minors. While Snapchat does have a young demographic, only a relatively small percentage of European Union

[78] "Making Smartphones and App Stores Safe for Kids: Federal, State, and Industry Measures," published on Nov. 16, 2023 by the Ethics and Public Policy Center and Institute for Family Studies: https://ifstudies.org/ifs-admin/resources/briefs/ifs-eppc-smartphonesappstoresbrief-nov23.pdf.

users fall within the 13-17 age category. The largest age category of European Union users is 18-24.

5.8.3 App Store Level Safeguards

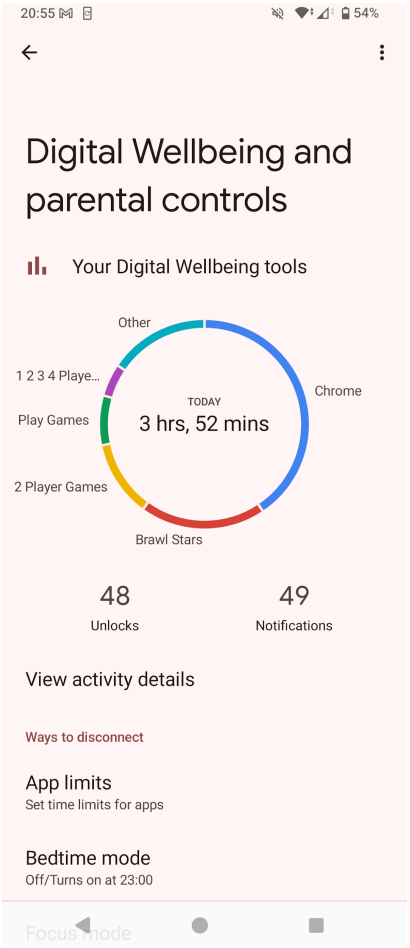
App stores contain and facilitate a vast range of apps presenting a wide array of different risks and representing the entire spectrum of online risks. They have a special place in the ecosystem and therefore a uniquely high risk. They should thus be considered as 'high-risk situations'. In addition to the specific risks presented by the individual apps hosted, app stores can generate higher and exponential risks for a minor than the ones created by each individual service (e.g. a minor accessing harmful information on different services and combining harms).

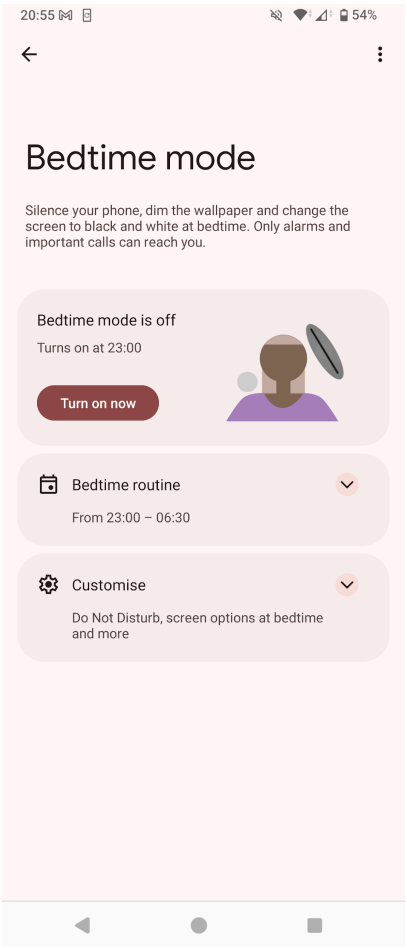
Larger app stores already recognise that they present unique risks and require additional mitigations. In fact, larger app stores already apply age gates to prevent users from downloading an app if their app store account age is below the app's minimum category specified by the app provider (and where applicable parents via the device operating system's account level family controls - see below). Like most online platforms, app stores usually rely on the app store account's declared age (which is often the same account used by the device operating system). We also have noted that app stores rely on predetermined age categories which do not necessarily capture nor fit the age categories defined by Snapchat and other developers at the app level. A classical example is the app store category 12+, which does not align with the age threshold of 13+ that is commonly specified by application services.

There are improvements which could be made by app stores providers for the benefit of the entire ecosystem including: (i) stronger age assurance at the app store / device OS account level and sharing that signal with developers to support their minor protection measures; and (ii) allowing developers to set a more precise minimum age.

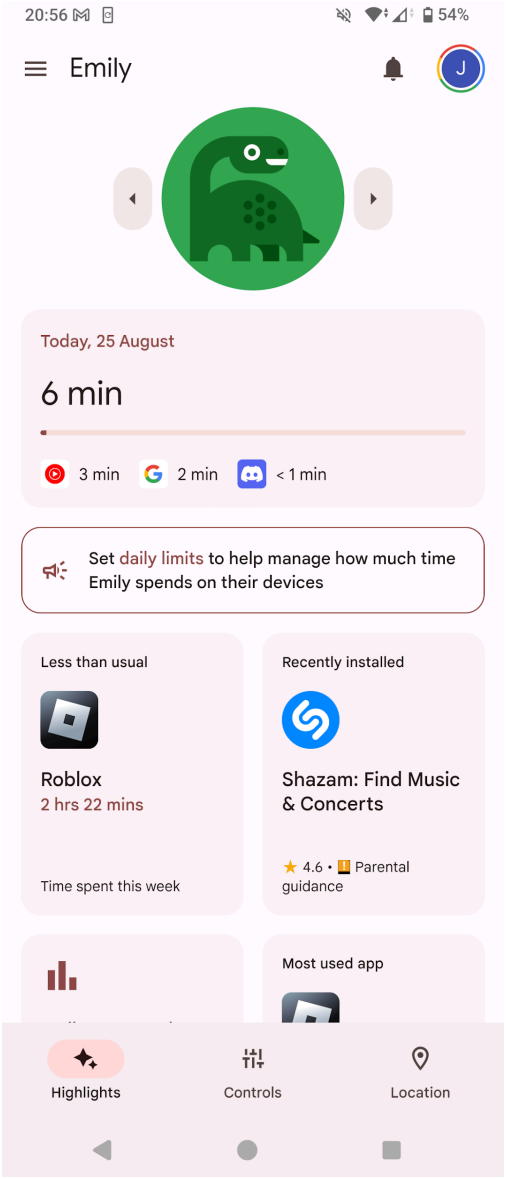
5.8.4 Device-Level Safeguards

Many additional controls are provided for teenagers, parents and other responsible adults. Many devices now come with wellbeing settings, such as bedtime mode that turn off device and app notifications and turn the screen the black and white to encourage sleep.





Additional controls are also provided via the device operating system's account level family controls (e.g. Google Family Link, Apple device parental controls and Family Sharing controls and Microsoft Family Safety). For example, via these controls, parents and other responsible adults are able to view usage, set time limits, and disable access for each app which the teenagers have on their device.



We have also noted that the providers that operate these family controls (who are also gatekeepers pursuant to the EU Digital Markets Act) provide deeper levels of visibility and control for their own first party services. This level of interoperability and access would be very helpful for our own Family Centre (which is explained below) as it would increase the awareness and accessibility for parents and other responsible adults who may not have a Snapchat account. As explained below, we are actively encouraging further multi-stakeholder dialogue to have drive solutions that provide equal access and interoperability across the industry.

5.8.5 Platform-Level Safeguards

There are several protections that we put in place at a platform level to mitigate the risk of malicious users of Snapchat. In particular, we have inference models in place that look at platform

wide meta-data signals to identify suspicious accounts. We use this information at a product level to implement additional safeguards for Teen and adult users.

5.8.6 Product-Level Safeguards

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.

Public Content

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. Teen stories are deleted by default. Their My Story view setting is defaulted to friends only. Friends lists are private.

Viewing Public Content

Teen accounts are restricted from access to certain content that is generally considered suitable for 13+ but may contain certain shocking or sensitive content some may not find appropriate.

Spotlight

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 upon reaching a threshold number of views, and before being even more widely distributed. These steps reduce the likelihood of Teens accessing illegal or violating content, or content that may negatively affect their rights, security and health.
We also aim to prevent older users from seeing content from younger users and to protect Teens from being contacted by older users. We seek to achieve this by, for example, implementing the following measures:

- We limit the recommendation of content created by Teens to older users
- Adults cannot comment on Teen's Spotlight content on Snapchat.
- Users can also choose to disable comments on any post.
- Teens are protected on Spotlight by not having their usernames displayed.

Map

As U18 users cannot have Public Profiles, they will not have their Public Stories featured on Snap Map when tagging a place or venue to a Public Story (which would occur for 18+ accounts).

Additionally, Map filters out suggestive content from being recommended to users ages 13-17 and age-gates certain types of locations to prevent them from showing on maps for minors, including bars and tattoo parlors.

Advertisements

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.

Advertisers must comply with our Ad Terms, Advertising Policies, and applicable national advertising codes. We prohibit ads that address or intend to appeal specifically to children under the age of 13.

Reporting and Blocking

- In-App Reporting: Teens have the ability to report abuse they may observe or experience within Snapchat. They can easily report Snaps, Chats, Stories, and Accounts by navigating to the clearly marked "Report" option in the menu on each of these feature screens or by pressing and holding on the content itself. Users follow our simple reporting flow and provide their reason for reporting and any additional comments that might be relevant. Reports are reviewed by our Trust & Safety teams that operate 24 hours a day, 7 days a week, and violating content and accounts are subject to enforcement. See the Enforcement section of this Report for more information.
- Blocking: All users have the option to Block another user. This prevents the friend from viewing friend content posted by, or sending Snaps and Chats to, the blocking Snapchatter. Since our 2023 Report, in an effort to prevent bullying and potential repeat harassment, we have introduced improvements to our blocking tools: Blocking a user will also now block new friend requests sent from other accounts created on the same device.[79]
- Removing Friends: All users also have the option to remove a friend from their friends list. Once removed, the Snapchatter will no longer be able to view content accessible only by friends and, by default, should not be able to Chat or Snap.

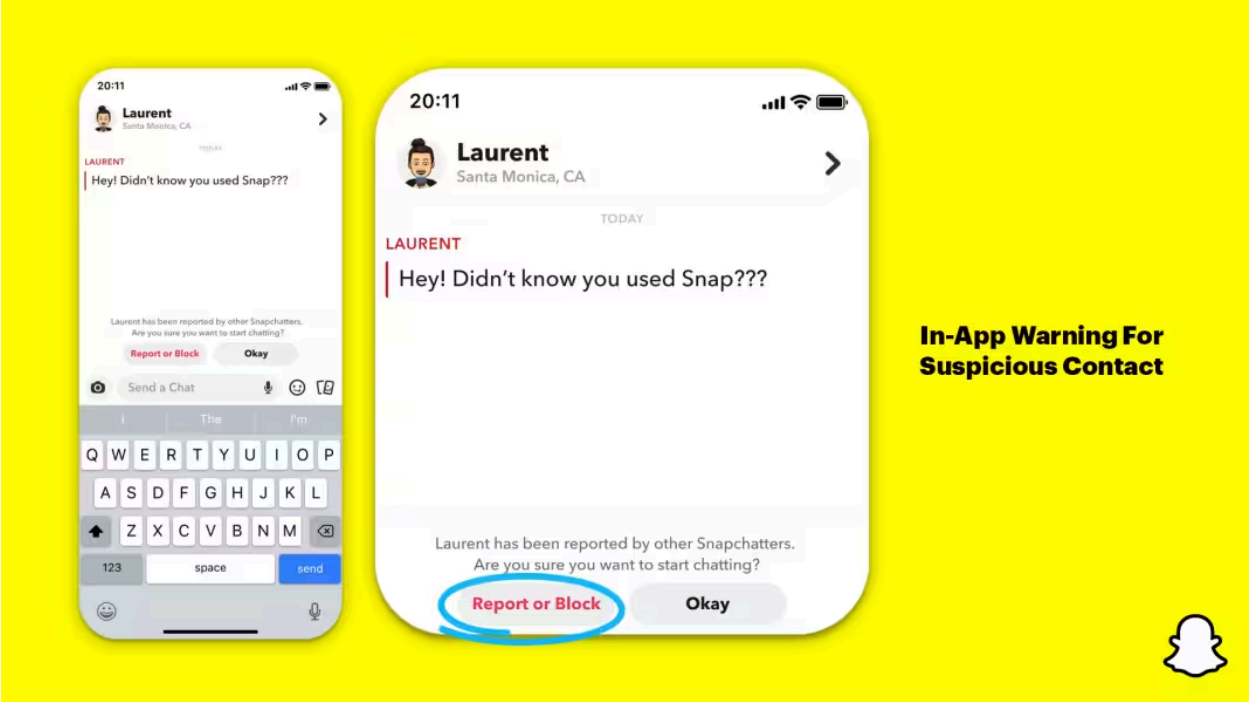
Friending

- Similar protections apply to prevent Teens from searching for unknown adults. We prevent delivery of a friend request altogether when Teens send or receive a friend request from

[79] https://values.snap.com/news/new-features-to-help-protect-our-community

someone they don't have mutual friends with, and that person also has a history of accessing Snapchat in locations often associated with scamming activity.[80]

- 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.
- No Public Friends Lists: 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).
- Friend Check-Up: Prompts Snapchatters to review their friend lists and remove those they are no longer in contact with, keeping their network up-to-date and focused on close friends.
- In-App Warning: We provide pop-up warnings: (1) when a teen receives a message from someone they don't already share mutual friends with or have in their contacts[81] and (2) if they receive a chat from someone who has been blocked or reported by others, or is from a region where the teen's network isn't typically located.[82]



[80] https://values.snap.com/news/new-features-to-help-protect-our-community.
[81] https://values.snap.com/news/new-safeguards-for-snapchatters-2023.
[82] https://values.snap.com/news/new-features-to-help-protect-our-community.

Family Center / Parent Tools

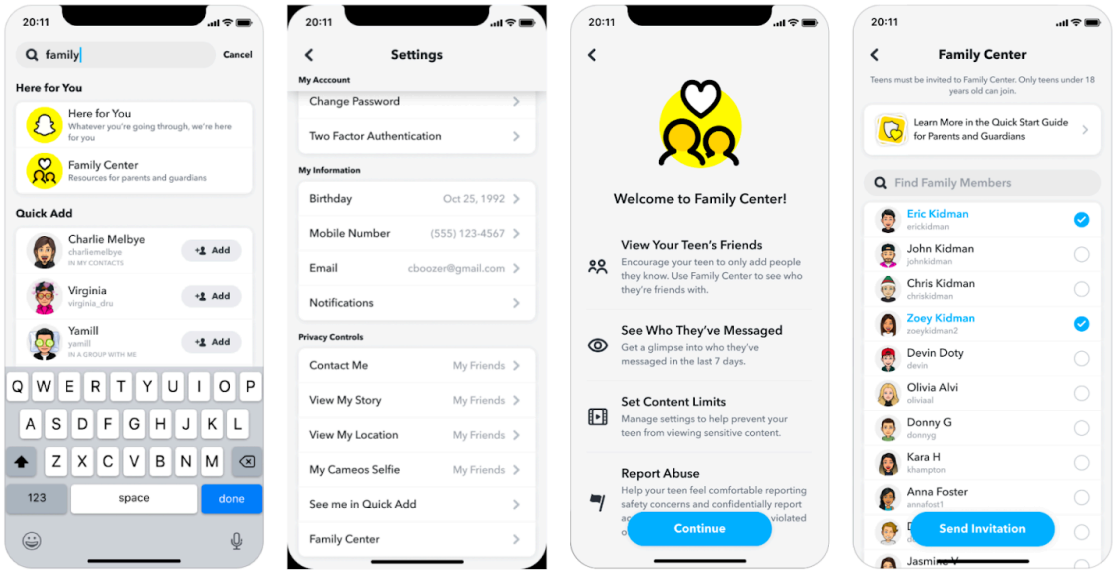
Our in-app parental supervision tool, Family Center, gives parents, caregivers, and other trusted adults visibility into their teens' friends list and who they have messaged with in the last seven days, as well as the the ability to: (i) restrict their teen's access to Spotlight and Discover content tagged as 'sensitive' by our moderation team, (ii) limit their teen's ability to engage with the My AI chatbot; and (iii) quickly request their teen's location (which the teen must approve before location is shared). Parents are also able to easily report accounts that may be in violation of our Community Guidelines and have access to helpful resources directly in the app.

Our goal in designing Family Center 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 continue to put care and time into establishing this balance 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, including members of our current Safety Advisory Board, and extensive cross-functional internal reviews, including by our Product Legal and Privacy Engineering teams.

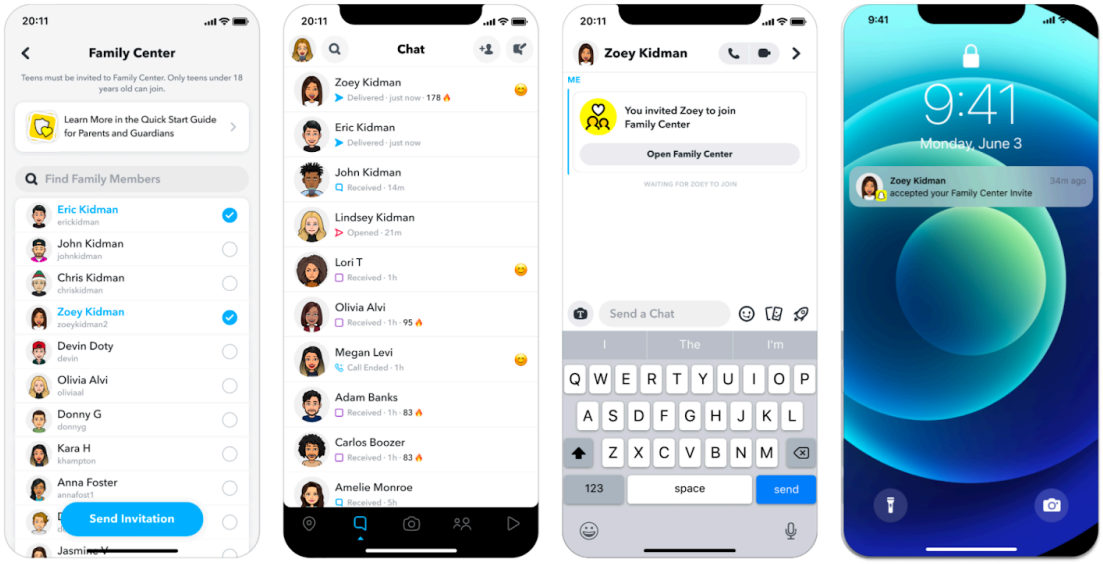
In their annual report,[83] Jugendschutz.net, the joint competence center of the German Federal and State governments for the protection of children and young people on the Internet, highlighted Family Center as a positive example in the area of parental tools and support on social media platforms. The 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.

[83] Jugendschutz, '*Jugendschutz im Internet - 2022 Bericht',* April 2023, url.

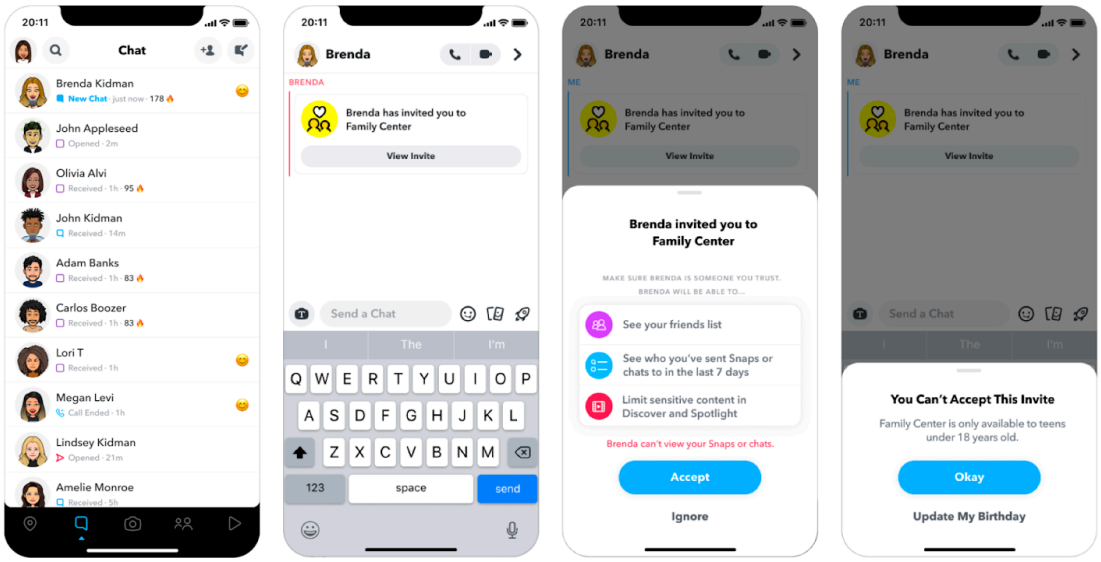
Finding Family Center



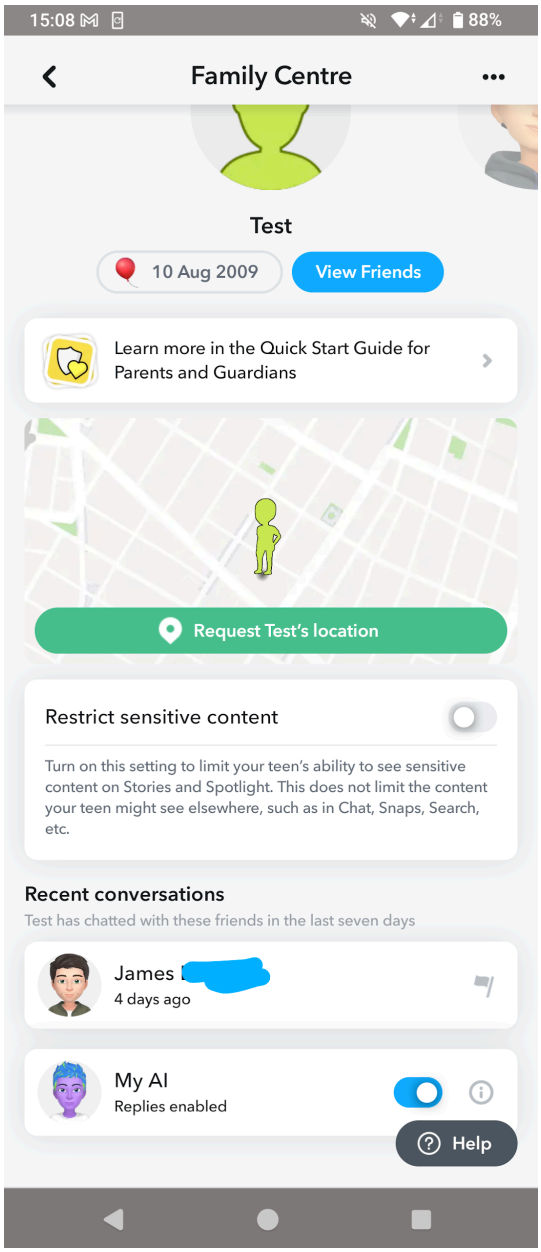
Inviting Teens to Family Center

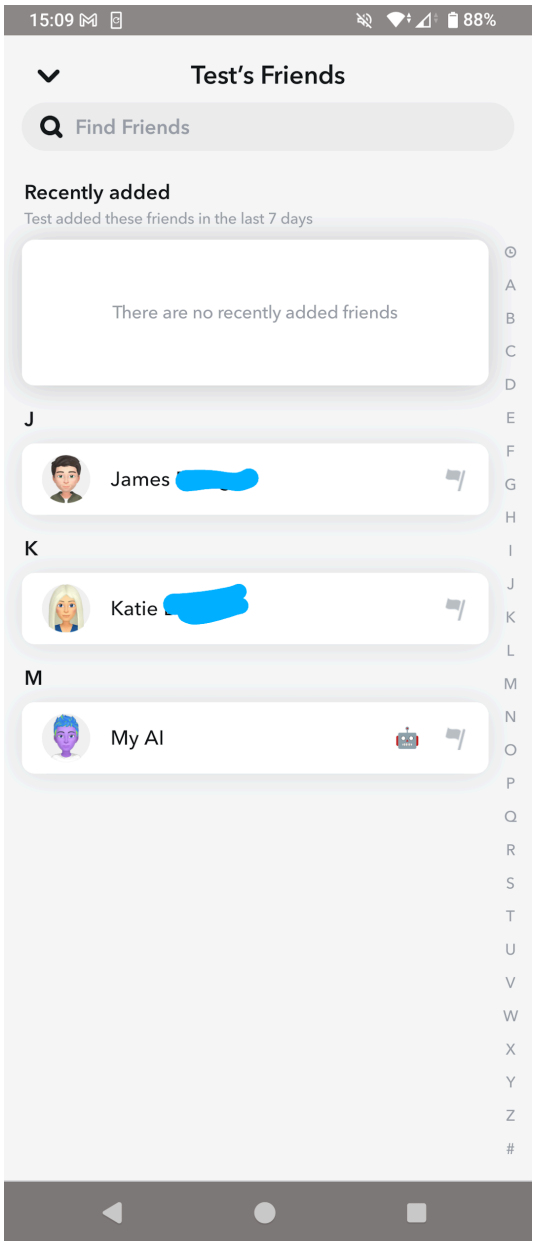


Joining Family Center



Using Family Center





5.8.7 Conclusion

We take the protection of Teens seriously on Snapchat. We've designed Snapchat to protect their privacy, safety, and security. Our key tenets include acting in the best interests of Teens, offering strict default settings for all users, and 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 help guide teens in their responsible use of our platform, attaching a heightened safety interest to Teens using our products, and establishing processes 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 continue to actively work with the Commission and others to establish an EU Age Appropriate Design Code and/or guidance on high level of privacy, safety and security under Art 28 of the DSA, and consider whether further mitigation measures may be reasonable, proportionate, and effective for online platforms, 'gateways,' and other online services.

5.9 Content Authenticity

5.9.1 Introduction

Snap is aware that there is intense interest and concern surrounding the ways in which advancements in generative AI technologies are impacting online platforms. Snap recognises the potential for AI generated or transformed content to be distributed through the inscope services of Snapchat, such as Spotlight and Discover, as explicitly called out in our Community Guidelines. Just like any other content distributed through these channels, this content may constitute illegal content or information that otherwise violates Snap terms and could contribute to the systematic risks outlined in Section 34 of the DSA.

5.9.2 Risk Assessment Results

Snap gave due consideration to the risks and harms that could arise from dissemination of user content in its risk assessment results section of this Report. In particular:

- Section 4.1.10 (Dissemination of harmful false information) - In this section, Snap recognised that "fake news," (online) "disinformation" and "deep fakes" had gained a lot of attention in the media and academic and political debate over the last years. We recognised that such content presented a risk of significant harm. This applied to all content formats, whether or not generated using AI tools. However, when considering evidence relating to Snapchat specifically, we found very low prevalence rates of this type of harmful content. We concluded that 'harmful false information' fell within the lowest likelihood and risk prioritisation category relative to other harms being monitored on Snapchat.
- Section 4.3.1 (Negative Effect on Democratic and Electoral Processes) - In this section, Snap recognised 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. Again, this applied whether or not generated using AI tools. However, when considering evidence relating to Snapchat specifically, we found only limited occurrence of content harmful to democracy. 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),[84] 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. We concluded that 'negative effect on democratic and electoral processes' fell within our lowest likelihood and risk prioritisation category relative to other harms being monitored on Snapchat.

- Section 4.3.2 (Negative effect on civil discourse) - In this section, Snap recognised that digital content platforms could contribute to negative effects on civil discourse. For example, we noted:
    - 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.

  However, when considering the evidence relating to Snap specifically, we again found a very low prevalence of content related to harming civic discourse relative to other categories being monitored. We concluded that 'negative effect on civil discourse' fell within the lowest likelihood and risk prioritisation category.

5.9.3 Mitigations

Although there is not a high prevalence of harmful false information, fraud and spam or impersonation, we take harmful information of this nature on Snapchat very seriously and Snap has implemented a number of mitigation measures.

Guidelines, policies, and practices

Snap maintains robust policies––applicable to both the dissemination and the creation of generative AI content––that function to mitigate risk and advance safety.

[84] Election Integrity Partnership (2020), url.

Creation

The features for creating generative AI content that Snap offers are not part of Snapchat's inscope services and are out of scope of this Report (save for certain commonplace ad creation tools). Nevertheless, outside of its DSA obligations, we note that Snap has developed several internal policies relating to generative AI. In particular,

(1) Content and Product policies: We have developed a suite of policies that disallow the generation of harmful content (including deceptive political content). Our policy and moderation teams work in partnership with engineering and data science colleagues to ensure that our AI products are responsibly trained on these policy parameters.
(2) Acceptable Use: We have similarly developed Acceptable Use Policies that prohibit the use of our AI tools to attempt to generate violative content at the prompt-level.

With regards to the Content and Product policies, we have drafted and implemented internal Generative AI Policies to govern the internal development of generative AI features, such as MyAI. Our Product and engineering teams refer to this policy as they train models or adapt models from third parties. Our Safety team applies this policy when testing new features.
These aligned very closely with the rules for content dissemination, which are explained below.

In addition, our generative AI tools feature a broad range of mitigation measures, depending on the tool, and include for example: specific transparency statements, abusive language detection and query related measures, age appropriate and/or canned responses, reporting mechanisms, off-by default settings, data minimisation, data sharing, testing and parental controls (see Section 5.8 on the Protection of Minors). Although out of scope of the DSA and this Report, risks and mitigations relating to our generative AI tools are assessed via our privacy and safety by design product reviews (see Section 6.3).

Dissemination

In the context of dissemination of content on Snapchat's online platform, in scope of the DSA, 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.

Our Community Guidelines and Terms of Service set out the rules on what content is allowed on Snapchat. They are focused on preventing harm to Snapchatters and the broader community from content and behaviour, whether or not caused by generative AI or any other form of IT tools (such as Photoshop). These rules apply to all content formats across our platform, including content that is AI-generated. While the rules are agnostic to content format or creative tools, the Community Guidelines specifically note: "We implement safeguards designed to help keep generative AI content in line with our Community Guidelines, and we expect Snapchatters to use AI responsibly. We reserve the right to take appropriate enforcement action against accounts that

use AI to violate our Community Guidelines, up to and including the possible termination of an account."

Our rules and internal enforcement guidance include clear provisions related to content risks, for example for civic discourse and electoral processes. In particular, our Community Guidelines 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).

Our Community Guidelines rules on false information refer to a more detailed Explainer that prohibits content that undermines the integrity of civic processes, or deep fake content or other media that is manipulated for false or misleading purposes. The Community Guidelines further explain that these prohibitions extend to the following types of harmful content:

- *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 rumours 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.

Sharing such content will violate Snap's Community Guidelines irrespective of whether it is AI-generated or user-generated, or whether it is generated on Snapchat or on another platform.

Snap has a suite of internal policies and guidelines to help our content review and trust and safety teams apply the Community Guidelines to user generated content disseminated via our online platforms (such as Spotlight and Discover). They provide more granular information for our content review teams. For example, we explain that obvious jokes, memes, satire and non-libelous comments about prominent social figures are OK; whereas false political narratives meant to undermine elections, or harmful / defamatory deepfakes, are NOT OK.

In addition, our platform does not widely distribute an unvetted feed of algorithmically curated political information; we disallow all political content[85] from Spotlight (our broadcast platform for User Generated Content) unless it's from trusted news partners and creators, and pre-moderate

[85] 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.

that surface to ensure that other such political content is not distributed. 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).

With regards to advertising, we do not require ads to label when advertisement includes generative AI content nor do require advertisers to disclose to us the tools they used to edit or create their ad creative. Instead, our approach is to subject all of our ads to a review process, and political ads are also subject to fact checking. Deceptive ads are rejected, irrespective of whether they use AI, photoshop, or other digital editing tools. Ads that are not deceptive, and otherwise comply with our Ad Policies, are approved to run (and if they are a political ad, they must include a "paid for by" disclaimer and are catalogued in Snap's political ads library).

User Guidance

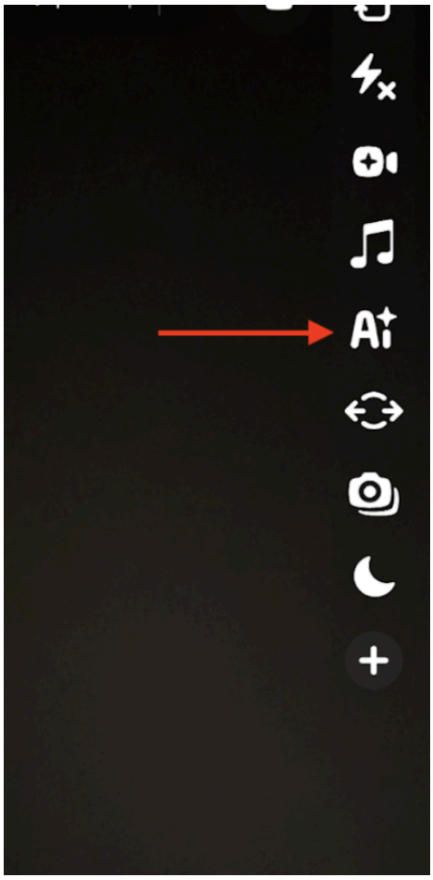
On-platform features for creating generative AI content are not part of Snap's inscope services and are out of scope of this Report (save for certain commonplace ad creation tools). Nevertheless, outside of its DSA obligations, Snap has released a generative AI support site that explains what generative AI is and provides additional transparency around Snap's practices with regard to generated images which are detailed below.

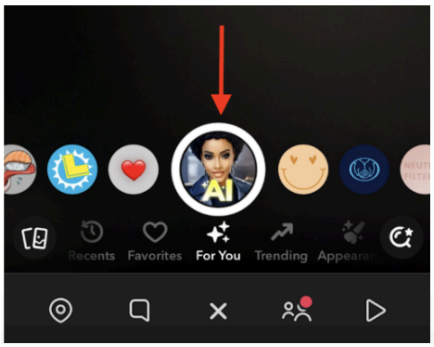
Snap considers that there limitations and factors that need to be considered when adding watermarks, labels and other marks to identify generative AI content:

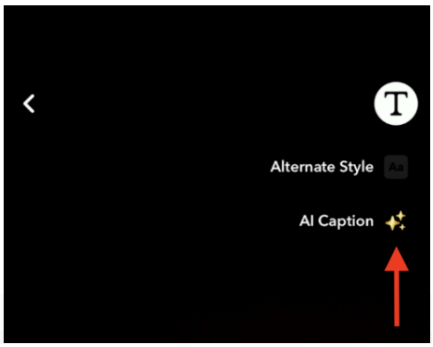
- **Noise:** On Snapchat, AI is often on a spectrum of assistance rather than being the primary feature. If an identifying mark must be added any time any feature that uses any tool that might be classified as generative AI, even if in a minor way, that would be overwhelming and users would likely start ignoring the watermark. Snap has adopted a focused approach.
- **Fakery:** whether visual or embedded in metadata, watermarks, labels and other marks can be counterfeited, altered or removed with a little technical skill. Snap therefore considers the potential for activity to create undue credulity for fake content, or to erode trust, and therefore avoids placing too great a reliance on watermarks, labels and other marks.
- **Privacy:** C2PA-style metadata watermarking could risk compromising a user's identity, location or other sensitive information. Snap takes this into account when considering whether to apply watermarks.

Given these limitations and factors, Snap often relies on other more proportionate, reasonable and effective measures to mitigate the risks presented by generative AI (such as content moderation). Nevertheless, Snap has chosen in some cases to indicate that a feature in Snapchat is powered by generative AI in a number of ways, including using the sparkle icon ✨, specific disclaimers, Context Cards, or tool tips. When users see these contextual symbols or other indicators in Snapchat, they should know that they are interacting with AI and not a human, or viewing content that has been produced using AI and does not depict real world scenarios.

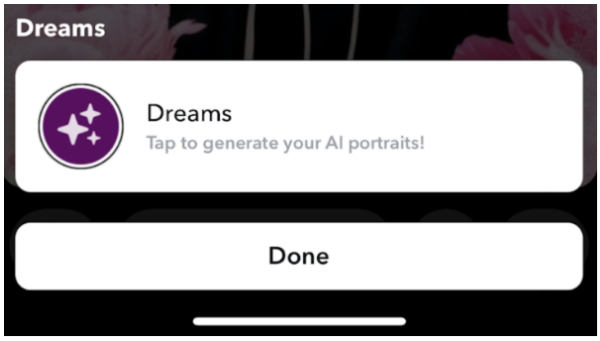
Here are examples of what these AI indicators look like in Snapchat:





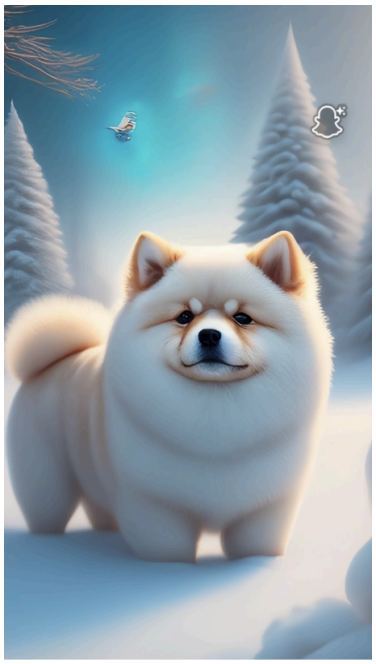


- When some AI generated images are shared with others on Snapchat, we may include a Context Card to let you know that the image was created with a generative AI-powered feature.



- Some generative AI-powered features, like Dreams and AI Snaps, allow you to create or edit images. When you export or save a generated image to Camera Roll, a watermark of a Snap Ghost with sparkles may be added to those images. The purpose of these watermarks is to provide transparency that the image was created with generative AI and is not real or based on real events, even if it is a realistic style. Not all AI generated images will include a Context Card or watermark. Images created with non-Snap products may not be labeled as AI generated.

Here's an example of what the watermark looks like:



Enforcement

Snap enforces these Community Guidelines fairly and consistently, using internal policies and guidelines, and applies outcomes that are commensurate with the severity of risk.

Accounts that we determine are used to perpetrate serious, high-severity harms will immediately be disabled. For other violations of our Community Guidelines, Snap generally applies a three-part enforcement process:

- Step one: the violating content is removed.
- Step two: the Snapchatter receives a notification, indicating that they have violated our Community Guidelines, that their content has been removed, and that repeated violations will result in additional enforcement actions, including their account being disabled.
- Step three: our team records a strike against the Snapchatter's account.

A strike creates a record of violations by a particular Snapchatter. Every strike is accompanied by a notice to the Snapchatter; if a Snapchatter accrues too many strikes over a defined period of time, their account will be disabled.

This strike system ensures that Snap applies its policies consistently, and in a way that provides warning and education to users who violate our Community Guidelines. The primary goal of our policies is to ensure that everyone can enjoy using Snapchat in ways that reflect our values and mission; we have developed this enforcement framework to help support that goal at scale.

Partnerships

Snap has closely followed the negotiations on the EU AI Act and plans to continue to actively engage and assess collaboration opportunities on the upcoming AI Pact, as well as on the drawing of the related codes of practice for providers of general-purpose AI models and those regarding the detection and labelling of artificially generated or manipulated content.

More broadly, tackling risks stemming from generative AI requires (among others) broad industry-wide technical solutions which have not been clearly identified so far. This is why Snap is actively engaging with its peers and industry experts in different fora to share best practices and advance the technical debate. These partnerships, industry collaborations and efforts include:

- *OpenAI Integration*: My AI is powered by OpenAI's ChatGPT, and Snap closely partners with OpenAI in relation to providing the My AI service, including sharing feedback on moderation of content.
- *Tech Coalition / Working Groups on Generative AI*: Snap is a member of the Tech Coalition's Working Group on Generative AI Content, and a member of the GenAI Briefing Subgroup. The Working Group on Generative AI Content meets regularly to facilitate dialogue and information- and idea-sharing around mitigating content-level generative AI

risks. The GenAI Briefing Subgroup meets periodically to plan expert briefings for Tech Coalition members on topics related to Generative AI risks; such briefings have included representatives from government, law enforcement, civil society, and the research community.

- *Tech Accord to Combat Deceptive Use of AI in 2024 Elections*: Snap was an initial signatory to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. This compact seeks to set expectations for how signatories will manage the risks arising from deceptive AI election content created through their publicly accessible, large-scale platforms or open foundational models, or distributed on their large-scale social or publishing platforms in line with their own policies and practices as relevant to the commitments in the accord. The Accord was announced at the Munich Security Conference in February 2024.
- *ITI AI Futures Initiative*: Through its membership in the Information Technology Industry Council (ITI), Snap has participated alongside other private sector actors in the AI Futures Initiative. Led by technical and policy experts spanning the tech ecosystem, the Initiative is a forum through which participants are developing action-oriented recommendations for AI policy and working to address emerging questions around AI. Deliverables to date have included the issuance of Global AI Policy Recommendations to help guide governments around the world as to develop responsible regulatory approaches to AI-related issues.
- *HackerOne - Red-Teaming Collaboration*: Snap partnered with HackerOne on red teaming exercises to test the strict safeguards Snap has in place around AI. Together with HackerOne, we made significant developments in the methodology for AI safety red teaming that has led to a more effective approach to surfacing previously unknown problems. We refer to the HackerOne blog for more details: https://www.hackerone.com/ai/safety-vs-security
- As an active member of the EU Internet Forum, Snap will support the upcoming dedicated working group on generative AI matters.
- We are also members of the Centre for Information Policy Leadership (CIPL) and the Future of Privacy Forum (FPF) which work with industry stakeholders (like Snap), NGOs and government agencies in each region to advance a broad array of information topics. CIPL has been a leader in AI matters for many years through its dedicated AI Project and specific Brazilian AI Project. Most recently, in Europe, CIPL has responded to the UK Information Commissioner's Office (ICO)'s consultations on Generative AI, and led various forums on Accountable Governance of AI and AI Regulation in Brussels and the UK. Similarly, FPF is working on AI Governance and other responsible Gen AI initiatives.

5.9.4 Conclusion

Content authenticity is a very challenging topic without a silver bullet. Snap is very conscious of the issues and has implemented a number of measures (including relevant measures identified in the Commission's guidelines concerning elections). Snap continues to carefully monitor

developments and industry practice, including regarding whether and how best to use prominent markings and other measures to distinguish content that falsely appears to be authentic or truthful.

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

5.10.1 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 completed report form with details of the potential violation via email to a dedicated, confidential email address. The email used is a high priority channel and reports are reviewed in less than 48 hours (with reports relating to the most serious harms prioritized and reviewed well within this timeframe). Once a decision has been reached, Snap informs the Trusted Flagger that appropriate action has been 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).

Upon receiving notice from a DSC about a proposed Trusted Flagger, we will let the DSC know if we have any questions or concerns. Typically, when we are considering accepting a new Trusted Flagger we take into account geographic coverage, area of expertise, anticipated volume of reports, among other factors. Once a Trusted Flagger is accepted we send them an onboarding package, which includes an overview document of the program, including our commitment to review reports in less than 48 hours (with reports relating to the most serious harms prioritized and reviewed well within this timeframe); instructions on how to file reports to our dedicated and confidential email address; a metrics document to keep track of how many reports they've filed and among which harm categories, as well as contact information in case they have questions or concerns about the program.

When a Trusted Flagger wants to file a report, they leverage our instructions on which categories of information they should include in their reporting email. After we receive an email, Snap's Trust and Safety teams review the report and take any appropriate enforcement action, or request additional information if required for full investigation. Once a decision has been made, Snap will inform the Trusted Flagger that appropriate action has been taken as a result of the report.

Our teams remain in contact with the Trusted Flaggers, including when we need to discuss any issues with their reports. Our team evaluates the reports submitted based on the completion of the form, the accuracy of the information provided, and whether or not the report leads to enforcement or other action. If our team identifies trends that are impacting the quality of the reports that cannot be resolved with a Trusted Flagger, we will communicate this information with the DSC to identify a resolution.

We are monitoring the Commission's publication pursuant to Article 22(5) of the entities that have been awarded the status of 'trusted flagger' pursuant to Article 22(2) DSA.

5.10.2 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

5.11.1 Overview and Approach

We invest significant resources in our community support teams who work to resolve queries and complaints received from Snapchatters and others. In line with DSA requirements (Article 21), we are informing our users in Section 19 of our Terms of Service, entitled Dispute Resolution and Arbitration, about the possibility of out-of-court dispute settlement in case they are not satisfied with the outcome of the internal complaint-handling system. As outlined above, we are also referencing the possibility of out-of-court dispute settlement in our notices.

As of 14 August, we are aware of four bodies that may have been established and certified by the DSCs. However, we note that the DSCs are required to notify the Commission that they have certified these bodies and the Commission is required to publish the bodies on their website under Article 21.8. However, as at 14 August 2024, we have not seen the Commission publish

such a list. As such we have not yet been able to confirm whether any bodies have been formally established and certified by their DSCs.

Once confirmed, users and others will have the option to contact such out-of-court dispute settlement bodies to raise their case. The relevant out-of-court dispute settlement bodies will then be able to reach out to Snap via our dedicated contact point (dsa-enquiries@snapchat.com) to start the out-of-court dispute settlement process. Snap will then engage, in good faith, with the selected certified out-of-court dispute settlement body with a view to resolving the dispute following Snap's policies and procedures.

5.11.2 Enquires

As of 14 August, we have received two queries from one body purporting to be a certified DSA out of court settlement body.

5.11.3 Conclusion

Snap remains committed to resolving user disputes effectively and in line with DSA requirements. Furthermore, we continue to 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

5.12.1 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. In particular, Snap is active member of the following groups:

- **EU Internet Forum** - Snap is an active member and contributor of the EU Internet Forum (EUIF), which provides a collaborative environment for EU governments, the internet industry, and other experts and partners to discuss and address the challenges posed by the presence of malicious and illegal content online. The EUIF aims at exploring possible responses against abuse and exploitation of online platforms by terrorists and violent extremists, as well as other malicious actors, including those that groom children for the purpose of sexual abuse and the production and dissemination of child sexual abuse

material online. Earlier this year, the scope of the EUIF work was expanded to tackle also the issues of drug sales online and the trafficking of human beings.

- **Technology Coalition** - Snap is also a member of the Technology Coalition, which is an alliance of global tech companies who are working together to combat child sexual exploitation and abuse online. The Tech Coalition coordinates industry's overall effort to combat child sexual abuse online. It provides resources, education, and capacity-building to tech companies, and serves as a resource for external stakeholders - from global policy-makers to members of the media - on what industry is doing to tackle this issue.
- **WeProtect Global Alliance** - Snap is a Board member of the Weprotect Global Alliance, which brings together the private sector, government and civil society to drive positive change to help protect children from sexual abuse online.
- **Alliance to better protect minors online** - Until the recent discontinuation of the initiative in July, Snap was also a member of the Alliance to better protect minors online. This self-regulatory initiative was aimed at improving the online environment for children and young people by steering debates and exchanges on the topic.
- **CIPL** - Snap is a member of The Centre for Information Policy Leadership (CIPL). This is a global privacy and data policy think and do tank based in Washington, DC, Brussels and London. We work with CIPL and other industry leaders, regulatory authorities and policy makers to develop global solutions and best practices for privacy and responsible use of data, including with respect to teenagers and young adults.
- **The Future of Privacy Forum** - Snap is a member of The Future of Privacy Forum (FPF). FPF is a non-profit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. We work with FPF on a range of matters, including developing best practices related to Augmented Reality (AR), Artificial Intelligence (AI), biometric data, children's rights, and more.
- **Centre on Regulation in Europe (CERRE)** - Snap is a member of CERRE, which is a not-for-profit think tank based in Brussels. Its goal is to support and inform about regulation in Europe and beyond. We work with CERRE on in-depth reports and issue papers that address the major regulation challenges and high-quality, policy-oriented research undertaken by top-level academics in the tech, media and telecom sector.

Other Cooperation with industry

Snap is 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.

  - **DOT Europe** - Coordination on EU privacy, security,safety, content policy issues
  - **ITI** - Coordination on EU privacy, security and safety policy issues

5.12.2 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 interventions where we see the highest risks.

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 6th 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 (DG Just) 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.

Article 28 Guidance / EU AAD Code

Since our VLOP designation, we engaged with different stakeholders at the European Commission[86] to flag our interest to contribute directly to the discussion and working group on the AAD Code.

We understand focus has shifted to preparing guidance on Article 28 of the DSA for online platforms on ensuring a high level of privacy, safety and security for minors. We have recently submitted our feedback to the Commission on its proposals for guidance on Article 28 for online platforms.[87] 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 any Article 28 guidance or European wide AADC should consider in particular the role of online platforms that are 'gateways' (such as device operating systems, app stores and web browsers) through which parents and Teens engage with such platforms. For example, it would make logical sense to further enhance age assurance at the App Store and device operating system levels to ensure a robust upstream solution and support the measures taken by third party services (like Snapchat).

Critically, Snap consistently focuses on the potential impact to key areas when evaluating the effectiveness of age-assurance measures. Such factors include the tradeoff between the safety of minors and compromising user privacy/data security, the accuracy and reliability of age estimation technology (particularly for younger or ethnic minority populations), the fairness of methods that may disadvantage vulnerable segments including users without official government IDs or bank accounts, and the harm to industry competitiveness from the exorbitant cost of adopting third-party technology at scale.

We see the proposed Article 28 guidance, or any European wide AADC, as an excellent opportunity to develop an industry-wide solution on age assurance, which could effectively support the DSA implementation (drawing support from DMA to ensure interoperability and portability is provided by gatekeepers). We believe that such a solution should be developed at device level by operating systems, which would create a signal and make it available to services

[86] 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.

[87] The feedback can be found here: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14352-Protection-of-minors-guidelines/F3496394_en

operating in their environments. Our proposed approach has attracted the interest of different regulators, including recently in the UK, France and Germany (amongst others), and we would be happy to further expand on that if helpful. We have also been proactive in driving industry wide discussion through initiatives such as the CIPL / WeProtect Multi-stakeholder Dialogue on Age Assurance.

5.12.3 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.
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 (and we work closely with our global trusted flagger network on these matters). 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, our Content Crisis Response Protocol plays an important role in providing a structure to our collaborative internal operations and efforts.

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.

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, including a DSA Governance Group and meeting; (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 Advisory Groups (Safety Advisory Board and Council for Digital Well-Being).

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.

6.1 Platform Principles-based 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. The DSA Compliance Team and Cross-Functional Working Groups review Snap's approach periodically to ensure it is in line with DSA requirements and global best practices.

6.2 DSA Compliance Team and Cross-Functional Working Groups

This section sets out Snap's approach to the establishment of governance mechanisms over Snap's compliance with the Digital Services Act (DSA), fulfilling the requirements set out in Articles 11, 12, 13, 41, and 43.

6.2.1 Introduction

This section outlines Snap's governance mechanisms relevant to compliance with the requirements of Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (the "Digital Services Act" or "DSA"), in particular with regard to the requirements outlined in Articles 11, 12, 13, 41, and 43.

6.2.2 DSA Independent Compliance Function

Snap has established an Independent Compliance Function (ICF) as part of the Snap Legal team, separate from Snap operations, with sufficient authority, stature, and resources that is responsible for coordinating, overseeing, and implementing Snap's Privacy and Regulatory Program. The ICF provides oversight to ensure the necessary internal processes, resources, testing,

documentation, or supervision are in place for compliance with the DSA and monitors Snap's compliance with the DSA, ensuring the identification and mitigation of risks associated with Snap operations.

6.2.3 Independent Compliance Function Leads

Snap has designated Independent DSA Compliance Officers who report to the Management Body and fulfill the tasks outlined in DSA Article 41 for the head of compliance. Snap's compliance officers possess the professional qualifications, expertise, experience, and capabilities necessary to fulfill the designated responsibilities. The Independent Compliance Officers are also responsible for monitoring Snap's compliance with commitments outlined in the relevant codes of conduct or crisis protocols.

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.

Snap's Board of Directors appointed Snap's compliance officers in their 7/24/23 meeting.

6.2.4 Compliance Officer Qualifications

Snap's Compliance Officers have numerous years of experience in data protection, privacy, compliance and governance.

6.2.5 Operation of the Independent Compliance Function

Responsibilities of the Independent Compliance Function

The Independent Compliance Function is responsible for the following activities:

- Cooperating with the relevant Digital Services Coordinator and the Commission for the purpose of DSA compliance;
- Providing oversight over the development of the Systemic Risk Assessment methodology and the conduct of the Systemic Risk Assessment, ensuring it is conducted on the basis of the best available information and scientific insights
- Ensuring that risks referred to in DSA Article 34 are identified and properly reported on and that reasonable, proportionate and effective risk-mitigation measures are taken pursuant to Article 35;
- Organizing and supervising the activities related to the independent audit pursuant to DSA Article 37;

- Providing oversight over the validation of controls leveraged to mitigate risks, evaluation of controls, and review of policies;
- Monitoring the compliance with DSA obligations;
- Reviewing and approving transparency reports;
- Informing and advising the management and employees about relevant obligations under the DSA;
- Where applicable, monitoring the compliance with commitments made under the codes of conduct pursuant to Articles 45 and 46 or the crisis protocols pursuant to Article 48 of the DSA.

Oversight and Monitoring of Snap's DSA Compliance

The Independent Compliance Function utilizes multiple ways to exercise oversight and monitoring over Snap's DSA compliance.
Any relevant issues or observations are discussed within the Independent Compliance Function as well as with relevant stakeholders and escalated to the Management Body and the CEO as needed. The Independent Compliance Function investigates root causes of issues or observations, develops mitigation plans and works with stakeholders and control owners to implement such corrective actions. The Independent Compliance Function may also escalate significant issues / observations to the Management Body and the heads of the respective function if the root cause is identified as relating to that function.

DSA Management Body

Snap has designated a Management Body which oversees and supports the independence of the compliance function and manages issues as escalated to the Body.

6.2.6 DSA Cross-Functional Governance Team

Snap established a Cross-Functional DSA Governance team inclusive of senior personnel, which includes Legal, Public Policy, Product, Engineering, Trust & Safety, and Information Security teams.

Given the high stakes related to DSA compliance, given the multi-faceted nature of DSA requirements and given cross-functional ownership and responsibilities, it is important to introduce a DSA governance structure that reflects the complexity of the DSA.

The purpose of this cross-functional DSA Governance Team is 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.

The DSA Governance Team is responsible for managing, overseeing, monitoring, assessing and adjusting Snap's DSA compliance program. The DSA Governance Team meets on a monthly basis (the Team and/or its members might meet more frequently if necessary).

6.2.7 Points of Contact

Designation, Publication, and Change Management

Upon designation, Snap's cross-functional DSA governance team established a process for designating and updating in a timely manner the publication of points of contact for regulatory authorities, recipients of the service, and the legal representative in a location that is publicly available and easily accessible.

Snap's Cross-functional DSA Governance Team reviews Points of Contact and Legal Representative designation. If a new Point of Contact or Legal Representation is appointed, the Data Compliance Officer or designee will work with the web team to update the above website accordingly.

Points of Contact are made publicly available in all official languages of the EU and are easily accessible on: https://values.snap.com/privacy/transparency/european-union

Point of Contact for the Authorities

Upon designation, Snap's cross-functional DSA governance team designated a single point of contact email address for communication with Member State authorities, the Commission, and the Board to enable direct and effective communication between Snap and the relevant authorities on matters related to the Digital Services Act.

Upon designation Snap's cross-functional DSA governance team ensured the easily accessible publication of the relevant information relating to the point of contact for regulatory authorities, including the languages to be used in such communications, on the Snap website.

Authorities can contact Snap at dsa-enquiries@snapchat.com, through our Support Site, which supports all official languages of the EU, and at

Snap B.V.
Keizersgracht 165, 1016 DP
Amsterdam, The Netherlands

The European Commission also received contact details to communicate directly with the Head of Snap's DSA Compliance Function on all matters.
Law Enforcement can contact through mechanisms described here:
https://www.snapchat.com/lawenforcement

Point of Contact for Users

Upon designation Snap's Cross-functional DSA Governance team designated a point of contact for users to contact Snap regarding DSA-related matters that is user friendly.

Upon designation, Snap made information on the customer service POC publicly available in a place easily accessible to the user in a place where they would be expected to be and up to date.

For general DSA inquiries, Snap can be reached through the dsa-enquiries@snapchat.com email address as well as by submitting a ticket through our Support Site. The support site is available to support users in all official languages of the EU.

Legal Representative

Snap's Cross-Functional DSA Governance team has designated a legal representative (Snap B.V.) in one of the Member States where the service is provided for complying with DSA obligations and enforcement matters. This designation includes allocation of resources and authorities sufficient to cooperate with relevant authorities and comply with decisions issued by the European Commission in relation to the DSA.

We have notified our Digital Services Coordinator of the contact information for our legal representative, including the name, mailing address, phone number, and email, and ensured the publication of the information in a publicly available location.

Snap has appointed Snap B.V. as its Legal Representative for purposes of the DSA. Snap's legal Representation can be contacted at

Snap B.V.
Keizersgracht 165, 1016 DP
Amsterdam, The Netherlands

6.2.8 DSA Supervisory Fee

In its latest assessment, the European Commission has determined that Snap does not meet the threshold for the supervisory fee, and Snap's current contribution is set at EUR 0.

6.3 Privacy and Safety by Design

6.3.1 DSA Risk Management

As explained in Section 3, 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 this is also an industry standard practice to ongoing risk management and found in many other laws requiring risk assessments (including guidance relating to DPIAs).

Snap has developed a number of practices to detect and manage risks to Snapchat's in-scope services. As explained in this Section 6, these practices include: (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, Safety Advisory Board, and Council for Digital Well-Being.[88] These are covered in other parts of Section 6 of this Report.

Additionally, Snap has a 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. Snap 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 the privacy, safety and/or security of its users. The privacy and safety by design process is a collaborative and cross-functional process, and stakeholders from Snap's legal and privacy engineering teams are embedded in key phases of the product's development.

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,

[88] https://values.snap.com/news/inaugural-council-digital-well-being.

prioritize them and agree on strategy and execution plans to resolve identified risks. 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.

6.3.2 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. Additional information can be found in Section 5.1.

Snap already had an extensive privacy and safety by design review process to assess privacy and safety risks in the design and development of Snapchat prior to the DSA coming into force, and this continues to be the case. As part of its privacy and safety by design program, Snap documents a review prior to new product and feature releases that materially affect the privacy, safety and security of our users.

6.3.3 Holistic Digital 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, for example, GDPR (in the form of Legitimate Interest Assessments (LIAs) and Data Protection Impact Assessments (DPIAs)) 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.

In order to meet Snap's annual and ongoing holistic risk assessment obligations, including with respect to the DSA, Snap continues to use a Digital & Data Impact Assessment ("DDIA") framework that combines our privacy, safety and security obligations into a single risk assessment. This was designed prior to the DSA coming into force, and will in future cover other aspects such as the EU AI Act.

The DDIA includes a template that serves as a vehicle to conduct the various risk assessments. This template supports the consideration of risk and mitigations related to a specific Snapchat product and includes guidance for the consideration of key factors and influencers on that risk, such as the performance of Snapchat recommender systems, the intentional manipulation of the platform, and regional and linguistic considerations.

The DDIA templates are implemented at the product level and cover a range of requirements beyond the scope of the DSA.
The 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 launches. The DDIA is completed dynamically depending on the nature of the Snapchat feature and impact of the change being assessed.

If reviewers determine that the change does require an update to the DDIA, they will work with Snap's Legal team (and other cross-functional Compliance team that they engage as needed) to update the relevant DDIA accordingly.

6.3.4 DSA Critical Impact Check

Snap recognises that, as well as carrying out our annual DSA Risk Report, it must also re-assess risk 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 this is also an industry standard practice to ongoing risk management and found in many other laws requiring risk assessments (including guidance relating to DPIAs).

As part of the DDIA update review, product reviewers consider whether the change amounts to a critical impact.
As a result, we are able to detect and manage our DSA risk assessment and mitigation obligations on an ongoing basis. Since our 2023 Report, we have not identified any deployed functionalities that were likely to have a critical impact on our assessment of risks and mitigations pursuant to Articles 34 and 35 of the DSA.

6.4 Prevalence Testing

A key measure we have in place to holistically detect and manage risk of illegal and other harmful content on an ongoing basis is prevalence testing i.e. testing the 'Policy Violating Prevalence' (PVP) of Stories accessible to the public via random sampling. The sampling allows us to estimate the percent of policy-violating views and monitor the presence of illegal and other violating content on Snapchat. Through this prevalence testing, we are able to uncover blindspots and prioritize efforts to close those gaps through improvements to our proactive detection mechanisms, infrastructure improvements and agent training.

6.4.1 Overall

Our overall PVP metrics demonstrate that the effectiveness of our proactive detection mechanisms, agent training and other content moderation and enforcement efforts has increased significantly since our 2023 Report.

6.4.2 Example mitigations

In our previous report, we identified the following examples of mitigations we have taken to reduce prevalence

An example of the further mitigations we have taken since our 2023 report, is an improvement to the efficacy and reduced latency of content enforcement mechanisms. This has involved measures to temporarily reduce exposure of content suspected of being illegal or otherwise violating our terms while in review. The vast majority of the content to which these brief temporary measures apply are subsequently enforced and as a result, this has helped us to further reduce the percentage of violating views of violating content.

6.4.3 Conclusion

In conclusion, prevalence testing continues to be an extremely valuable measure for our ongoing detection and management of content risks. Our ongoing efforts to improve our prevalence testing and our mitigations to reduce prevalence of illegal and other violating content has resulted in very significant decreases in PVP rates). As noted in the conclusion of our 2023 Report, there is always more to do as we ultimately aim to reduce prevalence across all our violating content categories to as close to zero as possible.

6.5 External Request Monitoring and Review

As noted in our 2023 Report, we produce a semiannual (every 6 months) Transparency Report, that captures our Community Guidelines enforcement data, law enforcement operations data, and copyright & trademark data. 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. Internally we also continue to review additional breakdowns of this data and, in preparation for our DSA compliance, we continue to review data relating specifically to the European Union's individual Member States.

We also continue to monitor advertising review rejections, advertising reporting and enforcements, 'privacy, data protection and DSA' requests and general community support requests. Since the DSA came into force for Snapchat on 25 August 2023, we have also monitored queries relating to compliance received via our dedicated dsa-enquiries@snapchat.com email address. This dedicated contact point is published on our website here, pursuant to Articles 11 and 12 of the DSA.

As with our 2023 Report, we have continued to review and use this external request data to support the conclusions reached in this Report.

6.6 Digital Well-Being Index (DWBI) Initiative

6.6.1 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 asked about the risks and potential harms teens and young adults are encountering online across all platforms, services and devices, not just Snapchat. We conducted the research in six countries – Australia, France, Germany, India, the UK, and the U.S, which includes three of the largest European countries, two of which are in the EU) – and also included parents of teenagers between the ages of 13 and 19.

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.

In our 2023 Report, we explained that we had repeated and expanded this research in 2023 ("Year Two"). For more about Snap's Digital Well-Being Index and research, see: Our website, as well as this explainer, the full research results, and each of the six country infographics: Australia, France, Germany, India, the United Kingdom and the United States. We took account of this and the previous year's research when conducting our assessment of risk and mitigations as highlighted in our 2023 Risk Report, and have continued to do so in this Report.

In this section we provide highlights from the Year Three research and Generation Z's views on reporting. In 2023, a total of 9,010 people participated in the study, and in 2024, respondents totaled 9,007 across the six countries referred to above. Our Year Two deep-dive findings were highlighted by several research institutions and organizations and, upon endorsement from leading child-safety NGOs, we have repeated this portion of the study in 2024.

6.6.2 Snap's Digital Well-Being Index - Year Three

In this Report, we have continued our research during 2024 ("Year Three"). In addition to repeating our 2023 work, our Year Three industry wide research has also investigated teen's and young adult's attitudes and sentiments around reporting problematic content to platforms and services, authorities and others. This offers insight into teens' and young adults' attitudes and sentiments around reporting problematic content to platforms and services, authorities and others.

6.7 Snap Advisory Groups

6.7.1 Introduction

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. The initiative 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 currently stands at 19 members, based in 10 countries and representing 11 different geographies and regions, 4 of the Members are EU-focussed. The board comprises 14 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. The SAB meets three times annually: twice virtually and once, in-person at Snap headquarters for an in-depth strategy session to help prepare Snap for the coming year's planning.

6.7.2 Updates

Since our 2023 Report, we have continued to progress our work with the Snap Safety Advisory Board and have also established a new Snap Council for Digital Well-Being.

Safety Advisory Board

In addition to the deep investment Snap is making in its SAB cohort, it periodically consults a cadre of some 50 safety experts from around the world on new product features and functionality, policies, and initiatives. 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. This year, collaborations focused on the U.S.-based National Center for Missing and Exploited Children (NCMEC) on the topics of sextortion and improving CyberTip reports to NCMEC, as well as smaller, executive-attended sessions with WeProtect Global Alliance, IWF, Thorn, and others. 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.

Snap Council for Digital Well-Being

Snap formally launched a new youth-focused program in January 2024 - Snap's first "Council for Digital Well-Being (CDWB)". The Council is a pilot program in the U.S. designed to encourage safer online habits and practices among teens with the aim of having these young people champion their knowledge and insights in their schools and communities.

For the first interaction of this program, we opened applications from U.S.-based teens, aged 13 to 16. The inaugural cohort was selected and announced in May. Following two virtual monthly meetings, the group traveled to Snap HQ in Santa Monica, California, USA, in July 2024 for the first Council Summit. The 2-½-day program consisted of breakout sessions for both the teen cohort and their parents/chaperones, full-group discussions, and guest speakers. The teens also got a glimpse into working at a technology company, as they were treated to a 90-minute "speed-mentoring" session with 18 Snap employees representing different roles and teams.

The Summit yielded interesting conversations and insights on topics such as online pitfalls, parental tools, and the differences and similarities between digital and in-person social dynamics. By the end of the Summit, the full group, chaperones included, was extremely motivated to be more involved in their own local communities and to act as ambassadors for online safety. We shared with them some of our outreach material to aid their efforts, including the below infographics on reporting and Snap Safety Milestones:

Safety Concerns

2024

Reporting a Safety Concern

Snap empowers people to express themselves, live in the moment, learn about the world, and have fun together. One of the most important things Snapchatters can do to help keep the service free of bad actors and potentially harmful content is to **reach out** to us when you encounter something that makes you uncomfortable. All you need to do is **press and hold** on the piece of content or the chat message and a menu will appear. Then, tap **"Report"** to see a menu of options.

Our safety teams work 24/7 to review reports made on Snapchat or through our Support Site and, once reviewed, our safety teams will take action on content and accounts that violate our Community Guidelines or Terms of Service. It's important to remember that reporting is confidential and the account-holder you reported won't be told who reported them. If you encounter anything that appears to be illegal or dangerous, or if you have reason to believe someone is at risk of harm or self-harm, **contact local law enforcement** immediately and then report it to Snapchat, as well.

You can read through Snapchat's Community Guidelines and our Terms of Service to familiarize yourself with what content is permitted on Snapchat. A good rule of thumb: if what you're saying could create an unsafe or negative experience for someone, it's better left unsaid.

Also, if you see something you don't like on Snapchat, but it may not violate the Community Guidelines, you can choose to **unsubscribe, hide the content,** or **unfriend** or **block the sender.**

Your Common Questions Answered

Is reporting on Snapchat confidential?

Yes. We do not tell other Snapchatters when you make a report.

Who reviews my submitted report?

When you report a concern on Snapchat, you receive a confirmation that your report has been submitted. Behind the scenes, our safety teams work 24/7. If the teams' review confirms a violation of our Community Guidelines or Terms of Service, the content will be removed and we may even lock or delete the account, and report the offender to authorities.

Does Snapchat tell the reported account who reported them?

No. The account holder that you reported won't be told who reported them. All reports are strictly confidential.

If I block or remove someone, will they know?

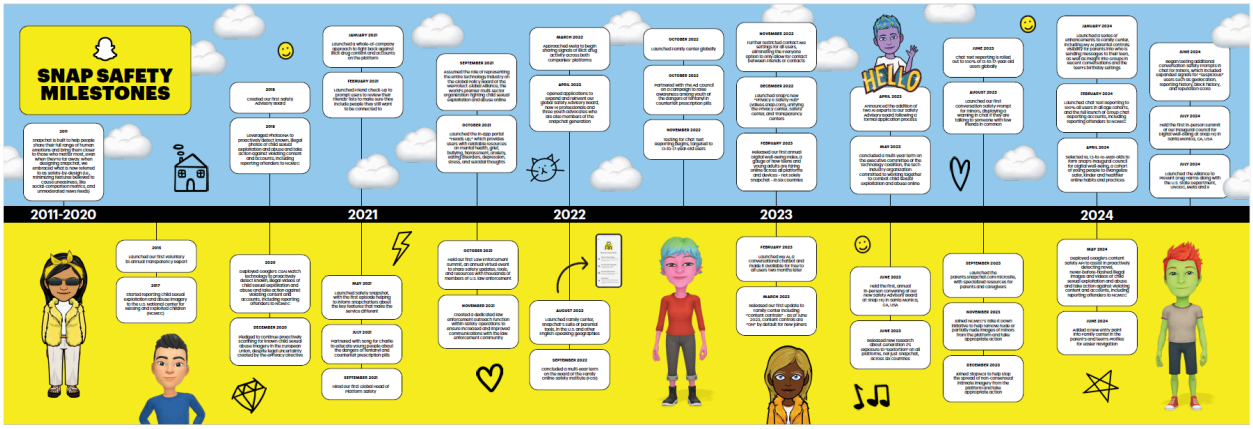
When you block or remove someone from your Friend List, they are not formally notified, but they may be able to infer this when their messages are no longer reaching you.

Does Snapchat alert me if someone reports me?

If we take action on your content that was reported, you may be alerted in our app or via email.

I reported something on Snapchat but it wasn't taken down. Why is this?

Not all reported content is removed. We remove content that violates our Community Guidelines or Terms of Service. If you see content that you don't like, but is permitted according to our Community Guidelines or Terms of Service, you can avoid seeing it by adjusting your privacy settings, hiding the content or blocking and removing the sender.



The program will continue through next summer, and we look forward to continued engagement with this cohort, their parents and chaperones, and the Snap SAB.

6.8 Audit

We recently completed the 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.

It is also worth noting, as we did in our 2023 Report, that Snap already has another annual audit process in place. Snapchat is subject to an independent third party audit pursuant to Snap's 2014 FTC Consent Decree in the United States. 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. This work involves stakeholders across the company, including Ops, Product, Engineering, Customer Support, Privacy Engineering, Legal etc, and provides another layer of external audit of Snap's practices, and further reassurance that Snap's practices are appropriate.

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(1); and (ii) the review of the specific mitigation measures that 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(1) 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 and Section 6 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:** Since our 2023 Report, we have observed a further substantial reduction in the prevalence of content that is illegal or otherwise violating Snap's Terms being disseminated on Snapchat's online services in general. We have observed this content to now be at a very low level compared to the prevalence of this content on websites and other online spaces. For the first time, none of our illegal or other violating content categories were observed from our testing to have a prevalence rate that fell within our highest relative prevalence category. Within this very low level of dissemination in general:

    a. We have continued to categorize 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 again confirmed we have reasonable, proportionate and effective mitigation measures for all three of these categories. As a result of these ongoing measures, all three are still assessed to fall within our extremely low likelihood category of the risks identified by Snap.

    b. We have continued to categorize 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 serious harm each may cause. We have again confirmed we have reasonable, proportionate and effective measures in place for

all five of these categories. As a result of these measures four are still assessed to fall within our extremely low likelihood category of the risks identified by Snap. In the case to terrorism content, we have noticed a slight increase in prevalence since our last Report, and, while this is consistent with the October 7, 2023 Hamas terrorist attacks and ongoing conflict in Gaza and Israel, we continue to carefully monitor this risk category to ensure the slight increase is temporary and prevalence remains very low. In the case of harassment & bullying content, although we have observed a significant fall in prevalence and reporting, from our testing of the in-scope services of Snapchat, we will continue to carefully monitor this risk category as we look to achieve further reductions.

c. We have continued to categorize 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 again confirmed we have reasonable, proportionate and effective measures in place for all eight of these categories. As a result of these measures six of these are still assessed to fall within our extremely low likelihood category of the risks identified by Snap. **With respect to fraud and spam and adult sexual content, since our 2023 Report we have observed a substantial reduction in prevalence rates and both now fall within a lower likelihood category. We are continuing to monitor both categories as we look to achieve further prevalence reductions (and to ensure the consistently higher proportion of ad rejections for fraud and spam are not an indication that further mitigating measures might be required with regards to ads).**

2. **Negative effects on EU Fundamental Rights:** We continue to categorize: (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 (c) 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 again confirmed we have reasonable, proportionate and effective measures in place for all of these categories. **We continue to actively participate in efforts to develop EU wide guidance to assess if further industry measures are needed to address risks to child rights.**

3. **Negative effects on Public Security:** We continue to categorize 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 again confirmed we have reasonable, proportionate and effective measures in place for all of these categories.

4. **Negative effects on Public Health:** We continue to categorize: (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 again confirmed we have reasonable, proportionate and effective measures in place for all of these categories. **As above, we continue to actively participate in efforts to develop EU wide guidance to assess if further industry measures are needed to address risks to the protection of minors.**

It is Snap's mission to reduce and maintain a lower prevalence of illegal and otherwise violating content on Snapchat's inscope services. We have been successful in increasing the granularity of data that we rely on, which we have incorporated into this Report. We will continue to progress this action to ensure that risks can be tracked with even greater precision across in respect of each of Snapchat's in-scope services.

Regarding our Mitigations, since our 2023 Report, we have worked hard to: (i) evaluate our mitigation measures against new guidance from the Commission (in particular the Guideline for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes and the updated EU Code of Conduct to counter illegal hate speech online) and (ii) provide additional information on our measures to reflect the increased use of generative AI technology such as updates to our content moderation policies (note that many of these measures relate to out of scope services on Snapchat). Where required by the DSA, we have made available our mitigations in all of the languages of the European Union and in all other cases in every language in which Snapchat is available.

Since our 2023 Report, we did not identify any deployed functionalities that were likely to have a critical impact on our assessment of risks and mitigations pursuant to Articles 34 and 35 of the DSA. As described in the Ongoing Risk Detection & Management section above, our DSA Governance Team continues to regularly evaluate the effectiveness of its measures as we look to maintain or further reduce prevalence, detect any new risks, assess any deployed functionalities for critical impacts and determine whether further mitigating measures might be required.

In summary, we have carried out our second annual risk assessment of Snapchat's in-scope services required by Article 34(1) of the DSA. We have observed further significant reductions in the prevalence across our illegal and otherwise violating content categories. We continue to conclude that we have reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified, as required by Article 35(1) of the DSA. There are three risk categories that we continue to monitor to ensure that remains the case.

8. Final Words

This is the second 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. Snap has continued to take a comprehensive approach to the obligations in Articles 34, 35 and 42. As in our 2023 Report, although there is still no settled legal definition of 'systemic risk', we have again adopted the position that all the risks identified in the DSA are systemic to online platforms (which is why they have been identified in the DSA). We are still then looking 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.

As noted in our 2023 Report, our position 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 this is demonstrated again in this Report which concludes that Snapchat represents an even lower risk profile than identified in our 2023 Report. This is due to its unique design and function, but also in particular to the efforts of our cross-functional teams who have worked hard to provide high levels of privacy, safety and security for all our users and further substantial falls in the risks specifically referred to in the DSA. We are particularly proud of the **significant decrease** in the overall prevalence rate for our illegal and harmful content categories.

We were pleased to receive positive feedback from the Commission on our 2023 Report. We look forward to again receiving feedback from the Commission on this second report, as well as the publication of our first risk assessment and mitigation report and our audit report.

Annex 1 - 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