

# Digital Services Act Risk Assessment and Mitigation Report 2024



---

# Contents

<b>1. Executive Summary</b>	<b><a href="#">3</a></b>
<b>2. Introduction</b>	<b><a href="#">4</a></b>
2.1 Pinterest	<a href="#">4</a>
2.2 The Digital Services Act	<a href="#">4</a>
2.3 What's changed since last year?	<a href="#">5</a>
2.4 Updates on 2023 Mitigation Measures	<a href="#">6</a>
<b>3. Risk Assessment Methodology</b>	<b><a href="#">8</a></b>
3.1 Understanding Pinterest's systemic risk landscape	<a href="#">8</a>
3.2 Risk Assessment Methodology	<a href="#">9</a>
<b>4. Pinterest's platform ecosystem</b>	<b><a href="#">10</a></b>
4.1 Influencing factor 1: Applicable terms and conditions and their enforcement	<a href="#">10</a>
4.2 Influencing factor 2: Content moderation systems	<a href="#">11</a>
4.3 Influencing factor 3: Design of recommender systems and any other relevant algorithmic systems	<a href="#">15</a>
4.4 Influencing factor 4: Systems for selecting and presenting advertisements	<a href="#">16</a>
4.5 Influencing factor 5: Data practises	<a href="#">16</a>
4.6 Influencing Factor 6: Intentional Manipulation	<a href="#">17</a>
<b>5. Systemic Risk Landscape</b>	<b><a href="#">18</a></b>
5.1 Systemic Risk 1: Illegal Content	<a href="#">18</a>
5.2 Systemic Risk 2: Negative effects on the exercise of fundamental rights	<a href="#">23</a>
5.3 Systemic Risk 3: Negative effects on civic discourse, electoral processes and public security	<a href="#">28</a>
5.4 Systemic Risk 4: Negative effects in relation to gender-based violence, the protection of public health and minors, and serious negative consequences to the person's physical and mental well-being	<a href="#">32</a>
<b>6. Conclusion</b>	<b><a href="#">37</a></b>
<b>Appendix</b>	<b><a href="#">38</a></b>

---

# 1. Executive Summary

Pinterest is a visual search and discovery platform where people find inspiration, curate ideas and shop products — all in a positive place online. At Pinterest, our mission is to bring everyone the inspiration to create a life they love, and that’s our guiding light for everything we do, from the way we draft and enforce our content and advertising policies to the design of our platform and new features we launch. Pinterest does not prioritise endless engagement or virality. Our users (“Pinner”) come to Pinterest to find ideas for recipes, beauty, fashion, home decor and more. We believe that safety and protecting our community is central to creating a positive and inspirational online platform. We have established [Community Guidelines](#) as our “rules of the road” that clearly outline what is and isn’t allowed on Pinterest, and we work hard to identify and remove harmful content from our platform.

The Digital Services Act (“DSA”) came into force on 16th November 2022 with the aim of giving better protection to users and to fundamental rights online, establishing a transparency and accountability framework for online platforms, and providing a single, uniform framework across the European Union (“EU”). The DSA requires Pinterest as a Very Large Online Platform (“VLOP”) to identify, analyse, assess and mitigate certain systemic risks stemming from the functioning and use of Pinterest in the EU.

Following our risk assessment framework, we’ve assessed that the overall risk that Pinterest poses to our users and to society remains **Low**. Our second annual systemic risk assessment considered similar factors as our first assessment in 2023, including the reasons that users come to Pinterest, the ways they use the platform and interact with others, and the limited opportunities for content to go “viral” on Pinterest. For our 2024 assessment, we also considered changes to functionalities on our platform, updates to our tools and systems, business strategies and priorities, and events in the external environment, such as global elections and the increase in the prevalence and use of Generative Artificial Intelligence (“GenAI”). As before, we also looked at the controls we have in place to prevent these risks from occurring, from our [Community Guidelines](#) and other related policies to the robust processes we use to detect and take appropriate action on harmful content and behaviour.

In carrying out this assessment, we conducted interviews and surveyed a broad range of internal Pinterest stakeholders, reviewed internal documentation and leveraged data — including information reported in our global and DSA transparency reports — and gathered input from a range of external stakeholders, such as users and industry groups. Based on this review, we identified potential risks associated with the design, functioning or use of Pinterest and assessed how these might impact EU users and the broader EU society. We’ve grouped these risks into four main categories:

1. Illegal content;
2. Negative effects for the exercise of fundamental rights;
3. Negative effects on civic discourse, electoral processes and public security; and
4. Negative effects in relation to gender-based violence, the protection of public health and minors, and serious negative consequences to the person’s physical and mental well-being.

We’ve also considered if and how Pinterest’s design, functionality or use influence these systemic risks. We considered Pinterest’s overall ecosystem, including: applicable terms and conditions and their enforcement, content moderation systems, design of recommender systems, systems for selecting and presenting advertisements, related data practices, and intentional manipulation.

Although our current overall assessment of the risk that Pinterest presents to EU users and society is **Low**, we’re continuously improving our control environment. We’re actively making enhancements to specifically address these four risk categories, including a continued focus on user safety, enhancing our current QA and training programmes, enhancements to our mental health supports, and additional measures to make Pinterest an inspiring and inclusive place for all our users.

User safety remains critical for Pinterest and is championed by our CEO, Bill Ready, both in public and within the company as a strategic priority. The safety of our users is key to the success of our business and we strive to constantly evolve and enforce better content safety policies and make deliberate choices to set an example as a more positive place online.

---

## 2. Introduction

### 2.1 Pinterest

#### Product Overview

Pinterest is a visual search and discovery platform for finding ideas like recipes, home and style inspiration, and more. People around the world, including in the EU, use Pinterest to bring ideas to life, shop products and discover inspiring content. Pinners come to Pinterest to discover Pins they love and save them to boards to keep their ideas organised and easy to find. There are billions of Pins on Pinterest, allowing our users to find ideas to spark inspiration. Pins can be images, videos or products.

Pins can be saved and organised on boards. Pinners can name their boards, arrange them on their profile however they want and organise their Pins in a way that makes sense to them. They can also invite other people to collaborate on group boards to find even more ideas. All Pinners have the option to set their profile to private, and private profiles won't appear in search results on Pinterest or in search engines.

A Pinner's home feed is where we suggest Pins, creators, products and merchants we think they'll love. The home feed will also show Pins from the creators and boards they've chosen to follow. Users can save, create, share and shop Pins, explore suggested topics and trends or search for topics of their own, and collaborate with others.

#### Pinterest's mission and content approach

At Pinterest, our mission is to bring everyone the inspiration to create a life they love. User safety is critical to our mission, especially when it comes to the safety of young people, and our content moderation practices and [Community Guidelines](#) are always evolving to be current, including to keep up with new behaviours and trends and to create a more positive place on the internet.

Pinterest is working towards making the internet a safer and healthier place for everyone, especially young people. In partnership with the Digital Wellness Lab at Boston Children's Hospital, Pinterest helped create and signed the [Inspired Internet Pledge](#), which provides a framework for companies to take "meaningful, measurable actions to support positive mental and emotional well-being outcomes both on and offline." The pledge is a call to action for technology companies to make the internet a safer and healthier place, and Pinterest is proud to be a founding signatory.

### 2.2 The Digital Services Act

The goal of the DSA is to "create a safer online experience for citizens to freely express their ideas, communicate and shop online, by reducing their exposure to illegal activities and dangerous goods and ensuring the protection of fundamental rights."<sup>1</sup> Articles 34 and 35 of the DSA lay out the requirements for Pinterest as a VLOP to identify, analyse, assess and mitigate certain systemic risks stemming from the design, functioning and use of Pinterest in the EU.

This report describes the results of Pinterest's second DSA systemic risk assessment and associated mitigation measures, including an overview of Pinterest's risk landscape, ongoing mitigation efforts to address those risks and updates to the further mitigation measures we identified in 2023. While most of the policies and controls we discuss in this report are global in nature, the scope of this report and our risk assessment is limited to the EU.

<sup>1</sup> [https://ec.europa.eu/commission/presscorner/detail/en/QANDA\\_20\\_2348](https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348)

## 2.3 What's changed since last year?

Our second annual DSA systemic risk assessment remains grounded in our bedrock risk assessment methodology. We've examined new risks posed by external events and emerging technology and have incorporated learnings from the first year of DSA enforcement. For example, this year, we considered new risks associated with global elections, especially those in the EU, and the growing prominence of GenAI. We also considered new Pinterest products and functionalities and provided an update on the risk mitigation measures identified in last year's report.

### Global Elections

There are more than 50 elections happening around the world in 2024. Within the EU, there have been national or local elections in 11 member states as well as elections for the European Parliament, and more elections are due to take place before the end of the year.

We know that our Pinnerers don't come to Pinterest for politics — they come for inspiration on topics like recipes, beauty and home design. Nevertheless, we understand the implications of harmful content related to elections and civic discourse and we strive to make sure that this type of violative content is identified and actioned on our platform. Pinterest has been on the forefront in this area — since 2018, we have banned political campaign advertising including political candidates, political parties and political issues with intent to influence an election (full information on this can be found in our [Advertising Guidelines](#)). In addition, our Civic Participation Misinformation policy prohibits false or misleading content on Pinterest that impedes an election's integrity or an individual's or group's civic participation, which includes activities like registering to vote, voting and being counted in a census.

We have elections-focused teams who closely monitor elections across the globe and work to take proactive and reactive measures in response to elections, ensuring our policies and systems remain current. We have also participated in roundtable discussions on election integrity hosted by the European Commission to review lessons learned from the European elections and plans for future improvements.

More information on this risk can be found in the discussion of Systemic Risk 3 ([Negative effects on civic discourse, electoral processes and public security](#)).

### Artificial Intelligence

Artificial intelligence ("AI"), and in particular, GenAI, is a technology that has significantly grown in prominence since our last DSA risk assessment. Pinterest has always used traditional forms of AI (machine learning or machine learning models) as a component of our recommender systems and content moderation models. We use human reviewers to train our machine learning models and we continually monitor and evaluate those models to make sure they're working as intended.

GenAI is a complex area and has many potential uses, from automating processes to generating new content. In completing this year's DSA risk assessment, we took into consideration the targeted and intentional ways in which Pinterest's product uses GenAI in the EU and what risks stem from such use. At this time, Pinterest does not have GenAI features that allow individual users to generate content.

Pinterest, in using GenAI, has also taken a number of steps to ensure the safety of our users. For example, for purposes of training GenAI models, we do not use images appearing in private Pins or user profile information, and our GenAI systems have a number of technical safety measures in place.

In looking at the risk of GenAI, we also have to take into account manipulated or edited content that might appear on our platform and the existing controls we have to identify this type of harmful content. Our [Community Guidelines](#) apply to all content on the Pinterest platform regardless of how it has been created. We've addressed intentional manipulation for each of our systemic risks below.

We have also partnered with the Digital Trust and Safety Partnership ("DTSP") and participated in their GenAI and Automations Working Group. The output for this Working Group will be to publish a paper on best practices for GenAI and automation in trust and safety.

## Further Mitigation Efforts for GenAI

In addition to the controls and mitigation measures already in place for our use of GenAI, we plan to introduce additional risk mitigation measures to guide us and our users in the area of GenAI. The first of these measures is to provide more transparency to our users and our partners around our approach to responsible use of GenAI. In addition to providing this transparency and the existing information in our [Privacy Policy](#) on machine learning, we're working on a Help Center article that will be available to our users and will give them details on what GenAI is, how Pinterest uses GenAI and how this might impact them.

We're continuing to enhance our content moderation systems, which also enhance our risk mitigation efforts for [Community Guidelines](#)-violating GenAI-created content.

## Updates to our platform

While our mission and values remain the same, we've added some new features to our platform to make sure that we continue to be a positive place online where people can find the inspiration to live a life they love.

Since our last risk assessment, we've expanded our inclusive product feature offerings with a new feature called body type ranges, which uses our body type technology to give users the choice to self-select what body types are featured in their search results for women's fashion and wedding ideas.

In addition, we're leveraging GenAI technology to support creative, improved campaign performance for advertisers by using GenAI technology to enable creation of scenic lifestyle backgrounds on product images.

We've also made it easier for users to share their inspiration with others with our new board sharing feature that lets users share dynamic videos of their favourite Pinterest boards to other social platforms.

All of these new features and functionalities have been taken into consideration for the purposes of this year's risk assessment.

## 2.4 Updates on 2023 Mitigation Measures

In our 2023 DSA Risk Assessment and Mitigation Report, we identified five areas for additional mitigation measures. Updates on those areas are provided below:

### 1. Recommender Systems

In last year's risk assessment, we noted that we had begun to implement methods to ensure that our recommender systems do not learn from harmful content or bad actors. Since then, we have further enhanced these systems.

Our Trust and Safety team provides our Recommender Systems teams with a daily output of users and content that have been identified as policy-violating. The Recommender Systems teams then exclude these users and content from model training. This means that Pins, boards and users that have been deactivated due to [Community Guidelines](#) violations are excluded from the data used by our recommender systems. This process was finalised in Q4 of 2023. Full details on actions taken on policy-violating content are detailed under [influencing factor 2](#).

### 2. Automated Models

The safety of our users is top of mind for Pinterest. Last year's risk assessment highlighted that investment in our machine learning models was a key component in identifying and actioning harmful content.

Since then, we have continued to invest in and make enhancements to machine learning for content moderation. Since the start of H2 2023, we have launched or updated 16 machine learning models covering categories such as adult content, depressive quotes, child safety, drugs, weapons and eating disorders. We have also expanded our adult content model to proactively remove adult content from our platform. Since the rollout of this updated model, we have seen a reduction in user reports for adult content. We plan to continue our investments in this area to fight policy-violating content on Pinterest.

### 3. Third-party experts

Pinterest consults with external third-party experts to identify new and emerging trends in harmful content on our platform. We mentioned last year that we had plans to expand our work with these experts and since then, we have made additional investments with existing third-party experts who provide us with information on possible policy-violating content. In addition to misinformation, elections and child safety, we have expanded the coverage of our external third-party experts to content areas including self-harm, illegal drugs and adult sexual services.

### 4. Quality assurance

In our 2023 risk assessment, we identified that we needed to enhance our quality assurance (“QA”) programme to ensure our enforcement decisions are consistent, accurate and in line with our content moderation policies.

Since last year, we have enhanced our process to calibrate enforcement accuracy between our content moderation outsourced partners, Pinterest’s content safety enforcement subject-matter experts, and the Pinterest Content Policy team. To this end, we introduced a formal Root Cause Analysis and Error Trend Analysis to identify and mitigate the reasons for incorrect enforcement decisions. In addition, we adjusted both the volume and the mix of different content actions in the quality sample to better represent the total volume of items reviewed. Our QA programme remains a priority for our Trust & Safety team and we continue to make investments in our QA team to enhance our current programme.

### 5. Supporting mental health and emotional well-being

Pinterest is committed to taking meaningful actions in the service of supporting more positive mental and emotional well-being for all people, especially young people. Since last year, we’ve continued our work to help make the internet a more positive place with a focus on public health, minors, and the physical and mental well-being of our users. We recently completed an external listening tour with mental health experts and a youth council on teen safety. We have also partnered with a research firm to run co-creation sessions with teens to better centre their lived experience and create a digital space that supports who they are and what they envision for themselves. We augmented our existing partnership with the Mental Health Foundation to continue to address suicide and self-harm content on the platform. We’re also taking steps to make Pinterest a more inclusive platform with the introduction of our body type technology and the launch of our body type ranges in March 2024. In addition, 70% of our year-to-date strategic giving budget has been awarded to organisations or projects focused specifically on supporting the mental and emotional well-being of young people. This is an area that Pinterest is committed to and we’ll continue our efforts in this space as we see the value and impact of our efforts.

Details on our mental health and emotional well-being efforts can be found in our latest [ESG Report](#). We include details on our existing, in-product mental health and well-being features, such as [compassionate search](#) and [search advisories](#) later on in this report.

---

## 3. Risk Assessment Methodology

### 3.1 Understanding Pinterest's systemic risk landscape

Pinterest considers a systemic risk to be a risk that a platform could be designed, functioning or used (or misused) in a way that could cause serious harm or have serious negative consequences for the platform's users in the EU.

Although there have been changes to the external environment and our product that will impact the 2024 DSA risk assessment, many of the controls we describe are similar to those described last year. We have included additional information where we have made improvements and updates.

#### Systemic Risk Categories

Similar to last year, in order to identify the risks that Pinterest could present to users in the EU, we reviewed and updated our risk register. We started this process by reviewing the risk register from 2023 to make sure that those risks remained relevant. We made updates to these existing risks where required. In addition, we looked at internal and external factors that could have an impact on our risk register, such as changes to the Pinterest product or new functionalities, external events such as global elections and the prominence of GenAI, and how these factors impact how harm can manifest on the platform.

We assessed each systemic risk individually and aggregated them into four risk categories:

1. Illegal content;
2. Negative effects on the exercise of fundamental rights;
3. Negative effects on civic discourse, electoral processes and public security; and
4. Negative effects in relation to gender-based violence, the protection of public health and minors, and serious negative consequences to the person's physical and mental well-being.

Given the volume of content on the Pinterest platform, detecting and taking appropriate action on harmful content cannot completely eliminate these four risk areas. There are inherent risks on any platform that deals with vast quantities of content and data. Pinterest takes a risk-based approach when it comes to content moderation, prioritising harms that pose the greatest potential risk and balancing fundamental rights with keeping users safe.

#### Influencing Factors

We've also considered if and how Pinterest's design, functionality or use influence these systemic risks. We've focused on the following influencing factors (the "influencing factors") and have taken other potential influencing factors—such as manipulation and amplification—into account where deemed relevant:

1. Applicable terms and conditions and their enforcement;
2. Content moderation systems;
3. Design of recommender systems and any other relevant algorithmic systems;
4. Systems for selecting and presenting advertisements;
5. Related data practices; and
6. Intentional manipulation

These influencing factors encapsulate Pinterest's full platform ecosystem and we've assessed each factor as part of this risk assessment.



## 3.2 Risk Assessment Methodology

Based on our understanding of Pinterest's systemic risk landscape and our risk register as described in the [Systemic Risk Categories](#) section, we evaluated our systemic risks by analysing a variety of sources, including:

- Interviews: We conducted ~30 interviews with a wide range of internal Pinterest stakeholders.
- Questionnaires: We collected information from 14 questionnaires sent to 30 internal Pinterest stakeholders.
- Documentation: We reviewed internal documentation, such as policies, procedures and other control documentation.
- Metrics and other data: We leveraged data and other information included in our most recent Global Transparency Report (H2 2023), which contains information on our efforts to keep our platform safe and inspiring, such as the number of Pin removals and account deactivations, and our most recent DSA Transparency Reports (reporting periods 25 September 2023 to 31st December 2023), which contain information specific to EU users content.
- Input from external stakeholders: We leveraged input from our users, risk experts, experts in specific types of harm, industry groups and independent civil society organisations. For example, Pinterest attended the EU Rights & Risks Forum, hosted by the DTSP and Global Network Initiative, which brought together representatives from VLOPs and Search Engines (VLOSEs), as well as civil society and academic experts from across Europe and other jurisdictions to discuss systemic risk assessments as provided for in the DSA. Over the course of two days of panels and workshops, Forum participants explored DSA risk assessments and their potential impact on fundamental rights. The Forum was designed to inform approaches to risk assessment and stakeholder engagement, but not to constitute the totality of any company's assessment or engagement.

For each risk on our risk register, we assessed:

- Inherent risk rating: the level of risk that exists if left untreated. To determine this rating for each risk, we considered:
  - Severity: the impact that it would have on user groups and EU society in general. Each risk was assigned a severity rating of **Marginal**, **Moderate**, **Significant** or **Critical**.
  - Probability: the likelihood that the impact will occur. Each risk was assigned a probability rating of **Unlikely**, **Possible**, **Likely** or **Almost Certain**.

Based on both the severity and probability ratings, each risk was assigned an inherent risk rating of **Low**, **Medium**, **High** or **Very High**. (See the [Appendix](#) for additional information on these ratings.)

- Control effectiveness: Using our control library, we identified the controls and safeguards in place to mitigate each risk and determined how effective the control environment is in mitigating the inherent systemic risk. In order to assess the effectiveness of controls, we used data and metrics where available, as well as preliminary audit observations identified by both our Internal Audit team and our independent external auditors who carried out testing of our DSA controls. Our Compliance team was responsible for overseeing this testing. We plan to use the results of the independent audit as part of the DSA risk assessment in future years. Each control was assigned a control effectiveness rating of **Ineffective**, **Somewhat Effective**, **Effective** or **Highly Effective**.
- Residual risk: the level of risk remaining once the controls and mitigations have been considered. Based on the inherent risk rating and control effectiveness ratings, each risk was assigned a residual risk rating of **Low**, **Medium**, **High** or **Very High**.

While we assessed each risk individually, we grouped the risks into four systemic risk categories (see above) and reported on the aggregate risk ratings and control effectiveness scores.

See the [Appendix](#) for additional information on our risk assessment methodology.

---

## 4. Pinterest's platform ecosystem

At Pinterest, our mission is to bring everyone the inspiration to create a life they love, and it's our "guiding light" in drafting and enforcing our content policies. Our moderation practices are always evolving to keep up with new behaviours and trends and to create a more positive place on the internet for the people on our platform. We continue to invest heavily in measures like machine learning technology to fight policy-violating content on Pinterest and work with outside experts and organisations, including civil society, charities, user groups and academics to inform our policies and content moderation practices. We are proud of our policies and practices because they're the right things for the people on our platform. They help Pinterest to be a more positive and inspiring place online.

Each element of Pinterest's ecosystem has been developed with safety as a guiding principle. As part of this risk assessment, we've analysed the potential impact these elements — or "influencing factors" — could have on each of the systemic risk categories. Before we dive into the results of the assessment, we'll first provide an overview of each of these elements.

### 4.1 Influencing factor 1: Applicable terms and conditions and their enforcement

#### Policies and guidelines

To help us cultivate a positive and inspired community, we develop and enforce content policies that help in our aim to ensure our platform is a positive place where people can find real-life ideas as to what to try, cook, wear or do next. This includes:

- [Terms of Service](#): terms users agree to when using Pinterest
- [Business Terms of Service](#): governs business access to and use of Pinterest
- [Community Guidelines](#): what we do and don't allow on Pinterest
- [Merchant Guidelines](#): requirements for merchants operating on Pinterest
- [Advertising Guidelines](#): standards for creating ads
- [Advertising Services Agreement](#): the terms advertisers agree to when advertising on Pinterest
- [Privacy Policy](#): information we collect, how we use it and users' options
- [Copyright and Trademark policies](#): information on how we expect Pinners to respect the intellectual property rights of third parties, and how rights holders can protect their rights on Pinterest
- [Enforcement](#): how we put our policies into practice, including any restrictions that we may apply to users' content or use of Pinterest
- [Developer Guidelines](#): guidelines for building with Pinterest materials

These policies and guidelines are applicable globally, meaning that they apply to all users, businesses, merchants and advertisers who use our platform. For our policies, we've built in nuances for local legislation and regulations, and they're translated and available in 45 languages. We want to make sure that our Pinners understand our policies the first time they read them, so we use plain language along with short summaries to make our policies accessible.

All of our policies and guidelines are available on our [Policy Page](#) and are also searchable through our [Help Center](#), making them easily available to Pinners and others. Our Policy team, alongside our Legal team, is responsible for drafting and maintaining Pinterest's policies and guidelines. We continue to engage with external parties to get input and feedback on any new policies or major updates to existing policies to make sure that new and updated policies do not have a disproportionately negative effect on certain user groups. These third parties can include non-profit organisations, independent experts, civic groups and our Pinners.

We also have internal-facing enforcement guidelines for our content review teams and automated content review models to identify and action violating content. These enforcement guidelines align with our [Community Guidelines](#) and other publicly-facing policies and are designed to provide more nuance and details for our content review teams and systems to properly identify prohibited content and take appropriate enforcement action. We monitor these enforcement guidelines closely and aim to make sure that they are effective and resulting in the correct enforcement actions.

We have an agile and streamlined approach to revising and updating these policies so that we can quickly adapt to new and emerging harm types and make any adjustments that need to be made based on the results of monitoring and QA.

We explain how we enforce our content policies in Influencing factor [#2](#).

## 4.2 Influencing factor 2: Content moderation systems

Pinterest has a robust content moderation enforcement framework in place to identify and take action on harmful content on our site. The foundation of this framework is our content policies, which are comprised of both our external policies and our internal enforcement guidelines. Building on this foundation, we use different methods to enforce these policies, including reporting by users and our own identification and enforcement models, which include automated, manual and hybrid enforcement. Further details on these are set out [below](#).

We take action on harmful, violating and/or illegal content by blocking, removing or limiting the distribution of content, and deactivating the accounts of individuals and groups that create or spread harmful content and behaviour. The type of action we take is based on how much harm the violating content or behaviour poses. We may deactivate an account either after a single violation if a policy violation is severe enough or if we determine that the account has repeatedly posted policy-violating or illegal content or displayed signs of unacceptable behaviour. For Pinner's who repeatedly post content that violates our policies and guidelines, we have a strike system in place, which could result in the deactivation of a Pinner's account. Our policies and the methods we use to enforce them are always evolving to keep up with new behaviours and trends. Additional information on potential enforcement action can be found on our [enforcement page](#).

### Reporting harmful content and behaviour

Pinterest provides several ways to report content that are easily located, directly accessible and always available.

#### Reporting Policy Violations

Our reporting options allow our users to tell us if they think something on Pinterest is in violation of our policies. Pinner's can report any Pin, account, board, comment or message they believe is in violation of Pinterest's policies and guidelines. This process is available across all Pinterest surfaces (i.e., website, iOS, Android) and in all Pinterest-supported languages.

There are two ways that Pinner's can make a report to us: in-product or through our [Help Center](#) using the “[Report something on Pinterest](#)” link. If a user is logged in, they can click on the three small dots located directly on the content (whether it's a Pin in the home feed, in a close-up of a Pin, on a board, account, comment or message). There are multiple reporting options for the user to choose from, including policy-violating content and intellectual property infringements (e.g., copyright, trademark) in addition to local EU law violations for EU users, as described [below](#). Depending on the reason for the report, the user may be asked for additional details and to confirm the report. Where suitable, their report will then be directed to the appropriate review team. The report will be reviewed against our policies and for illegality when applicable. Our enforcement guidelines assist review teams to properly identify violating, prohibited or illegal content and take the appropriate enforcement action. The review teams receive specialised training — both at onboarding and ongoing training — to keep up to date with new and emerging trends in harmful and violating content as well as any changes to our policies.

If a user reports content for violation of our copyright or trademark policies, they are directed to a dedicated reporting form designed to request all the information required for the Intellectual Property Operations team to review the report. The copyright and trademark reporting forms are also available to non-users in our [Help Center](#). If a violation is confirmed, the relevant specialised operations agent will take action on the content as appropriate.

## Reports under local law

For people in the EU, we additionally provide a [dedicated reporting channel](#) where they can report content on Pinterest they believe to be illegal under EU or Member State law. This reporting channel is available both in-product and through the [Help Center](#), and is available to users and non-users alike. The form is available in the EU languages of countries in which Pinterest is available. For in-product reports, users may choose the option to “Report Pin for EU local law violation.” When choosing this option, they will be directed automatically to a dedicated reporting form asking them to provide more information about their report and why they believe the content to be illegal. When reporting via the in-product reporting flow, the reporting form will be automatically populated with the URL to the reported content. The reporting form can also be accessed from our Help Center.

Reports submitted via this designated channel are reviewed by agents on the Trust & Safety team who are specially trained to review certain policy violation reports and/or reports under local law. The content is reviewed, and if the content is found to violate our policies, action is taken on the content globally. If content is determined to violate local law but not removed for violation of our policies, the content is blocked in the relevant jurisdiction or region where it is considered unlawful.

We also have dedicated channels through which government authorities may request the removal of content they consider to be unlawful or in violation of our policies. These reports are submitted via a designated email alias ([abuse@pinterest.com](mailto:abuse@pinterest.com)) and are reviewed by specialists on the Law Enforcement & Government Operations team. Where content is found to be violative of our policies, appropriate action is taken globally; if not otherwise policy-violating, content that is confirmed to be locally unlawful is blocked in the relevant jurisdiction. The Law Enforcement & Government Operations team is also trained and tasked with responding to legal and emergency requests from law enforcement for user information. This team aims to protect Pinners and their data by balancing our legal obligations with user privacy.

We have several controls in place to ensure the integrity of our reporting channels and we use these controls to identify and restrict or prevent the processing of reports from people who abuse our reporting channels. For example, to prevent abuse, we may limit the number of reports that one person can submit in a specific time period. In accepting our [Terms of Service](#), our users agree to submit reports in good faith and not misuse any reporting or appeals channel by making baseless reports or appeals.

## Enforcement Systems

In addition to our reporting tools and channels, we enforce our policies through various enforcement models, which include automated tools, manual review and a hybrid approach that incorporates elements of both. These systems may use machine learning as well as logic-based rules. Where appropriate, we may take into account information provided by trusted third parties and industry tools.

### Automated Enforcement Tools

Our automated tools use a combination of signals to identify and take action against potentially violating content. For example, our machine learning models assign scores to content added to our platform. Our automated tools can then use those scores to perform appropriate enforcement actions. We use our automated tools for many specific types of harmful and policy-violating content: adult content, child safety (including sexualization of minors), civic and electoral integrity, graphic violence, hate speech, illegal drugs, medical misinformation, self-harm, and spam.

We are constantly working to improve our automated models. We use the latest modelling techniques and continuously iterate on these models. By adding new data and exploring new technical breakthroughs, we maintain or improve their performance over time to effectively address violative content.

To control the quality of a machine learning model, we evaluate the performance of that model through offline analysis and online experiments before launching. We look at a number of metrics to review the effectiveness of our models, both prior to and subsequent to launch of the model. We regularly review data and other relevant information to continually identify areas for improvement that our teams investigate and address as appropriate.

Where and when our automated tools detect policy-violating content, it will remove, limit the distribution or send the content for manual review. To balance the fundamental rights of users, we have a process for Pinners to appeal if they think that a content restriction taken for a policy violation has been in error.

## Manual Enforcement Actions

We manually act on some Pins and boards through our human review process. Pins actioned through this process may include those identified internally, those reported to us by third parties and those reported to us by users and non-users.

Pinterest uses manual review teams to receive and take appropriate action on reports of harmful and illegal content and behaviour on our platform. We have in-house moderators and we work with global external partners that support scalability in different languages and time zones. Agents go through a robust training programme at onboarding, followed by regular ongoing and ad hoc training. The agents also have access to PinU, our self-guided training portal. When there are updates to policies or guidelines, additional training is provided to agents.

The team consists of specialists trained in Pinterest's policies. In the event of a complex report, a team member can escalate questions to Pinterest team leads. These team leads may in turn consult further with specialists at Pinterest and designated review partners that have expertise in the laws of the EU and its member states.

Though Pinterest is primarily a visual platform, all review agents speak at least one EU Member state language. We also have external resources available to assist with any languages not spoken by agents.

Our external partners are responsible for following quality processes implemented by Pinterest's QA programme. This programme focuses on verifying the accuracy and consistency of enforcement decisions. Since last year, we've made some enhancements to our QA programme, with our goal being to further enhance this programme to make sure our enforcement decisions and actions are consistent, accurate and in line with our policies and enforcement guidelines. Further details on these enhancements can be found [here](#). We monitor the results of our QA programme and we may implement additional controls, enhance our training programme or make changes to our enforcement guidelines based on the results.

## Hybrid Enforcement Actions

Hybrid actions include those where a team member determines that a Pin violates policy and automated systems help expand that decision to enforce against machine-identified matching Pins. Depending on the volume of matching Pins, a hybrid action may result in a number of Pins actioned or none at all.

## Other enforcement mechanisms

### *Third-party experts*

Pinterest engages with external third-party experts to provide additional content moderation support. These third parties are experts in specific harm types, they keep us informed of industry trends and help us detect whether those trends are occurring on Pinterest. The new and emerging trends and other signals we receive from these experts are reviewed and where appropriate, we build these trends and signals into our content moderation tools. We may also carry out targeted enforcement efforts based on signals provided by these third parties and update our moderation tools based on the results of those enforcement efforts.

Since last year's risk assessment, we've made additional investments with our existing third-party experts to expand the scope of their coverage and entered into additional partnerships with new third-party vendors to include additional types of harm. Further details on these new measures are [here](#).

### *Managed list of sensitive terms*

We also maintain a list of sensitive terms and phrases (Sensitive Terms List or "STL"), which is used to block search results or prevent content from appearing in recommendations where it may violate our policies, such as terms associated with child safety, self-harm, suicide, drug abuse and eating disorders. In response to searches containing certain terms, where appropriate, we display an advisory that connects users with resources if they or someone they know are in crisis or provides useful and authoritative information about an important topic. Our STL is continually expanding as we identify online trends, both internally and with the support of third-party experts. We're also expanding our work with external expert third parties so that our users have access to higher quality resources if they are in crisis. Further information on this can be found [here](#).

## Notification and appeals

### Notification

The Reports and Violations Center (“RVC”) is the central place for Pinners to see updates on content that they have reported as well as content restrictions on the Pinner’s account based on our policies or local law.

Pinners will receive a daily email alerting them to new violations or updates to content they have reported. The email directs the Pinner to the RVC, which provides additional information and, where applicable, a detailed Statement of Reason for content restrictions.

If a Pinner believes that we’ve made the wrong decision on a report or a content restriction, they can submit an appeal within six months of being notified of our decision.

Additional notifications may be sent via email rather than through the RVC. For example, people who report content they believe to be locally unlawful through our designated reporting channel first receive an acknowledgement of our receipt of their report via email, and then receive another email with a more detailed response outlining the results of our review of the report. Certain other content restriction decisions for which users receive Statements of Reasons are currently also sent via email, such as notifications of the outcome of intellectual property reports (copyright infringement, trademark infringement) and notifications of content restrictions based on our [Advertising Guidelines](#) and [Merchant Guidelines](#).

### Decision appeals process

Appeals are how users can tell us if they think we made an enforcement error. Appeals can be submitted via the [Help Center](#), by clicking the one-click appeal link in an enforcement notice email that was sent to the user, or, where applicable, directly through the RVC for content restriction decisions or decisions on reports they’ve submitted. We review appeals requests and update our enforcement decision if we determine that we made a mistake. Appeals availability may vary for some product features or in some localities; in addition, some Pinners may have additional appeal options or mechanisms under their local law.

Similar to reports, we may limit appeals; for example, we may suspend the processing of appeals from people who frequently submit unfounded or abusive appeals, and we may limit the number of times that a particular decision can be appealed. We may also use automation to handle appeals more efficiently, for example, by expanding a decision made on one Pin to other machine-identified similar Pins. We also inform Pinners in the EU of their opportunity to seek further review by a certified out-of-court dispute settlement body.

## Content moderation system integrity

Pinterest has processes in place to ensure that any new functionalities or products that are developed are assessed for any potential risk and impact on user safety. For any new product or functionality that is being launched, we have a Launch Readiness programme to ensure that our Trust & Safety team can assess the potential impact on user safety and Pinterest’s content moderation tools and recommend changes where necessary. As required by the DSA, for any new product or functionality that is being launched in the EU, we also have a process where the Compliance and Legal teams work with cross-functional internal partners to conduct a risk assessment on the functionality if it is likely to have a critical impact on systemic risk in the EU.

Pinterest also has processes in place to regularly review and update our [Community Guidelines](#) and other content policies to ensure that new and emerging types of harm are considered by our overall content moderation system. Our newly-established Risk Intelligence Engineering team is focused on proactive identification of Trust and Safety risks on our platform. This team monitors internal and external data sources and translates data anomalies into actionable insights that inform our policies, guidelines and machine learning models. When updates are made to our policies and guidelines, our policy, operations and engineering teams work in tandem to disseminate these changes throughout the content moderation ecosystem. This can include delivering new training and updating training decks, enforcement guidelines and automated models.



## Transparency reporting

Pinterest publishes biannual [global Transparency reports](#), which outline the actions we take to uphold our [Community Guidelines](#). The transparency reports contain information on our content moderation efforts globally as well as insights into the volume of information and removal requests we receive from law enforcement and government entities. Only the data required for each transparency report is gathered — all data is anonymous and no sensitive data is used for the purposes of these reports.

In addition, pursuant to the DSA's requirements, Pinterest publishes DSA transparency reports at least every six months that provide information on our content moderation activities in the EU. Similar to our global transparency reports, all data is anonymous and no sensitive data is published.

### 4.3 Influencing factor 3: Design of recommender systems and any other relevant algorithmic systems

Rather than optimising for view time, Pinterest is focused on creating positive outcomes for users by helping them go from online inspiration to offline reality. Pinterest's users are looking to plan what's next, and our goal is to help them take action. People use Pinterest to visualise their future, from everyday decisions like what to make for dinner, to celebrating special moments like birthday parties and holidays, to planning for life milestones like a new baby. This is reflected in Pinterest's most common use cases, which include food and drink, fashion, home decor, crafting, and do-it-yourself projects.

Pinterest's recommender systems are designed to prioritise high-quality and inspirational content. We do this by tuning our recommender systems to prioritise explicit user signals — not just views alone — for example, “saves,” in which a user saves a Pin to a board. We don't prioritise unlimited views and time spent on our platform. Saved Pins are more likely to be inspirational and useful content that the user intends to return to later, such as step-by-step guides, self-care ideas or inspirational quotes. Because of the choices we make in designing our Recommender Systems, Pinterest's subsequent recommendations to the user will be tuned to this type of content.

Our Pinner have the ability to fine-tune the recommendations they receive. Pinner can tune their home feed to customise their preferences. For example, users can limit specific topics of interest when they no longer wish to see those recommendations. Pinner can also hide individual Pins from their home feed or unfollow the board, topic or account that the Pin came from. In addition, Pinner in the EU have the ability to opt out of personalised organic recommendations that use inferred signals.

As well as surfacing positive content, we also tune our machine learning models to de-emphasise low quality content. In addition, Pinterest utilises a workflow that captures Pins, boards and users that have been deactivated due to content safety policy violations and excludes them from the data used by recommendation models to help prevent these systems from recommending similarly unsafe content. Pinterest also uses a machine learning model that is designed to identify content that — while not policy-violating — may be harmful, such as depressive quotes or racy content. We then filter or downrank this content from recommendation surfaces. If a user does click on a Pin that has been identified by this model, related Pins are not displayed beneath the Pin.

Pinterest utilises non-engagement signals, such as in-app surveys (where users can tell us about their experience on the platform) or independent assessments of content quality, which are usually generated by manual labelling. These non-engagement signals provide a balance to engagement signals in our content ranking and also help us put our values into action. For example, our industry-leading inclusive products have relied heavily on non-engagement signals, as they provide recommendations based on direct user feedback on the body type, hair pattern or skin tone they want to prioritise in their feed. Pinterest is passionate about this work, publishing a “field guide” earlier this year to assist other companies interested in adopting this technology.

## 4.4 Influencing factor 4: Systems for selecting and presenting advertisements

Our [Advertising Guidelines](#) help our advertisers promote inspiring content on Pinterest in line with our mission, and all advertisers are subject to these guidelines. These guidelines include information on the categories of ads that are prohibited and restricted. They also contain country-specific guidelines that must be followed when targeting users in those countries.

We want Pinterest to be a positive and inspiring place for everyone, so there are some areas where our Advertising Guidelines go further than what is prohibited by law. For example, we don't allow political campaign advertising, ads that body shame or ads containing adult content.

Pinterest is transparent when it comes to advertising, with ads clearly labelled as 'Sponsored' or 'Promoted By' to distinguish them from organic content. Pinterest also allows Pinners to see why they are being shown an ad via the Why Am I Seeing this Ad ("WAISTA") feature. WAISTA provides Pinners with information about who is presenting an ad and the main parameters used to determine why they were shown an ad. Pinterest's Ads Repository makes all ads served in the EU in the last year publicly available and provides additional information about an ad, such as how it was targeted to audiences. This allows for additional transparency on the ads being served on the Pinterest platform.

Pinterest uses a mixture of manual review and other controls to enforce our [Advertising Guidelines](#). Pinterest takes a risk-based approach to reviewing ads. A portion of ads are reviewed prior to being served on our platform. Most reviews are manual and ads that violate our Advertising Guidelines will be rejected. In addition to the manual review of ads, we use tooling to auto-review duplicate ads.

Unlike standard ads that can be content that drives users to any type of landing page an advertiser wants to promote, shopping ads drive directly to a shopping experience. Shopping ads are derived from a product catalogue that allows users to directly purchase a product. Due to the volume and relative risk level of these ads, they are not manually reviewed prior to being served on the platform. However, additional controls are in place to prevent these ads from containing illegal or harmful content. This includes user reports, which will trigger a manual review of an ad. Users repeatedly hiding an ad will also trigger a review. Pinterest also deploys machine learning models to detect certain categories of prohibited products that appear in ads. Once identified, these will be manually reviewed and taken down if they are confirmed to violate Pinterest's [Advertising Guidelines](#).

## 4.5 Influencing factor 5: Data practices

Pinterest's [Privacy Policy](#) governs our data practices. We gather data from our Pinners so that we can show them personalised content and ads we think they'll be interested in. We will only use that information where we have a proper legal basis for doing so.

When our Pinners sign up for and use Pinterest, they choose to share information with us, such as their name, birthdate, email address, photos, Pins, comments and other information they choose to share.

We collect technical information from our Pinners when they use Pinterest including log data, cookie data, device information, and clickstream data and inferences. We also get information about our Pinners and their activities outside of Pinterest from our affiliates, advertisers, partners and other third parties we work with. We use all of the data that we collect from our Pinners to show them content that is relevant, interesting, inspirational and personal to them. Additionally, we use this information to keep Pinterest and our Pinners safe.

Our Pinners have choices about how we use their information. Pinners can edit information in their profile at any time, link or unlink their Pinterest account from other services, choose whether Pinterest uses information from their engagement with advertisers to personalise the ads they see and close their account at any time. Pinners can adjust their cookie settings and choose how and whether their photos and other data is shared with Pinterest.

Details of the data we collect and the way we use it are set out in our [Privacy Policy](#), which is easily accessible. Logged-in users of Pinterest can also adjust their privacy settings according to their preferences in their privacy and data settings. This allows users to manage the data Pinterest shares with external parties, update their cookie preferences, adapt their personalisation settings, request a copy of the information Pinterest collects about them and delete their account or information, consistent with applicable laws.



Pinterest has a number of different systems in place to both process and store Pinner data. Some of these systems are proprietary to Pinterest and others are provided by third parties. Where third-party systems are in use, Pinterest has controls in place to prevent Pinner's personal data from being accessed by these third-party providers. Pinterest also has a number of controls in place to prevent external data breaches, including a bug bounty programme, penetration exercises and open source scanning. Internally, Pinterest has an Acceptable Use Policy that governs the ways in which employees and contractors can access Pinner data, including limiting access as narrowly as possible to those with a legitimate business need.

## 4.6 Influencing Factor 6: Intentional Manipulation

Our [Community Guidelines](#) govern our practices for dealing with intentional manipulation and outline that we do not permit spam or engaging in spammy behaviour, fake accounts or fabricated or meaningfully manipulated visual or audio content that erodes trust or causes harm.

Unfortunately, bad actors may seek to manipulate the Pinterest platform, including using spam attacks, bad actors using fake accounts or by distributing false or misleading content. Account Takeovers ("ATOs") can also occur on the Pinterest platform, which is where attackers gain access to existing accounts (for example, through stolen login credentials). Rather than creating fake accounts to spam users, attackers can take over existing accounts.

Pinterest utilises machine learning technology and has built automated models that swiftly detect and act against spam. These models are iterated on a regular basis by adding new data and exploring new technical breakthroughs to either maintain or improve their performance over time to effectively address spam. Logic-based rules and machine learning models are used to detect potential manipulation by analysing patterns in real-time, daily and weekly intervals. When these accounts are identified, they are deactivated. Users have the ability to appeal these decisions. Users can also report content and profiles for suspicions of spam.

Pinterest's most recent DSA Transparency Report (reporting period 25th September 2023 to 31st December 2023) shows that we deactivated over 500,000 accounts for engaging in violative behaviours, such as distributing spam. We received just over 16,000 appeals as a result of these deactivations, demonstrating the effectiveness of our controls in identifying and actioning spam.

If we think a Pinner's account has been taken over by a spammer (an ATO), we take measures to protect the account. This means that we log the Pinner out of the existing session, invalidate their password and send them an email asking them to reset their password. Once they reset their password and log back in, they will regain access to their account.

For intentional manipulation in the form of fabricated or meaningfully manipulated content that is harmful, our content moderation tools are equipped to identify and action all types of harmful content regardless of how the harmful content is created. Full details on our content moderation systems can be found under [influencing factor 2](#).

# 5. Systemic Risk Landscape

In order to understand the [systemic risks](#) that Pinterest might pose to EU users and society, we identified and assessed a wide range of individual risks associated with each of the systemic risk categories. We have reported on these risks at a rolled-up, aggregated level across four categories of systematic risk.

This section provides a summary of each of these systemic risks, the controls we have in place to mitigate these risks and what we plan to do to further mitigate these risks. In assessing each of the individual risks, we have analysed the:

- 1. Inherent risk: the level of risk that exists if left untreated. We have considered the severity of the risk and the probability of the impact occurring;
- 2. Controls and other mitigation approaches currently in place to address each risk; and
- 3. Residual risk: the level of risk leftover once the controls and mitigations have been considered.

We have also considered if and how Pinterest’s design, functionality or use influences these systemic risks. These [influencing factors](#) encapsulate Pinterest’s full platform ecosystem and we have assessed each factor as part of this risk assessment.

## 5.1 Systemic Risk 1: Illegal Content

### Summary



### Risk Overview

We have not made changes to the definition of “illegal content” that we used in our 2023 systemic risk assessment; we still use the DSA’s definition, which is “any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law.”

This is a broad concept and can potentially manifest in multiple ways on the Pinterest platform. To assess the risk of illegal content, we have looked at the categories of our [Community Guidelines](#) that most align to the concept of illegal content and are most likely to include potentially illegal content. When looking at the risk of illegal content appearing on the Pinterest platform, we have focused on the following types of policy violations: adult content, adult sexual services, child safety, dangerous goods and activities, graphic violence and threats, harassment and criticism, hateful activities, and violent actors. We have also expanded the types of harm within these categories to capture new and emerging harm types with a focus on child safety and adult content.

While these policies are not limited to illegal content (i.e., they often will be broader and stricter than what may be permitted under local law), they may be seen as signals that indicate the potential risk of illegal content on Pinterest. For these categories of policy violations, we have considered these risks individually, assessing the likelihood that this content appears on the Pinterest platform, the volume of this content and the differing severity levels that each type of content could cause for users. In this assessment, we have reported on the risk of illegal content at an overall level, however, we acknowledge that within this broad category, risk levels differ.

Inherent risk rating

Inherent Risk - Probability

Similar to last year, we have assessed the probability of the risk of illegal content stemming from the design, functioning or use of Pinterest in the EU as **Possible**. We know from our detection mechanisms such as user reports, that it is possible for users to be exposed to policy-violating or illegal content on Pinterest, however briefly, but the volume and spread of such content is relatively low. We also examined how the way this risk presents itself on the platform might have changed since our last risk assessment, and expanded the types of harm within this category to capture new and emerging harm types.

For example, as described in our most recently published DSA Transparency Report (reporting period 25th September 2023 to 31st December 2023), we removed the following numbers of user-reported violations of a particular content policy in the EU:

Content Policy	Number of EU user reports that resulted in removal of Pin(s)
Adult Content	7,411
Adult Sexual Services	3
Child Safety	1,233
Dangerous Goods & Activities	231
Graphic Violence & Threats	728
Harassment & Criticism	161
Hateful Activities	960
Violent Actors	149

Although these policies are not limited to illegal content and these metrics do not necessarily reflect content determined to be illegal, these metrics do indicate that the number of users exposed to illegal content are low.

In addition to reports of violations of particular content policy, we also receive reports of illegal content through our [European Union Illegal Content Reporting Form](#). Between 25th September 2023 and 31st December 2023, as a result of reports received through our illegal content reporting form alleging that certain content was illegal according to the laws of the EU and/or its member states, we removed 247 Pins, boards, accounts and comments for policy violations and 30 Pins, boards, accounts and comments were blocked within the relevant territory for illegal content. All of these reports were reviewed manually. These metrics indicate a very low number of Pins with potentially illegal content in the EU and therefore a low probability of the likelihood of exposure to our EU users.

We also receive requests from government entities to remove content on Pinterest that may be illegal in their country and/or a violation of our [Community Guidelines](#). We review the requests to determine if the content identified violates our Community Guidelines or local law. Our teams take action on violations, ranging from removing the content globally to blocking the content within the relevant country if it appears to violate local law but does not violate our policies. During the reporting period 25th September 2023 to 31st December 2023, we did not receive any government removal orders from member states of the EU.

## Inherent Risk - Severity

Similar to last year, we have assessed the severity of the inherent risk of illegal content if left unchecked stemming from the design, functioning or use of Pinterest in the EU as **Significant** primarily because even limited exposure to illegal content for a small number of users could lead to harm or consequences.

However, the rapid and widespread dissemination of content or content “going viral” is not common on Pinterest. This is critical to understanding the severity of this risk for EU users and society. In order to understand how uncommon this is on Pinterest, we use the concept of “reach,” which is one of our key indicators of user experience. To calculate this metric, we start by looking at each policy-violating Pin removed in a reporting period. Then, we count the number of unique users that saw each of those Pins during the reporting period for at least one second before it was removed. For example, we can look at our most recent Global Transparency Report (covering the period July 2023 - December 2023), and of the Pins removed for violating our Graphic Violence & Threats policy during the period, 90% were seen by 10 or fewer users during the reporting period and 91% of Pins removed for violating our Hateful Activities Policy were seen by 10 or fewer users during the reporting period.

Based on these metrics across our policies that aim to combat illegal content, rapid dissemination of this content is not common on Pinterest due to the nature of the platform.

## Overall inherent risk rating

Based on the probability and severity ratings, we have assessed the inherent risk of illegal content stemming from the design, functioning or use of Pinterest posing harm to users and EU society as **Medium**. There has been no change to the inherent risk rating since the 2023 risk assessment.

## Controls and mitigation efforts

Pinterest’s first line of defence for mitigating the dissemination of illegal content is our Terms of Service, which states that users will not post “User Content” — anything that a user posts or otherwise make available on Pinterest — that infringes the intellectual property rights of others (e.g., copyright infringement, trademark infringement or counterfeit), or that is otherwise unlawful. Consistent with those Terms, our [Community Guidelines](#) outline the types of content and behaviour prohibited from the Pinterest platform. There have been no changes to this part of our Terms of Service or our Community Guidelines since the 2023 risk assessment.

Similarly, our [Advertising Guidelines](#) outline the types of content prohibited in ads on Pinterest. Whilst these guidelines are global, they have been crafted to reflect certain types of content considered illegal for advertising in certain countries.

In addition to these external guidelines, Pinterest has detailed internal enforcement guidelines that delineate how to take appropriate action on violating content on Pinterest, including removing, limiting distribution of and identifying permitted content. These policies and guidelines drive our overall content moderation approach, including how automated models are used, the types of content that users can report, our manual review process and our enforcement approach. See Pinterest’s [platform ecosystem](#) for more detail on these mechanisms.

## Control Effectiveness Rating - Overall

In assessing the effectiveness of these controls we have taken into consideration the controls’ design and implementation and also the enhancements made to our machine learning models (further information [here](#)). In addition, we’ve taken into account various other factors, including the low number of reports for policy violations of categories of our [Community Guidelines](#) that most align to the concept of illegal content. For example, our most recent DSA Transparency Report (reporting period 25th September 2023 to 31st December 2023) shows that we proactively removed over 16,000 Pins for violations of our Dangerous Goods & Activities policy and removed 231 Pins as a result of user reports. Similar to last year, we have assessed our control effectiveness rating as **Effective**.

## How influencing factors affect this risk

In addition to considering the inherent risk rating and the controls we have in place, we considered how each of the influencing factors could affect the systemic risk of the dissemination of illegal content.

### Applicable terms and conditions and their enforcement

In drafting our policies and guidelines, we have worked to strike the balance between ensuring that our policies are global, easy to understand and being broad enough to cover a wide range of harmful content and behaviour. This allows our users to easily understand what is and isn't permitted on Pinterest, and it means that our policies are adaptable as new trends and types of harm emerge. While some policies such as our [Advertising Guidelines](#), contain country-specific restrictions, we address country-specific definitions of illegal content on a case-by-case basis when government authorities, Pinners or other third parties report content that they believe may be illegal in their country.

### Content moderation systems

Pinterest's content moderation systems are driven by our policies and guidelines. We have multiple mechanisms in place to detect and enforce our policies against policy-violating and illegal content. We use the latest modelling techniques and continually iterate on these models to make sure that we are keeping up with new and emerging types of harm. Our automated models are used for specific types of policy violations and we continue to leverage this technology to further expand coverage.

In addition to our standard user reporting process, Pinners and non-Pinners in the EU can report content for suspected illegality. Pinterest reviews these reports and removes globally or blocks access to content in the country or countries where it is illegal.

A key element of our content moderation system is our human review process, which has an increasingly robust QA programme to ensure that the decisions made by review agents are consistent, accurate and in line with our content moderation policies.

### Design of recommender systems and any other relevant algorithmic systems

Pinterest's recommender systems are designed to show our users content we think will be relevant, interesting, inspirational and personal, based on explicit and implicit signals that we receive from users. If a user actively searches for illegal content, there is the possibility that our recommender systems will work to show more of this content to users.

We have several controls in place to prevent this from occurring, and our recommender systems serve as a key tool in preventing the spread of this type of low-quality content. Our overall content moderation system seeks to identify and remove policy-violating and potentially illegal content from the platform, both proactively and in response to user and other third-party reports. Pinterest utilises a workflow which captures Pins, boards and users that have been deactivated due to content safety policy violations and excludes them from the data used by recommendation models to help prevent these systems from recommending similarly unsafe content.

In addition, our managed list of sensitive terms prevents a user's search from returning any results for certain terms that we consider likely to be policy-violating or otherwise harmful.

Even with these controls, there is still the possibility that illegal content could be recommended to users. This is particularly the case if users search for content that in and of itself is not illegal, but they seek to use this content for inappropriate or illegal means.

Given the risk that illegal content can pose to users and EU society, we work hard to continuously improve our efforts to ensure that our recommender systems don't contribute to the dissemination of illegal content.

## Systems for selecting and presenting advertisements

Pinterest uses its [Advertising Guidelines](#) to let advertisers know what they can and cannot advertise on our Platform. All advertisers must agree to adhere to these guidelines as a condition of advertising on Pinterest. These guidelines contain information on categories that may align to otherwise illegal content, including adult content, counterfeit goods, endangered species and live animals, illegal drugs, and illegal products and services. As well as global requirements, the guidelines list country-specific guidance that prohibit certain types of ads from being targeted to certain regions.

Pinterest has multiple controls in place to enforce these guidelines, including automated models to detect prohibited content and users reporting policy-violating ads. In addition, a portion of ads undergo a proactive review.

## Data related practices

Similar to last year, we do not consider our data practices as specifically impacting the dissemination of illegal content.

## Intentional manipulation

We don't typically see spam attacks, fake accounts or ATOs specifically aimed at disseminating illegal content, however we have controls in place to swiftly detect and act against this kind of intentional manipulation. We do not consider the risk of intentional manipulation to specifically impact illegal content.

We've also considered the intentional manipulation of content through the use of GenAI or other manipulation efforts. Our content policies and enforcement guidelines contain details on manipulated content and our content moderation tools are designed to identify and action harmful and policy-violating content regardless of how this content is created or manipulated.

## Residual risk rating

Based on the inherent risk rating and the effectiveness of the controls we have in place, we have assessed the residual risk of the dissemination of illegal content posing harm to users and EU society as **Low**. There is no change to the residual risk rating since the 2023 risk assessment.

## Further mitigation efforts

Though we have assessed this risk as **Low**, we are always working to enhance our controls because of the potential harm that illegal content could cause to our users. We're working on the following enhancements:

- 1. Agent Training Team:** Pinterest plans to make additional investments in the area of agent training by increasing headcount for our Agent Training team and further enhancing our current agent training programme.
- 2. Quality Assurance Team:** Pinterest plans to make additional investments in our QA programme. We will be increasing headcount in this programme and making further enhancements to our QA programme.
- 3. Risk Intelligence Team:** Pinterest's Risk Intelligence Engineering team plans to carry out enhanced monitoring of data sources, such as user reports across all policy areas to monitor and investigate significant fluctuations and identify internal gaps and emerging trends.

## 5.2 Systemic Risk 2: Negative effects on the exercise of fundamental rights

### Summary



### Risk Overview

We have used the Charter of Fundamental Rights of the EU as our guide in assessing the risk that Pinterest could have actual or foreseeable negative effects on the exercise of fundamental rights.

Pinterest is primarily a visual platform and for that reason we have focused on the fundamental rights that are most relevant. These include, but are not limited to, freedom of expression and information, the right to non-discrimination, media freedom and pluralism, respect for private and family life, protection of personal data, human dignity, the rights of the child, the right to protection of property including intellectual property, and consumer protection.

We have considered these fundamental rights individually, however, the risk ratings below reflect our assessment of the risks’ impact on these fundamental rights in aggregate. In assessing these risks, we kept in mind the balance that must be struck between fundamental rights, other competing requirements of the DSA regarding content and user safety, and Pinterest’s mission and responsibility to inspire and protect our users.

### Inherent risk rating

#### Inherent Risk - Probability

Similar to last year, we have assessed the probability of the risk that the design, functioning or use of Pinterest negatively impacts the fundamental rights of users or EU society as **Possible**. We know from our detection mechanisms such as user reports, that it is possible for users to be exposed to content or behaviour on the Pinterest platform which might negatively impact the exercise of fundamental rights in the EU. The increased number of global elections was factored in when assessing the probability of this risk, however, we found that though we increased the probability scores for the relevant risk areas, it did not raise the probability score in aggregate. Fundamental rights are threaded throughout our [Community Guidelines](#), like in our policy preventing hate speech on the platform. Our content moderation approach seeks to balance the fundamental rights of users — for example, freedom of expression — with preventing harmful content and behaviour from appearing on the platform.

#### Inherent Risk - Severity

Similar to last year, we have assessed the severity of the inherent risk that the design, functioning or use of Pinterest negatively impacts the fundamental rights of users or EU society as **Significant** in the absence of controls. There has been no change to the severity rating since last year’s risk assessment. Negative impacts to fundamental rights could lead to harm or consequences for users and EU society. For example, it could lead to the violation of a user’s personal data or the suppression of a user’s right to express themselves freely.



## Overall inherent risk rating

Based on the probability and severity ratings, we have assessed the inherent risk that the design, functioning or use of Pinterest negatively impacts the fundamental rights of users or EU society as **Medium**. The overall inherent risk rating has not changed since last year.

## Controls and mitigation efforts

Our overall content moderation systems and controls work together to make Pinterest an inspirational and positive place on the internet that also protects the fundamental rights of users and other members of EU society.

### Freedom of expression and information

Pinterest users are free to express themselves and inspire others on our platform within the bounds of our [Community Guidelines](#). Our Community Guidelines outline the type of behaviour we do and don't allow on Pinterest and are always grounded in user safety. While some users may disagree with policy stances we have taken, for Pinterest, user safety is critical to our business. Nevertheless, we have built nuance into our moderation systems to ensure that content can be reviewed with context and understanding, and users have the ability to appeal decisions if they disagree with our enforcement decisions.

Also, at times, we receive requests from EU government agencies to remove content on Pinterest that may be illegal in their country. The fundamental rights of our users are paramount — we diligently review these requests and we only take action on content that we have confirmed is policy-violating or illegal.

### Non-discrimination

Pinterest is a place for inspiration, not discrimination. We have a number of policies that address discriminatory content and behaviour on our platform. This includes our Hateful Activities policy, which prohibits hateful content and the people and groups that promote hateful activities on Pinterest. Hateful activities include slurs and negative stereotypes, caricatures and generalisations, as well as support for hate groups and people promoting hateful activities.

In an effort to create belonging on Pinterest, we intentionally make the content surfaced on our platform more diverse and inclusive. Our Inclusive Product team oversees this effort and has put measures in place to make sure that fairness and bias are considered in our models and recommender systems. We also look at our models' performance to see whether they are behaving as intended. We want our systems to surface a diverse set of results — we don't want our Pinner to work hard to feel they are represented.

Our algorithms diversify our search results, related feeds and new user homefeeds by skin tone and body type, which directly increases representation. For Pinner who want to see more specific refinements, we've developed unprecedented, industry-leading inclusive features, such as hair pattern search, skin tone ranges and body type ranges. We want it to be easy for our Pinner to find content that's relevant to them.

We also want Pinterest to be inclusive for all types of sight ability. We partnered with LightHouse for the Blind and Visually Impaired to better understand how we could make Pinterest more useful for people with different levels of vision. In 2018, we made updates across our apps and website for Pinner with disabilities. For some surfaces, these updates included closed captioning, alternative text, better screen reader support and colour contrast sensitivity improvement. As a result, it's much easier for users to browse, search and save ideas on Pinterest.

### Media freedom and pluralism

We do not limit media or news organisations from joining, having accounts or creating Pins on Pinterest, except as required by law (for example, sanctioned state-controlled media organisations) and subject to our [Terms of Service](#) and [Community Guidelines](#).

In addition, we know from research that Pinterest isn't typically a platform where users come to seek news or current affairs. The top three categories that monthly users say they come to Pinterest for are Craft and DIY, Home Design and Decor, and Food and Drink.



## Protection of personal data

Pinterest highly values the protection of our users' personal data. Pinterest's [Privacy Policy](#) explains to users the personal information we collect, how we use it and the choices that users have related to this, which includes how we use data to personalise a user's experience on the platform and the information that we obtain on users from our partners and advertisers. We have [Help Center](#) articles that elaborate on our Privacy Policy. We have multiple options for users to choose how their personal data is used. Detail on these options is provided in Pinterest's [platform ecosystem](#).

At times we receive legal requests from law enforcement for Pinterest user information. We diligently review each request and only produce data for those that meet the requirements in our Law Enforcement guidelines and in accordance with our [Privacy Policy](#) and legal obligations.

## Respect for private and family life

We give users options when it comes to engaging on the platform privately. Boards and Pins can be private, shared with a limited number of other accounts or visible to the public. The profiles of users under the age of 16 are set to private as the default and only option. Users can also report content for privacy violations, for example, if a Pin contains private contact information, personal or sensitive information, or is a private photo. Moreover, users can report and/or block other users if they believe they are being harassed.

A user can close their account at any time. When you close your account, we'll deactivate it, remove your Pins and boards from Pinterest, and delete your account data (subject to our standard data retention policies and legal requirements).

## Human dignity

Our Community Guidelines outline the content and behaviour that is and isn't allowed on Pinterest. We have specific policies, including our Harassment & Criticism policy, to help people engage on Pinterest in a positive, inspirational and respectful way. We also prohibit content that insults, hurts or antagonises individuals or groups of people, including manipulated images intended to degrade or shame, shaming people for their bodies or assumed sexual or romantic history, sexual remarks about people's bodies, solicitations or offers of sexual acts, and mocking someone for experiencing sadness, grief, loss or outrage.

Pinterest isn't a place to insult, hurt or antagonise individuals or groups of people, and this type of behaviour is not tolerated. Respectful criticism is of course permitted, but we may limit the distribution of or remove insulting content that violates our policies to keep Pinterest a positive, inspiring place on the internet.

## Intellectual property

Pinterest respects intellectual property rights, and we expect our Pinners to do so as well. Our Copyright and Trademark policies set out the ways that Pinterest protects the intellectual property and fundamental rights of our users. We respond promptly to claims of copyright and trademark infringement on Pinterest. It's our policy, in appropriate circumstances and at our discretion, to disable or terminate accounts that repeatedly or seriously infringe or are repeatedly charged with infringing copyrights or other intellectual property rights.

## Consumer protection

Although Pinterest is not involved in facilitating the purchase, sale or delivery of goods, we want people to have good experiences shopping for products they find on Pinterest. Merchants are responsible for making sure they follow all relevant laws, regulations and industry codes when they use our service. Merchants are also responsible for handling and responding to all purchases, deliveries, customer service questions, complaints, problems and disputes. These requirements are set out in our [Merchant Guidelines](#).

Pinterest has a Verified Merchant programme to help shoppers discover and buy from verified brands. A verified merchant gets a badge on their profile and product Pins showing that their brand was verified by the Pinterest team. Verified merchants must adhere to specific requirements set out in our guidelines, and we also monitor shopping experience quality of verified merchants. If we detect excessive user reports, merchants may be suspended from the programme. In addition to our Merchant Guidelines, our [Advertising Guidelines](#) include information on unacceptable business practices.

## Control Effectiveness Rating - Overall

Similar to last year, and based on their design and implementation, we have assessed these controls as **Effective**. This rating is based on a number of factors, including ongoing monitoring of the controls discussed above, the accuracy and coverage of our automated models, our controls designed to protect intellectual property, and ongoing monitoring of users' reports associated with content of this nature. Our most recent DSA Transparency Report (reporting period 25th September - 31st August 2023) shows that we proactively removed over 20,000 Pins for violation of our Harassment & Criticism policy and over 16,000 Pins that violated our Hateful Activities policy. We've also made improvements to our product with the introduction of our body type ranges feature, demonstrating Pinterest's commitment to creating belonging. However, given the potential harm that this risk can pose to users and EU society, we work hard to continuously improve our efforts in these areas.

## How influencing factors affect this risk

In addition to considering the inherent risk rating and the controls we have in place, we considered how each of the influencing factors could broadly affect the exercise of fundamental rights. In the following analysis, we consider fundamental rights as a whole.

### Applicable terms and conditions and their enforcement

Our policies and guidelines provide transparency to users and allow them to decide whether Pinterest is a platform for them. By having clear policies and guidelines about the type of content and behaviour that is permitted on Pinterest, we are able to moderate content in an unbiased way. Our policies are designed to balance the fundamental rights of users with protecting the safety of our users. When we make updates to our policies and guidelines, we often engage external experts to ensure that we are not disproportionately impacting a specific group of users.

Our policies and guidelines, and the enforcement of these, impact several of the fundamental rights discussed above, including freedom of expression and information, non-discrimination and human dignity.

### Content moderation systems

Our overall content moderation approach works to detect and take appropriate action on harmful content that could impact fundamental rights, such as human dignity or non-discrimination. We have controls in place to ensure that these processes are accurate and without bias, including continually improving our detection measures, a QA programme for our human review processes, training for our review agents and an appeals process. Freedom of expression is taken into account and our [Community Guidelines](#) seek to balance the fundamental rights of users with preventing harm to our users.

### Design of recommender systems and any other relevant algorithmic systems

Pinterest's recommender systems are designed to show our users content we think will be relevant, interesting, inspirational and personal. This impacts several of the fundamental rights discussed above including freedom of expression and information and protection of personal data. We are transparent with users about how we use this information and users are given the option to "opt out" of personalised recommendations based on inferred signals, which limits the type of personal data that we use.

### Systems for selecting and presenting advertisements

Our [Advertising Guidelines](#) prohibit targeting audiences based on sensitive categories, such as race, religious beliefs or political affiliations, among other things. Pinterest's systems for presenting advertisements are designed to respect the right to protect personal data, respect for private and family life, and non-discrimination.

### Data related practices

Our [Privacy Policy](#) and internal data privacy and security policies work together to ensure that we collect, use and store personal data in an appropriate way and that we maintain the security of our users' data. Details on these controls are provided in Pinterest's [platform ecosystem](#). These practices help to protect users' fundamental rights, in particular regarding personal data and respect for private and family life.

## Intentional manipulation

We don't typically see spam attacks, fake accounts or ATOs specifically targeted towards the area of fundamental rights, however, we have controls in place to swiftly detect and act against intentional manipulation. We do not consider the risk of intentional manipulation to specifically impact the exercise of fundamental rights.

As well as intentional manipulation through spam attacks or ATOs, we've also considered the intentional manipulation of content through the use of GenAI or other manipulation efforts. Our content policies and enforcement guidelines contain details on manipulated content and our content moderation tools are designed to identify and action harmful and policy-violating content regardless of how this content is created or manipulated.

## Residual risk rating

Based on the inherent risk rating and the effectiveness of the controls we have in place, we have assessed the residual risk that the design, functioning or use of Pinterest negatively impacts the fundamental rights of EU users or society as **Low**. There has been no change to the residual risk rating since last year.

## Further Mitigation Efforts

Though we have assessed the residual risk as **Low**, we are always working to make improvements to this area. We're working on the following enhancements:

- 1. Body type diversity:** We plan to improve the diversity of different men's body types in men's fashion search results.
- 2. User Reporting:** Pinterest will be introducing a new user reporting function so that our community can report suspected underage users, which will help us in our efforts to keep our most vulnerable users safe and protected.

We're also implementing [additional controls](#) around our use of GenAI and providing our users with information about how we use GenAI. These enhancements will add to the ways we protect fundamental rights, particularly in relation to personal data and respect for private and family life.

Additional investments in our [Agent Training team](#) will also enhance our detection and enforcement against content that may negatively harm the fundamental rights of users. We are also making investments in our [Quality Assurance Team](#) to enhance our current QA programme. As well as these measures, our [Risk Intelligence Engineering Team](#) will be carrying out enhanced monitoring of data sources to identify policy grey areas and emerging trends across all content policies.

### 5.3 Systemic Risk 3: Negative effects on civic discourse, electoral processes and public security

#### Summary



#### Risk Overview

When assessing the risk that Pinterest’s design, functioning or use could lead to negative effects on civic discourse, electoral processes and public security, we considered the various ways in which this risk could manifest, including misleading information about election dates, how to correctly fill out a ballot, who is allowed to participate in an election or census, and fabricated or manipulated content (e.g., content generated with AI tools) creating the appearance of someone doing or saying something they didn’t with the clear intent of influencing public sentiment about a political figure and/or election. A change to the risk this year is the number of global elections that took place and/or are set to take place, including, amongst others, the European Parliament elections. We’ve set out below the additional measures we put in place to account for this change.

We also looked beyond elections to consider how users engage in civic discourse on the platform and whether Pinterest’s design has a negative effect upon this discourse. The assessment below reflects our analysis of these individual risks in aggregate.

#### Inherent risk rating

##### Inherent Risk - Probability

Similar to last year, we have assessed the probability of the risk that the design, functioning or use of Pinterest negatively impacts civic discourse, electoral processes or public security in the EU as **Possible**. We know from our detection mechanisms such as user reports and deactivations, that it is possible for users to be exposed to this type of content on Pinterest, however, generally, users do not come to Pinterest for election content and the number of users exposed to this type of content is low.

This is supported by data from our most recent DSA Transparency Report (reporting period 25th September to 31st December 2023) showing that we removed a total of 137 Pins in the EU related to our Civic Participation Misinformation policy.

##### Inherent Risk - Severity

Similar to last year, we have assessed the severity of the inherent risk that the design, functioning or use of Pinterest in the absence of controls negatively impacts civic discourse, electoral processes or public security in the EU as **Significant**, primarily because even limited exposure to content or behaviour of this nature for a small number of users could lead to harm or consequences.

However, people do not come to Pinterest for political content, and as an image-based visual inspiration platform, the possibility of rapid and widespread dissemination of content or “going viral” is low. This inherent severity risk rating of **Significant** is based on the risk for EU society more broadly.

Our most recent Global Transparency Report (H2 2023) shows that in Q3 2023, we removed fewer distinct images and Pins that violated our Civic Participation Misinformation policy compared with previous quarters. The decrease in Q3 2023 compared with previous quarters was the result of fewer user reports for this policy and fewer matches of violative content identified. In Q4 2023, Pin and distinct image removals returned to the baseline from Q1 and Q2 2023.

### Overall inherent risk rating

Based on the probability and severity ratings, we have assessed the inherent risk that the design, functioning or use of Pinterest negatively impacts civic discourse, electoral processes and public security in the EU as **Medium**. There has been no change to this risk rating since last year.

### Controls and mitigation efforts

As with other types of harmful content and behaviour, Pinterest’s first line of defence for mitigating negative effects on civic discourse, electoral processes and public security are our [Community Guidelines](#) and other relevant policies. Our Community Guidelines apply to all content categories and provide guardrails for appropriate civic participation on the Pinterest platform. We also have enforcement guidelines for our content review teams.

Our Civic Participation Misinformation policy prohibits misinformation, disinformation and mal-information, and the individuals or groups spreading or creating it on Pinterest. This policy prohibits false or misleading content on Pinterest that impedes an election’s integrity or an individual’s or group’s civic participation, including registering to vote, voting and being counted in a census.

Our Violent Actors policy also prohibits violent content, groups or individuals. We limit the distribution of or remove content and accounts that encourage, praise, promote or provide aid to dangerous actors or groups and their activities. This includes extremists, terrorist organisations, gangs and other criminal organisations. We work with industry, government and security experts to help us identify these groups. For example, since 2019, Pinterest has been a member of the Global Internet Forum to Counter Terrorism (“GIFCT”), a non-governmental organisation designed to prevent terrorists and violent extremists from exploiting digital platforms.

Our Conspiracy Theories policy prohibits conspiracy theories about civic participation and content that turns or encourages turning individuals, groups of people, places or organisations into targets of harassment or physical violence, such as hate-based conspiracy theories and misinformation about mass atrocities.

Civic discourse is a broad concept and we have other policies that also work together to prevent negative effects. This includes our Graphic Violence & Threats policy, which covers threats against voting locations, census or voting personnel or participants and our Hateful Activities policy, which covers intimidation of vulnerable or protected group voters or participants.

These policies and guidelines drive our overall content moderation approach, including how automated models are used, the types of content that users can report, our manual review process and our enforcement approach.

Pinterest does not allow political campaign ads and this is outlined in our [Advertising Guidelines](#). People come to Pinterest for inspiration for recipes, fashion, home decor, crafting and do-it-yourself projects — people don’t come to Pinterest for political content.

Although people don’t come to Pinterest for political content, we understand that we must be responsible when it comes to election integrity and civic engagement. We have elections-focused teams who proactively and reactively address major elections around the world and decide what level of action is required based on the risk presented by each election. In order to identify the risks associated with each election, we look at the number of users in each member state, any previous instances of intentional information manipulation and any other known risks. The type of additional mitigation measures that will be put in place will be dependent on these risks.

Our proactive measures include having our external third-party experts provide additional support including by running targeted reports on election activity, using these experts to notify us of harmful trends and to search for these trends on our platform. We also have internal specialists monitoring election-related narratives based on the work of our external third-party experts. The experts' findings may lead to terms being added to the STL to limit or block the dissemination of harmful content on the platform. Our external partners also introduced dedicated election resources to support our enforcement efforts.

In addition, during election periods in certain countries, when Pinners search on Pinterest for topics like "voting" or "elections," we'll show them a search advisory at the top of search results directing them to non-partisan authoritative voting resources. The purpose of these advisories is to help Pinners searching for election-related terms easily access authoritative resources they can trust for more information about the elections happening in their country.

### **Control Effectiveness Rating - Overall**

Similar to last year and based on the design and implementation of these controls, we have assessed these controls as **Effective**. This rating is based on a number of factors, including the enhancements made to our machine learning models and the low number of users exposed to content that violates our Civic Participation Misinformation policy. Our most recent DSA Transparency Report (reporting period 25th September 2023 to 31st December 2023) shows that for violations of this policy, we proactively removed 132 Pins in the EU and removed five Pins as a result of user reporting. This rating is also based on the ongoing monitoring of the accuracy and coverage of our automated models and ongoing monitoring of users reports associated with content of this nature.

### **How influencing factors affect this risk**

In addition to considering the inherent risk rating and the controls we have in place, we considered how each of the influencing factors could negatively affect civic discourse, electoral processes and public security.

### **Applicable terms and conditions and their enforcement**

Civic discourse, electoral integrity and public security are broad concepts, and our policies and guidelines reflect this. We have multiple policies that work together to help prevent this risk from occurring on Pinterest. These policies and guidelines are key to our enforcement of content that might negatively contribute to civic discourse, election integrity or public security, and they guide our overall content moderation approach.

### **Content moderation systems**

Given the complexity of this risk area, Pinterest partners with external third-party experts to provide us with further support on risk areas like misinformation, election integrity and political issues specific to particular geographies. These experts help us better understand how these risks can manifest on Pinterest and they provide us with signals (like trending keywords) that we build into our overall content moderation system, including our managed list of sensitive terms. Our overall content moderation system also works to prevent the risk that Pinterest causes a negative effect on civic discourse, electoral integrity or public security. We have multiple mechanisms in place to detect and enforce violating content, such as user reports and manual reviews. See Pinterest's [content moderation ecosystem](#) for more detail on these mechanisms.

### **Design of recommender systems and any other relevant algorithmic systems**

Our systems recommend content primarily based on a user's previous activity and we know that users do not primarily come to Pinterest to participate in civic discourse or find information about elections. As such, the design of our recommender systems and the nature of our platform's purpose do not have a significant negative effect on political discourse or civic engagement.



## Systems for selecting and presenting advertisements

Pinterest's advertising system is designed to help decrease the risk that Pinterest could negatively impact civic discourse, electoral integrity or public security. Our [Advertising Guidelines](#) prohibit political campaign advertisements. We do not allow advertising for the election or defeat of political candidates running for public office, including fundraising for political candidates or parties, political parties or action committees, political issues with the intent to influence an election, legislation (including referenda or ballot initiatives), and merchandise related to political candidates, parties or elections. We also do not allow advertisers to target certain audiences, including based on political affiliation. Pinterest has multiple controls in place to enforce these guidelines, including manual review of ads, automated models to detect prohibited content and users reporting policy-violating ads.

## Data related practices

We do not consider our data practices to specifically impact civic discourse, electoral integrity or public security.

## Intentional manipulation

We've also looked at how intentional manipulation can affect the area of civic discourse, electoral processes and public security. Similar to the area of fundamental rights, we typically don't see spam attacks, fake accounts or ATOs targeted towards this area of risk, however, our controls work to detect and act against intentional manipulation when it occurs.

An additional factor that we have considered for this year's risk assessment is the area of the intentional manipulation of content itself through the use of GenAI or other manipulation efforts to produce fabricated, manipulated or edited content, such as deepfakes and shallowfakes. Our [Community Guidelines](#), including our Civic Participation Misinformation policy, as well as our internal enforcement guidelines contain details on our approach and efforts to action any policy-violating content. Our content moderation tools are equipped to identify and action this type of harmful content when it appears on the platform regardless of how the harmful content is created.

## Residual risk rating

Based on the inherent risk rating and the effectiveness of the controls we have in place, we have assessed the residual risk that Pinterest's design, functioning or use could lead to negative effects on civic discourse, electoral processes or public security in the EU as **Low**. This rating is the same as last year. Though there has been increased political action in the EU due to the number of elections scheduled for 2024, we're satisfied that the additional measures we implemented in response to this plus our belief that people do not come to Pinterest for political content keeps the residual risk rating for this risk at **Low**.

## Further mitigation efforts

Because of the potential harm that this risk can pose to users and EU society, we work hard to continuously enhance our policies and processes. Though people don't come to Pinterest for political content, we understand our responsibilities when it comes to civic engagement and we are making the following enhancements: We are making additional investments in our [Agent Training](#) team, which will enhance our detection and enforcement against content that could result in negative effects on civic discourse, electoral processes and public security. In addition, we are also making investments in our [Quality Assurance Team](#) to further enhance our current QA programme. Also, our [Risk Intelligence Engineering team](#) will be carrying out enhanced monitoring of data sources to identify policy grey areas and emerging trends across all content policy areas.

## 5.4 Systemic Risk 4: Negative effects in relation to gender-based violence, the protection of public health and minors, and serious negative consequences to the person’s physical and mental well-being

### Summary



### Risk Overview

In looking at the ways that these risks could manifest on our platform, we have looked at the volume of harmful content related to these risks, such as health misinformation, content promoting physical or mental harm, hate speech, harassment, or child sexual exploitation (“CSE”) content.

We also considered Pinterest’s design and whether this contributes to these risks — for example, are there adequate safeguards for minors, does the platform tap into users’ addictive behaviours, and do we provide users with options for how they engage on the platform.

The assessment below reflects our analysis of these individual risks in aggregate.

### Inherent risk rating

#### Inherent Risk - Probability

Similar to last year, we have assessed the probability of the risk that Pinterest’s design, functioning or use negatively impacts the protection of public health and minors, a person’s physical and mental well-being, or gender-based violence in the EU as **Possible**.

We know from our detection mechanisms such as user reports, that it is possible for users to be exposed to this type of content on Pinterest, however, data indicates that the number of users exposed to this type of content is low. Our most recent DSA Transparency Report (reporting period 25th September 2023 to 31st December 2023) shows that we received 32 user reports that resulted in Pins being removed for violating our Medical Misinformation policy, 595 user reports that resulted in Pins being removed for violating our Self-Injury & Harmful Behaviour policy, and 1,233 user reports that resulted in Pins being removed for violating our Child Safety policy.

#### Inherent Risk - Severity

Similar to last year, we have assessed the severity of this inherent risk if left unchecked as **Significant**, primarily because even limited exposure to content or behaviour of this nature for a small number of users could lead to harm or consequences.



However, rapid dissemination of content “going viral” is not common on Pinterest and this is critical to understanding the severity of this risk more broadly. We have looked at the concept of “reach” to understand this. If we take data from our most recent Global Transparency (H2 2023) report we can see that of the Pins removed for violating our Health Misinformation policy in Q4 of 2023, 95% were seen by fewer than 10 users during that reporting period. Also, looking at our Self-Injury and Harmful Behaviour policy, of the Pins removed in Q4 of 2023 for violation of this policy, 95% were seen by fewer than 10 people.

### Overall inherent risk rating

Similar to last year, based on the probability and severity ratings, we have assessed the inherent risk that the design, functioning or use of Pinterest negatively impacts the protection of public health and minors, a person’s physical and mental well-being, or gender-based violence in the EU as **Medium**.

### Controls and mitigation efforts

Our content moderation system is one of the main ways that we mitigate this risk. Our systems are capable of quickly detecting and removing harmful content. The other main way that we mitigate this risk is through the design of our platform. It is difficult to amplify risks to public health, minors, physical and mental health, and gender-based violence in a “viral” manner. Pinterest’s Recommender Systems are designed to prioritise high-quality and inspirational content — we have specific search features such as compassionate search, to guide users to well-being practices if they are trying to manage difficult emotions, and our search advisories give our users direct access to suicide prevention lifelines.

We’ve continued our work in this area to help make Pinterest a safer place for our users, focusing on public health, safety and health of minors, and the physical and mental well-being of all of our users. Further details on the steps we’ve taken in this area can be found [here](#).

Our [Community Guidelines](#) are the foundation upon which our content moderation systems are built — they outline what we allow on Pinterest and the type of content that we prohibit on our platform. In addition to our Community Guidelines, we have several enforcement guidelines that provide additional guidance and clarification to our content review teams and systems. Specifically pertaining to this risk we have our Child Safety policy, Dangerous Goods & Activities policy, Graphic Violence & Threats policy, Adult Sexual Services policy, Harassment & Criticism policy, Hateful Activities policy, Medical Misinformation policy, and Self-Injury & Harmful Behaviour policy. These policies and guidelines drive our overall content moderation approach, including how automated models are used, the types of content that users can report, our manual review process and our enforcement approach.

Pinterest has a longstanding commitment to creating a safe and positive place online, particularly for more vulnerable users, such as minors. We are committed to continuously designing and implementing measures that allow us to keep our young users safe and investing in our policies, products and partnerships to support the well-being of our community. We make changes in response to new behaviours and trends we’ve observed online, adopt technology available to combat harmful content, enhance brand safety for our advertising partners, and also simply because it’s the right thing to do for the people on our platform.

Given the potential severity and harm associated with these risks, Pinterest has the following controls in place to protect our most vulnerable users:

- We do not allow users under the minimum age of consent in their country to sign up for Pinterest. Where and when we discover any accounts of an underage user, we deactivate them.
- Private only: The profiles of users under the age of 16 are set to private as the default and only option, and in the EU, the accounts of users aged 16 and 17 are set to private by default with the option to switch to a public account. Private profiles are undiscoverable on Pinterest search and search engines, and profiles, boards and Pins for users with private profiles are only visible to followers approved by the user. Users with private profiles are able to connect with family and friends by sending a unique profile link, and all users have the ability to review and remove followers.
- Removal of existing followers: For teens under the age of 16, all existing followers were removed in August 2023, so that those users could start with a clean slate to decide who can follow them.

- **No contact without consent:** Teens under the age of 16 can only send and receive messages from mutual followers who have been accepted through a unique profile link that expires after three days or when they connect with five new followers using this unique link, whichever comes first. Teens who are 16 and 17 can only receive messages from mutual followers, and can only receive message requests from users they follow.
- **User reporting:** The reporting reasons in Pinterest's in-product content reporting tool were updated to include not only sexually explicit content, but also any content that sexualises minors. We have also enabled more nuanced reasons to report users and boards. Though users could always report content or accounts, the community is now able to flag a more detailed and nuanced list of behaviours at both the user and board level. This includes: impersonation, saving normally appropriate content in a potentially sexualised manner, among many others.
- **[Help Center](#) resources:** The Pinterest Help Center provides information to parents or caregivers of teens on Pinterest. It explains our minimum age requirements, provides privacy resources and specifies ways for parents to notify us if they suspect their underage child has a Pinterest account so it can be deleted.
- **Age change verification:** If a user who previously entered their age as under 18 attempts to edit their date of birth on the Pinterest app, Pinterest requires them to submit a government ID and/or selfie to a third-party partner to confirm their age.
- **Parental support:** Pinterest offers a parental passcode feature for users under the age of 18, which allows parents and caregivers to set up a 4-digit passcode to lock certain settings and have oversight of their teen's Pinterest account. These settings include those related to account management, privacy and data, public/private profile status, and social permissions, such as messaging, mentions, comments and shopping recommendations.

We know, however, that this work is never done, so we continuously re-evaluate the effectiveness of our safety features and how they fit into our broader goal of creating a positive, inspirational product experience. Part of that process is listening to feedback from our users, and accordingly, in partnership with an external third party, we held a workshop with teen users to see what they thought of teen safety features. While, of course, individual users' opinions differ, several trends are discernable among younger users concerning both online safety in general as well as the specific measures recently implemented by Pinterest.

Teen users tell us that they are interested in finding more positive content online, and rather than using Pinterest to communicate, they use it to tap into creativity and commemorate important moments in their lives. When it comes to online safety, teens tell us that they are interested in having the ability to create their own online experiences by having control over their actions and interactions. This includes maintaining privacy over their online activity so they can feel free to express their aspirations and explore their identities. While some teen users recognise the value of age verification to protect their online experience, limitations on their online access can make them feel isolated and undervalued. Accordingly, they express a preference for features that empower them to show up as good digital citizens.

We're not just concerned with the safety of the teens that use our platform — we also want to support their mental health and emotional well-being and the mental health and emotional well-being of all our users. We've designed our product to help further this mission. For example, we have no filters on beauty. Beauty filters that alter a person's appearance when posting online can change the way teens think about themselves. We've taken a stand and don't have those kinds of filters on Pinterest. Our virtual Try On tool is a compelling way to play with eye makeup and lipstick colours, but it won't alter the user's face because we believe they look great just the way they are.

To further our efforts supporting mental health and inspiring our users, we have created specific search features. Pinterest's compassionate search feature includes a collection of evidence-based well-being practices that someone can do to improve their mood if they are feeling anxious, sad or trying to manage difficult emotions. For example, if someone searches for "stress relief," they might choose the "redirect your energy" activity, which suggests practices like journaling for perspective, drawing a nature scene or making a playlist. If they select "accept your emotions," they'll be guided through steps to practise self-compassion. For people who may be experiencing thoughts of suicide or need someone to talk to immediately, we continue to provide search advisories with direct access to suicide prevention lifelines.

## Control Effectiveness Rating - Overall

Similar to last year, and based on their design and implementation, we have assessed these controls as **Effective**. This rating is based on a number of factors, including the enhancements made to our machine learning models, the ongoing monitoring accuracy and the data in our most recent DSA Transparency Report (reporting period 25th September 2023 to 31st December 2023). For example, we removed over 8,000 Pins for violation of our Health Misinformation policy and over 1,000,000 Pins for violation of our Self-Injury & Harmful Behaviour Policy, while we removed just over 500 Pins due to user reports for violations of both this policy.

## How influencing factors affect this risk

Similar to the other systemic risks, we've considered how each of the influencing factors affects risks to public health, minors, physical and mental health, and gender-based violence.

### Applicable terms and conditions and their enforcement

Per our [Terms of Service](#), if you're based in the European Economic Area, you may only use Pinterest if you are over the age at which you can provide consent to data processing under the laws of your country, and we require a date of birth for new and existing accounts for all users regardless of age. We have recently expanded our age verification process. If someone who previously entered their age as under 18 attempts to edit their date of birth on the Pinterest app, we will require them to send additional information to our third-party age verification partner to confirm its legitimacy.

We have specific policies and guidelines that address public health and the mental and physical health of our users, including minors. These policies drive our content moderation and enforcement approach.

### Content moderation systems

Our overall content moderation system works together to detect and enforce content which might negatively impact public health, minors, physical and mental health, and gender-based violence. Given the severity of this risk, we invested in research to understand how these risks could manifest on the platform and have built specialised controls and product features as a result. We have also made further investments in our Minor Safety Operations team, creating a dedicated team for this area. The members of this team are minor safety specialists who perform operational and investigatory review of content and behaviour that potentially violates Pinterest's Child Safety policy identified via reactive and proactive detection. We have also made additional investments with external third-party experts who monitor child safety trends on our platform.

### Design of recommender systems and any other relevant algorithmic systems

We have made deliberate choices to engineer a more positive place online and prevent our platform from negatively impacting mental health. In order to make sure that the Pinterest platform itself is additive, not addictive, we tune our algorithmic systems to prioritise explicit signals — for example, “saves” — over just views alone. When people see something on Pinterest that they want to act on, they hit “save.” By prioritising what gets “saved” in the content, the images and videos that are top performing don't distract users from their life (like car crash videos or conspiracy theories), but actually help users improve it (like step-by-step guides, self-care ideas, inspirational quotes and how-to videos). Building our algorithmic systems to prioritise high quality and inspirational content enables us to create a more positive environment that can support the mental health of our users.

### Systems for selecting and presenting advertisements

We do not want ads on our platform that might create a negative impact for our users. To protect the physical and mental health of our users, including minors, our [Advertising Guidelines](#) place restrictions on certain categories of advertisements, including drugs and paraphernalia, sensitive content (such as excessively violent or profane content), tobacco, alcohol, gambling products and services, and body shaming language and imagery.

In addition to these restrictions, we limit how ads can be targeted to certain audiences. Ads cannot be targeted based on sensitive health or medical conditions, among other things. In addition, ads cannot be targeted or served to minors based on profiling and currently are not being served to minors in the EU at all.

Our guidelines are restrictive to ensure that Pinner safety comes first and we have controls in place to enforce these guidelines, detailed in Pinterest's [platform ecosystem](#).

## Data related practices

Users may only use Pinterest if they are over the age at which they can provide consent to data processing under the laws of their country. If a user is the appropriate age to create a Pinterest account, their data is collected and used per the practices set out in Pinterest's [platform ecosystem](#).

## Intentional Manipulation

Similar to the other systemic risks, we've looked at how intentional manipulation can affect gender-based violence, the protection of public health and minors, and serious negative consequences to the person's physical and mental well-being. We typically don't see spam attacks, fake accounts or ATO's targeted towards this area of risk, however, our controls work to detect and act against this type of intentional manipulation when it does occur.

We also looked at how intentional manipulation of content itself, either through the use of GenAI or other manipulation efforts to produce manipulated or edited content impacts on this area of risk. Our content moderation tools do not distinguish between harmful content that is manipulated or edited and harmful content that has not been manipulated. Our content moderation tools work to identify and action all policy-violating and harmful content regardless of how it was created.

## Residual risk rating

Based on the inherent risk rating and the effectiveness of the controls we have in place, we have assessed the residual risk that Pinterest could negatively impact the protection of public health and minors, a person's physical and mental well-being, or gender-based violence in the EU as **Low**. There has been no change to the overall risk level since last year's risk assessment.

## Further mitigation efforts

Though we have assessed this risk as **Low**, the safety of our users is top of mind and we are always looking for ways to enhance our controls. We're working on the following enhancements in this area:

- 1. Search Advisory:** We are partnering with a new external expert to improve our search advisory resources for people who may need someone to talk to immediately. This new partnership will allow us to make sure our resources remain high quality and will also allow for different contact formats for our users.
- 2. Parental Support:** We've recently made changes to strengthen our parental support measures, including taking steps to verify the details of parents or caregivers looking to make adjustments to account management, privacy, profile status and social permission settings.

We are also planning for further improvements in this area, by making additional investments in our [Agent Training team](#) for our agents who carry out manual review. In addition, we will make additional investments in our [Quality Assurance team](#) to further enhance our current QA programme. Our plans to introduce a new [user reporting function](#) for our users to report suspected underage users will enhance the measures we already have in place for minor safety. As well as these measures, our [Risk Intelligence Engineering team](#) will be carrying out enhanced monitoring of data sources to identify policy grey areas and emerging trends across all content policy areas.

---

## 6. Conclusion

Our second annual DSA Risk Assessment and Mitigation Report builds on our previous analysis of the systemic risks that stem from the design, functioning or use of Pinterest and its systems, and highlights the areas where we can further mitigate those risks to protect our users. We want Pinterest to be a safer place online for all of our users. We will continue to invest in our content moderation systems and explore new ways and invest in technology to fight policy-violating content on our platform. We are always evolving to identify and adapt to new behaviours and trends, and we'll continue to work hard to fulfil our mission to bring everyone the inspiration to create a life they love.

# Appendix

## Risk assessment methodology

To determine the risk that the design or functionality of Pinterest could cause harm to users and EU society, we first developed a risk register. This register was developed in consultation with a wide group of stakeholders and utilised Pinterest’s existing understanding of how harm can manifest on the platform as well as focusing on the systemic risks identified by the DSA. We assessed each risk individually and then reported on these risks at an aggregated level. To assess each risk we used the following formula:

*Inherent risk rating x control effectiveness score = residual risk rating*

### Inherent risk rating

We considered:

- **Severity:** the impact that it would have on user groups and EU society in general; and
- **Probability:** the likelihood of this impact occurring.

Each element was giving a score using a four-point scale and the inherent risk rating was obtained using the following formula:

*Severity score x Probability score = Inherent risk rating*

#### Severity score rubric

Severity	Definition	Score
Marginal	Could lead to minor harm or consequences to users /society based on the number of people impacted and/or the type of harm.	1
Moderate	Could lead to moderate harm or consequences to users/society based on the number of people impacted and/or the type of harm.	2
Significant	Could lead to significant harm or consequences for users/society based on the number of people impacted and/or the type of harm.	3
Critical	Could lead to critical harm or consequences for users/society based on the number of people impacted and/or the type of harm.	4

#### Probability score rubric

Severity	Definition	Score
Unlikely	Could occur in exceptional/extraordinary circumstances.	1
Possible	Could occur in uncommon/unusual circumstances.	2
Likely	Could occur in relatively common circumstances.	3
Almost Certain	Nearly certain to occur.	4

### Inherent risk rating rubric

<i>Inherent risk</i>	<i>Score</i>
Very High	13 - 16
High	9 - 12
Medium	5 - 8
Low	1 - 4

### Control effectiveness

We identified the controls and safeguards in place to mitigate each risk and determined how effective the control environment is in mitigating the inherent systemic risk. We considered the design of the control and, where available, we looked at metrics and data to understand the effectiveness of the control. We did not perform control testing as part of the DSA risk assessment, although we plan to implement control testing as part of future DSA risk assessments.

### Control effectiveness rubric

<i>Rating</i>	<i>Definition</i>	<i>Score</i>
Highly effective	Control designed and implemented to reduce the risk almost entirely	0.25
Effective	Control designed and implemented to reduce most aspects of the risk	0.50
Somewhate effective	Control designed and implemented to reduce some aspects of the risk	0.75
Ineffective	Control has a very limited impact on reducing the risk	1.0

### Residual risk rating

The resulting cumulative systemic risk exposure was determined by multiplying the inherent systemic risk score against the control effectiveness score to determine the residual risk rating.

### Residual risk rating rubric

<i>Residual risk</i>	<i>Score</i>
Very High	13 - 16
High	9 - 12
Medium	5 - 8
Low	1 - 4

