Digital Services Act Risk Assessment and Mitigation Report 2023
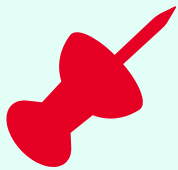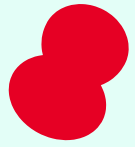
Contents

1. Executive summary

Pinterest is the visual inspiration platform people around the world, including in the European Union (EU), use to find ideas, shop personalised products, and discover the most inspiring content. 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. Not all content is inspiring, so we have Community Guidelines to outline what we do and don't allow on Pinterest, and we work hard to identify and deactivate harmful content from our site.

The Digital Services Act (DSA) came into force on 16 November 2022. The objective of the Act is to "give better protection to users and to fundamental rights online, establish a powerful transparency and accountability framework for online platforms and provide a single, uniform framework across the EU."[1] The Act 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 Pinterest poses to EU users and EU society is low. This assessment takes into account the impact that these risks might have if they occur and the probability of this impact occurring. We considered the reasons that users come to Pinterest, the ways that they use the platform and interact with others, and the limited ways that content can go "viral" on Pinterest. We also looked at the controls that 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.

We evaluated our platform ecosystem and the associated systemic risk that Pinterest could present to EU users and the broader EU society. We conducted interviews and surveyed a broad range of internal Pinterest stakeholders; we reviewed internal documentation and leveraged data, including information reported in our global transparency reports; and we 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; and related data practices.

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 risks categories, including updates to our recommender systems, expanding our machine learning tooling and infrastructure, increasing our use of third-party experts, enhancing our quality assurance programme, and championing the Inspired Internet Pledge, created by the Digital Wellness Lab at Boston Children's Hospital in collaboration with Pinterest.

User safety is critical for Pinterest. We've made and continue to make consistent, people-first decisions to protect our users. We strive to constantly evolve and enforce better content safety policies, and make deliberate choices to engineer a more positive place online

[1] https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

2. Introduction

2.1 Pinterest

Product overview

Pinterest is the visual inspiration platform people around the world, including in the EU, use to find ideas, shop personalised products, and discover the most inspiring content. Users ("Pinners") come to Pinterest to discover Pins they love, and save them to boards to keep their ideas organised and easy to find. On Pinterest, users can browse their home feed where they will find images of things, people, places and products ("Pins"), as well as people and businesses we think users will love, based on their recent activity. Users can explore suggested topics and trends or search for topics of their own; save, create, share, and shop Pins; 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, and it's our guiding light in drafting and enforcing our content policies. Not everything on the internet is inspiring, so we have guardrails for what's acceptable on Pinterest and what isn't allowed. Our moderation practices are always evolving to keep up with new behaviours and trends and to create a more positive place on the internet.

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.[2]" 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 first DSA systemic risk assessment and associated mitigation measures, including an overview of Pinterest's risk landscape and ongoing mitigation efforts to address those risks. 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.

[2]https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

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 and the broader EU society.

Risk categories

To determine the risk that Pinterest could present, we first developed a risk register in consultation with a wide group of stakeholders, utilising Pinterest's existing understanding of how harm can manifest on the platform as well as systemic risks identified by the DSA. These risks cover a wide range of areas like policy development and accessibility, product design and safety processes, detection and enforcement mechanisms, and our advertising and recommender systems. We also considered categories of harm, such as child safety, misinformation, exploitation and harassment.

We assessed each risk individually and aggregated them in four risk 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.

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 to any platform that connects users and 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; and

5. Related data practices.

These influencing factors encapsulate Pinterest's full platform ecosystem and we've assessed each factor as part of this risk assessment. See Section 4 for additional detail on these factors.

3.2 Risk assessment methodology

Based on our understanding of Pinterest's systemic risk landscape and our risk register described in Risk Categories, we evaluated our systemic risks by analysing a variety of sources, including:

- **Interviews**: We conducted ~40 interviews with a wide range of internal Pinterest stakeholders;

- **Questionnaires**: We collected information from eight questionnaires sent to 28 internal Pinterest stakeholders;

- **Documentation**: We reviewed internal documentation, such as policies, procedures and other control documentation;

- **Metrics and other data**: We leveraged data, such as the reach of Pins deactivated for violating specific policies, the number of actioned user reports, and other information included in our global transparency reports; and

- **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 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:** 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 this DSA risk assessment, although we plan to implement control testing as part of DSA risk assessments in the future. Each control was assigned a control effectiveness rating of Ineffective, Somewhat effective, Effective, or Highly effective.

- **Residual risk:** the level of risk leftover 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've 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 to inform our policies and content moderation practices. We are proud of our policies and practices because they're the right thing for the people on our platform and broader society. 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 of 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 for what to try next, cook next, wear next, 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; and

- **Enforcement Page**: how we put our policies into practice, including any restrictions that we may apply to users' content or use of Pinterest.

These policies and guidelines are applicable globally, with some nuances built in for local legislation or regulations. Our policies and guidelines are translated into 45 languages.

We use plain language along with short summaries to make sure that our Pinners understand the policies the first time they read them. All of our policies and guidelines are available on our Policy Page and are also searchable through our Help Center, making them easily accessible to Pinners and others.

Our Policy and Legal teams are responsible for drafting and maintaining Pinterest's policies and guidelines. We often engage with external third parties to get their input and feedback on any new policies or major changes to existing policies and to make sure that new and updated policies do not have a disproportionately negative effect on certain user groups. These external third parties can include non-profit organisations, independent experts, civic groups and our Pinners.

We also have internal content policies and enforcement rules for our content review teams. These internal policies build upon our Community Guidelines and other publicly facing policies to provide more nuance and details for our content review teams and systems to properly identify prohibited content and take appropriate enforcement action. Our process for revising all of our content policies is nimble and streamlined so we are able to quickly adapt to new risks and harm types.

In Influencing factor #2, we explain the ways we enforce our policies.

4.2 Influencing factor #2: Content moderation systems

Consistent with our policies, we have a robust framework in place designed to identify and take action on harmful content on our site, and our content policies and moderation practices are always evolving to keep up with new behaviours and trends.

We may block, limit the distribution of, or deactivate content and the accounts, individuals, and groups that create or spread harmful content and behaviour, based on how much harm the content or behaviour poses. We may also remove an account after a single instance of a severe policy violation or if we determine that the account has repeatedly posted policy-violating or illegal content. Pinterest has a strike system in place to address Pinners who repeatedly post content that violates our policies and guidelines. Repeated violations can result in temporary or permanent removal of the Pinner's account. Our Enforcement page provides additional information on these potential enforcement actions.

Reporting harmful content and behaviour

For people in the EU, Pinterest provides several ways to report content that are easily located, directly accessible, and always available.

Reporting policy violations

Reports are how users can tell us if they think something on Pinterest is in violation of our policies. Pinners 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. Reports can be submitted in-product or via the Help Center. 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). Once the user selects the option to report, they are prompted to choose the reason for which they would like to report the content. Depending on the policy reason for the report, the user may be asked for additional details to direct their report. Once they confirm their report, it is routed to a member of one of our review teams to review against our policies. Our internal content policies and enforcement guidelines assist these review teams to properly identify prohibited content and take appropriate enforcement action. These teams are specially trained in our content policies and team members are globally located, including in the EU.

Our in-product reporting flow also includes an option to report content for intellectual property violations, such copyright or trademark infringement. If a user clicks this reporting reason from the menu of report reasons, they are directed to a dedicated reporting form designed to request all the information required for our 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 specialist operations agent will take action on the content as appropriate. As noted, content may be deactivated or limited in distribution, depending on the violation type and severity.

Reports under local law

For people in the EU, we additionally provide a dedicated reporting channel where users and non-users can report Pinterest content they believe to be illegal under EU or Member State law. This reporting channel is accessible both in-product and via the Help Center, and is available to users and non-users alike. The form is available in all 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 in our Help Center.

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

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) and are reviewed by specialists on our Law Enforcement & Government Operations team. Where content is found to be violative of our policies, appropriate action will be taken globally; if not otherwise policy-violating, content that is confirmed to be locally unlawful will be blocked in the relevant jurisdiction. The Law Enforcement & Government Operations team is also trained and tasked with responding to legal process from law enforcement requesting user information. This team aims to protect Pinners and their data by balancing legal obligations with user privacy.

To strengthen the integrity of our reporting channels we may take actions to 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. As part of our Terms of Service, users agree to submit reports in good faith and to not misuse any reporting or appeals channel by making baseless reports or appeals.

Enforcement systems

We enforce our policies through automated tools, manual review and hybrid approaches that combine 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 actions

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.

Our automated tools are used for 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. Where a system detects policy-violating content, it will deactivate or limit the distribution of the content. 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 actions

We manually act on some Pins 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.

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 external partners based across two global locations, which supports scalability in different languages and time zones. Agents go through a robust training session at onboarding, followed by regular ongoing and ad hoc training. Agents also have access to PinU, Pinterest's self guided internal training portal. When there are updates to policies or guidelines, additional training is provided.

These external partners are responsible for following quality processes implemented by Pinterest's quality assurance programme. This programme focuses on verifying the accuracy and consistency of enforcement decisions. The results of the quality assurance audits are monitored and could lead to the implementation of additional controls, such as enhanced training or changes to our enforcement guidelines.

Hybrid 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

Another way that Pinterest enforces against potentially harmful or policy-violating content and behaviour is by engaging with 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 these trends occur on the Pinterest platform. 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 proactive sweeps of our platform based on signals provided by these third parties and update our moderation tools based on the results of those sweeps.

Managed list of sensitive terms

We also maintain a list of sensitive terms 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. Our list of sensitive terms is continually expanding as we identify online trends, both internally and with the support of third-party experts.

Notification and appeals

Notification

The Reports and Violations Center (RVC) is the central place for EU 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. This email directs the Pinner to the RVC, which provides additional information and where applicable, a detailed Statement of Reasons 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 will first receive an acknowledgement of our receipt of their report via email, and then will receive, also via email, a more detailed response outlining the results of our review of the report. Certain other content restriction decisions for which users will 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, users may also appeal content restriction decisions or decisions on reports they've submitted directly through the RVC. We review appeal 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 similar Pins. We also inform Pinners 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. 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. 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 updating training decks and delivering new training, updating enforcement guidelines, and updating 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 will publish a transparency report by 25 October 2023, and at least every six months thereafter, that will provide information on our content moderation activities in the EU.

4.3 Influencing factor #3: 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, inspiring, and personal. We look at a Pinner's activity on the platform—such as the content they save, Pins they view, and other Pinners they follow—to recommend new and relevant content, such as Pins, boards or other Pinners that we think users may like. We also use explicit signals that Pinners provide to us such as age, gender, and selected topics of interest, to recommend content, as well as information from their engagement with advertisers. Our algorithmic systems are designed to prioritise user action and engagement signals (such as Saves) so that recommendations are relevant to the user. In addition, Pinterest does not use dark patterns to nudge users to stay on the platform.

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

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.

There are areas where our Advertising Guidelines go beyond what is prohibited by law, as we want Pinterest to be a positive and inspiring place for everyone. For example, we prohibit weight loss ads and ads that body shame.

Pinterest is transparent when it comes to advertising, with all ads clearly labelled as 'Promoted By' to distinguish them from organic content. Pinterest also allows Pinners to see why they are being shown an ad via the WAISTA (Why Am I Seeing This Ad?) 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 information about each 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.

Where standard ads can be content that drives to any type of landing page an advertiser wants to promote, such as blogs and recipes, shopping ads drive directly to a shopping experience. Shopping ads are derived from a product catalogue which 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 data practices are governed by our Privacy Policy. 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 to use Pinterest, they voluntarily 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.

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 Pinners can close their account at any time. Pinners can control cookies and choose how and whether their photos and other data is shared with Pinterest.

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 Pinners' 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.

5. Systemic risk landscape

Based on our understanding of the risk that Pinterest might pose to EU users and EU society, we assessed a wide range of individual risks. We have reported on these risks at a rolled-up aggregated level across four categories of systematic risk.
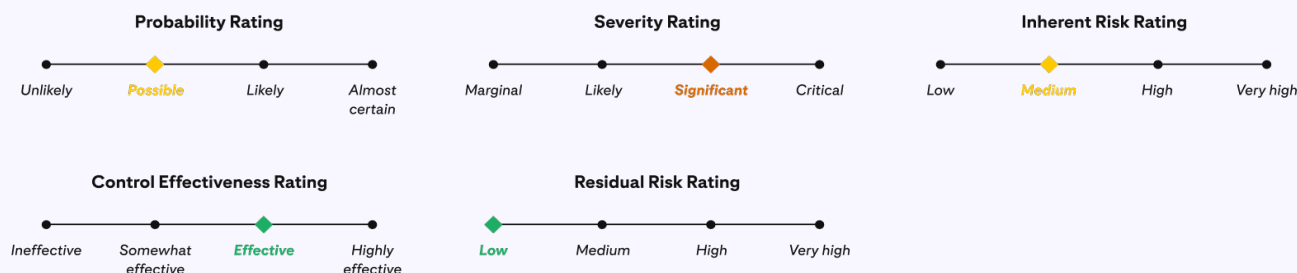
This section provides a summary of these systemic risks, how Pinterest is working to prevent these risks, and what we plan to do to further mitigate these risks. We have analysed the:

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 Illegal content

Summary

Probability Rating

Unlikely · **Possible** · Likely · Almost certain

Severity Rating

Marginal · Likely · **Significant** · Critical

Inherent Risk Rating

Low · **Medium** · High · Very high

Control Effectiveness Rating

Ineffective · Somewhat effective · **Effective** · Highly effective

Residual Risk Rating

Low · Medium · High · Very high

Risk overview

We have adopted the DSA's definition of illegal content as "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.[3]" 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, but not limited to, 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.

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 - 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 or EU society. 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.

[3]DSA, L 277/42

Inherent risk rating

Probability

We have assessed the probability of the risk of illegal content stemming from the design or functioning 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 of such content is relatively low. For example, as described in our most recent Transparency Report, in Q4 2022, we took appropriate action on the following numbers of reported violations of a particular content policy:

Number of user reports (resulted in a Pin deactivated)	Content policy
122,375	Adult Content
116	Adult Sexual Services
4,940	Child Safety
2,398	Dangerous Goods and Activities
8,690	Graphic Violence and Threats
4,968	Harassment and Criticism
11,180	Hateful Activities
1,526	Violent Actors

In addition to these content policies, we also report on government content deactivation requests in our most recent Transparency Report. Pinterest received a total of 15,284 content removal requests from government entities from July through December 2022, all of them from outside the United States. We deactivated content for 8,873 of those requests for violating our Community Guidelines and restricted content on an additional 2,977 requests. Content for 3,430 requests was inactive by the time it was reviewed in response to the government removal request. This can happen when, for instance, the content was deactivated in response to a user report prior to Pinterest receiving the government removal request.

While 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 relatively low compared to the billions of Pins on Pinterest.

Severity

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, rapid dissemination or content "going viral" is not common on Pinterest and this is critical to understanding the severity of this risk for EU society more broadly. We have looked at the concept of "reach" to understand this. To calculate this metric, we start by looking at each policy-violating Pin deactivated 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 1 second before it was deactivated. Taking the Dangerous Goods and Activities policy as an example, of the Pins deactivated in Q4 2022, more than 99% were seen by fewer than 10 users in the reporting period. Taking our Child Safety policy as another example, of the Pins deactivated in Q4 2022 for Child Sexual Exploitation content, 96% were seen by 100 or fewer users in this reporting period. Looking at 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.*

Controls and mitigation efforts

Pinterest's first line of defence for mitigating the dissemination of illegal content is our Terms of Service, which explicitly state that users of Pinterest may not post content that is in any way unlawful. Consistent with those Terms, our Community Guidelines outline the types of content and behaviour prohibited from the Pinterest platform and reiterates that users may not "do anything or post any content that violates laws or regulations." 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 counties.

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 content to be deactivated, content where distribution needs to be limited, and content which is allowed.

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.

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 low number of users exposed to illegal content, as reported in our H2 2022 Transparency Report, ongoing monitoring of the accuracy and coverage of our automated models, and ongoing monitoring of user reports associated with illegal content.

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 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 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.

Whilst 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. Our automated models are used for some types of content (such as adult content, graphic violence, hate speech, and illegal drugs) and we will be leveraging this technology to further expand coverage. See Further Mitigation Efforts for more information.

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 deactivates or blocks access to content from the country or countries where it is illegal.

A key element of our content moderation system is our human review process, which has a robust quality assurance programme to ensure that the decisions made by review agents are consistent, accurate and in line with our content moderation policies. We are continuously improving our quality assurance programme to further mitigate the systemic risk of illegal content. See Further Mitigation Efforts for more information.

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.

There are two main controls that we have in place to prevent this from occurring. The first is our overall content moderation system which seeks to identify and deactivate illegal content from the platform, both proactively and in response to user and other third-party reports. The second is our managed list of sensitive terms, which prevents a user's search from returning any results for certain terms which 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 and our efforts include ensuring that our recommender systems don't contribute to the dissemination of illegal content. See [Further Mitigation Efforts](#) for more information.

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 which prohibit certain types of ads from being targeted to certain regions.

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 as specifically impacting the dissemination of illegal content.

Intended manipulation

As well as the influencing factors listed above, we've also considered how intended manipulation could impact the risk of illegal content.

Unfortunately, bad actors may seek to manipulate the Pinterest platform. This includes spam attacks or bad actors using fake accounts, and there are multiple ways for this to manifest on the platform. For example, spammers may seek to make money from Pinners clicking on links that point to a spam website they own, with display ads or other monetization on the website. Or spammers may spread malware links and then monetize the network of infected devices, or spread phishing links and then monetize the stolen user credentials. Account Takeovers (ATOs) can also occur on the Pinterest platform which is where attackers gain access to existing accounts (for example through leaked 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 intended manipulation. 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, as outlined in the [decisions appeals process](#) section. Users can also report content and profiles for suspicions of spam.

As reported in Pinterest's H2 2022 Transparency Report, of the Pins deactivated in Q4 for spam, more than 99% were seen by 100 or fewer users in this reporting period.

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*.
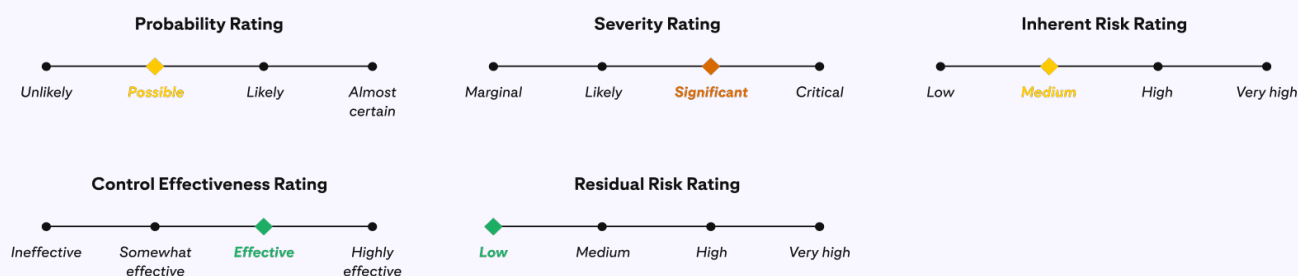
Further mitigation efforts

Despite the low residual risk rating, we are continuously improving our control environment because of the potential harm that illegal content can pose to users and EU society. These efforts include the following:

1. **Recommender systems** - Our recommendations systems learn from user actions and engagement. It is critical that we don't use the engagement of bad actors to train our models. We are continuing to implement methods to address this and prevent bad actor's actions and content from influencing our recommendations.

2. **Automated models** - We continue to invest heavily in efforts like machine learning. Our work continues to improve existing models and build new models to fight policy-violating content on Pinterest.

3. **Third-party experts** - We are expanding our work with third-party experts to inform our policies and content moderation practices. This ongoing effort will allow us to better leverage external experts in these fields.

4. **Quality assurance** - We are working on enhancing our quality assurance programme. This includes developing new tooling functionalities to increase the size and variance of the sample size selected for quality assurance. With these changes, our goal is to further enhance our quality assurance programme to ensure our enforcement decisions are consistent, accurate and in line with our content moderation policies.

5.2 Negative effects on the exercise of fundamental rights

Summary

Probability Rating				Severity Rating				Inherent Risk Rating			
Unlikely	**Possible**	Likely	Almost certain	Marginal	Likely	**Significant**	Critical	Low	**Medium**	High	Very high

Control Effectiveness Rating				Residual Risk Rating			
Ineffective	Somewhat effective	**Effective**	Highly effective	**Low**	Medium	High	Very high

Risk overview

In assessing the risk that Pinterest could have actual or foreseeable negative effects for the exercise of fundamental rights, we have used the [Charter of Fundamental Rights of the EU](#) as our guide. We have specifically focused on the fundamental rights that are most relevant to Pinterest's platform, including but 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[4], 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.

---

[4]Considered as part of Systemic Risk #4.

In assessing these risks, there is a balance to 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. How we have struck this balance is illustrated in our Controls and mitigation efforts section.

Inherent risk rating

Probability

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. Fundamental rights are threaded throughout our Community Guidelines, for example 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.

Severity

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. 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 it could lead to 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*.

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, which provide guardrails for the type of content and behaviour permitted on Pinterest and are always grounded in user safety and societal good. Whilst some users may disagree with policy stances we have taken, for Pinterest, user safety is critical. 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 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. For example, we've developed unprecedented inclusive features such as Hair Pattern Search and the Skin Tone Ranges, because no one should have to work harder 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 Pinners with disabilities. For some surfaces, these updates include 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 & DIY, Home Design & Decor, and Food & Drink[5].

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. 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 allowed and disallowed on Pinterest. And we have specific policies to assist people to engage on Pinterest in a positive, inspirational, and respectful way. This includes our Harassment and Criticism Policy; and we prohibit content that insults, hurts or antagonises individuals or groups of people. This includes 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.

[5]Pinterest Brand Equity Survey, US, UK, Germany & France, n=1,365, March 2023.

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 in the EU, 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 shows 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.

Overall

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 ongoing monitoring of the controls discussed above, such as 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.

However, given the potential harm that this risk can pose to users and EU society, we work hard to continuously improve our content moderation ecosystem. See [Further Mitigation Efforts](#) for more information.

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. For example, when making updates to our Adult Content policy we engaged with GLBTQ Legal Advocates and Defenders, the National Black Justice Coalition and the National Center for Transgender Equality to prevent our policy from negatively impacting or discriminating against certain groups.

Our policies and guidelines, and the enforcement of these, impact several of the fundamental rights discussed above, including freedom and 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 as well as a quality assurance 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 user's 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, respect for private and family life.
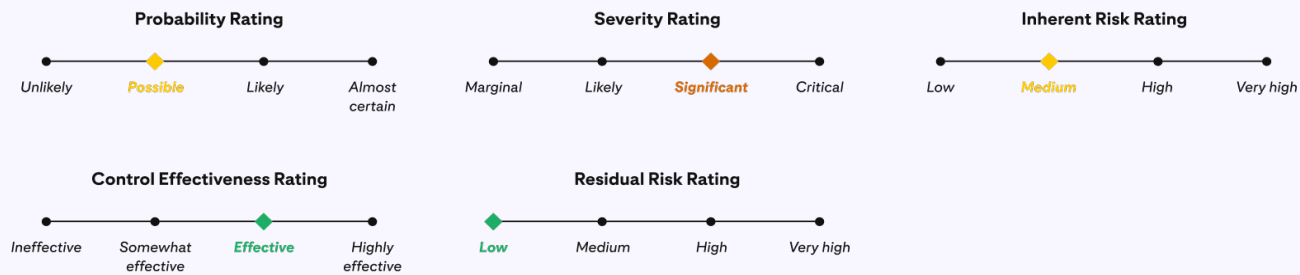
Residual risk rating

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 users or EU society as *Low*. However, given the potential harm that this risk can pose to users and EU society, we are working hard to continuously improve.

Further mitigation efforts

Due to the potential harm that this risk of impacting fundamental rights can pose to users and EU society, we are in the process of implementing additional controls related to automated models, third-party experts, and quality assurance, all of which will help to detect and enforce our guidelines against content that may negatively harm the fundamental rights of users. The Illegal Content section of this report has more details on the improvements we are making in each of these three areas.

5.3 Negative effects on civic discourse, electoral processes and public security

Summary

Probability Rating

Unlikely — **Possible** — Likely — Almost certain

Severity Rating

Marginal — Likely — **Significant** — Critical

Inherent Risk Rating

Low — **Medium** — High — Very high

Control Effectiveness Rating

Ineffective — Somewhat effective — **Effective** — Highly effective

Residual Risk Rating

Low — Medium — High — Very high

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, and who is allowed to participate in an election or census. 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

Probability

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, that it is possible for users to be exposed to this type of content on Pinterest; however, the volume is low compared to the billions of Pins on Pinterest. Per our most recent Transparency Report, in Q4 2022, we actioned 159 user reports that resulted in a Pin deactivated for violating our Civic Misinformation policy.

We know from research that Pinterest isn't typically a platform where users come to seek information on elections or to participate in civic discourse. The top three categories that monthly users say they come to Pinterest for are Craft & DIY, Home Design & Decor, and Food & Drink[6].

Severity

We have assessed the severity of the inherent risk that the design, functioning, or use of Pinterest, in 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, rapid dissemination or content "going viral" is not common on Pinterest and this is critical to understanding the severity of this risk for EU society more broadly. We have looked at the concept of "reach" to understand this. Taking the Civic Misinformation policy as an example, of the Pins deactivated in Q4 2022, 98% were seen by fewer than 10 users in the reporting period. Looking at these metrics across our policies that aim to combat negative effects on civic discourse, electoral integrity or public security, rapid dissemination of this content is not common on Pinterest.

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*.

[6]Pinterest Brand Equity Survey, US, UK, Germany & France, n=1,365, March 2023.

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 provide guardrails for appropriate civic participation on the Pinterest platform.

We also have internal content policies and enforcement rules for our content review teams. Our Civic Participation Misinformation policy builds upon our Community Guidelines to provide more details for our content review teams and systems to properly identify prohibited content and take appropriate enforcement action.

Our Violent Actors policy also prohibits violent content, groups or individuals. We limit the distribution of or deactivate 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.

Pinterest's Conspiracy Theories policy also works to prevent negative impacts on electoral processes by prohibiting content intended to delegitimize election results on the basis of false or misleading claims.

Civic discourse is a broad concept and we have other policies that also work together to prevent negative effects. This includes our Graphic Violence and 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. See Pinterest's platform ecosystem for more detail on these mechanisms.

We prohibit political campaign ads (see Systems for selecting and presenting advertisements for additional information). In addition, for certain elections, when Pinners search on Pinterest for topics like 'voting' or 'elections', we'll show them a search advisory directing them to non-partisan authoritative voting resources.

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 low number of users exposed to harmful content of this nature, as reported in our H2 2022 Transparency Report, 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.

To further enhance our ability to make enforcement decisions that are consistent, accurate, and in line with our content moderation policies, we are in the process of making improvements to our Quality Assurance programme. See Further Mitigation Efforts for more information.

Content moderation systems

Given the complexity of this risk area, Pinterest partners with an external expert to provide us with expert support on risk areas like misinformation, election integrity, and political issues specific to particular geographies. This expert

helps us better understand how these risks can manifest on Pinterest and they provide us with signals (like trending keywords), which 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 to find information about elections, so 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 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 as specifically impacting civic discourse, electoral integrity or public security.
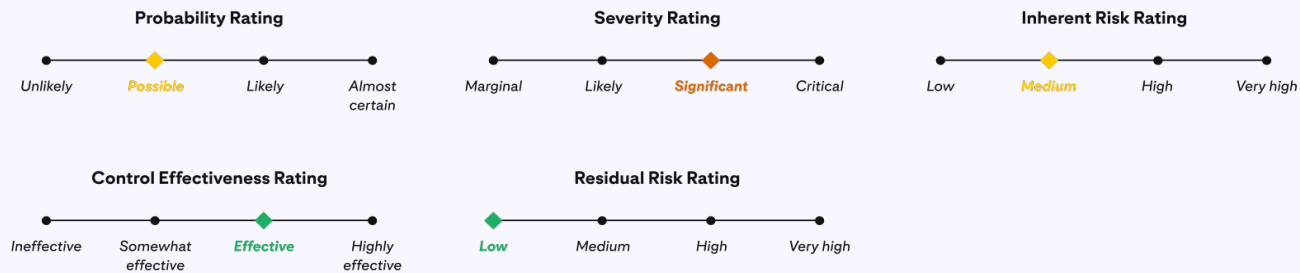
Residual risk rating

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*.

Further mitigation efforts

Because of the potential harm that this risk can pose to users and EU society, we work hard to continuously improve. More information about the steps we're taking can be found in the Illegal content section of this report.

5.4 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

Probability Rating

Unlikely — **Possible** — Likely — Almost certain

Severity Rating

Marginal — Likely — **Significant** — Critical

Inherent Risk Rating

Low — **Medium** — High — Very high

Control Effectiveness Rating

Ineffective — Somewhat effective — **Effective** — Highly effective

Residual Risk Rating

Low — Medium — High — Very high

Risk overview

We have considered the potential ways that these risks could manifest on Pinterest. We 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 user's addictive behaviours, and do we provide users with options in how they engage on the platform. The assessment below reflects our analysis of these individual risks in aggregate.

Inherent risk rating

Probability

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, the volume is relatively low compared to the billions of Pins on Pinterest. Per our most recent Transparency Report, in Q4 2022, we actioned 219 user reports that resulted in a Pin deactivated for violating our Medical Misinformation policy, we actioned 3,071 user reports that resulted in a Pin deactivated for violating our Self Injury and Harmful Behaviour policy and we actioned 4,940 user reports that resulted in a Pin deactivated for violating our Child Safety policy.

Severity

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 or content "going viral" is not common on Pinterest and this is critical to understanding the severity of this risk for EU society more broadly. We have looked at the concept of "reach" to understand this. Taking the Medical Misinformation policy as an example, of the Pins deactivated in Q4 2022, 91% were seen by fewer than 10 users in the reporting period. Taking the Self Injury and Harmful Behaviour policy, of the Pins deactivated in Q4 2022, 95% were seen by fewer than 10 users in the reporting period. Looking at these metrics across our policies that aim to combat these types of harmful content, rapid dissemination of this content is not common on Pinterest.

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 protection of public health, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence in the EU as *Medium*.

Controls and mitigation efforts

We mitigate this risk in two main ways - our overall content moderation system, which quickly detects and removes harmful content, and the design of our platform, which makes it difficult to amplify risks to public health, minors, physical and mental health, and gender-based violence in a "viral" manner.

With respect to our content moderation system, our Community Guidelines serve as the foundation of the type of content and behaviour that is prohibited on Pinterest. In addition, Pinterest has detailed internal content policies that provide additional guidance and clarification to our content review teams and systems, such as our Child Safety policy, Dangerous Goods and Activities policy, Graphic Violence and Threats policy, Exploitation guidelines, Harassment and Criticism policy, Hateful Activities policy, Health Misinformation policy, and Self Injury and 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. See Pinterest's platform ecosystem for more detail on these mechanisms.

Given the potential severity and harm, we have implemented additional controls designed to address these risks. Pinterest conducted research of its teen users to understand the risks they face on social media through surveys, online journaling and in-depth interviews. We learnt from this study that teens and their parents are looking for more safety features that put them in the driver's seat. This research helped inform the additional controls we have implemented to protect teens and minors, including:

- **Private by default:** The accounts of Pinterest users under the age of 18 will default to private, which means their accounts won't be discoverable by others. Users aged 16-17 have the option to switch to a public account if they choose to do so.

- **No contact without consent:** Currently, boards and Pins created by teens under 16 are not visible or accessible to anyone but the user and their collaborators. This content is not incorporated into our recommender systems.

- **Help Center resources:** The Pinterest Help Center provides information to parents 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.

- **User reporting:** We have enabled more nuanced reasons to report users and boards. Though users could always report content or accounts, the community is now able to flag for a more detailed or nuanced list of bad behaviours at both the user and board level. This includes impersonation or saving of normally appropriate content in a potentially sexualized manner, among many others.

- **Parental support:** We are developing more options for parents and guardians who want to support teens under 18 online. For example, parents now have the ability to require a passcode for their teen to change certain settings related to account management, social permissions, privacy, and data.

We want to do more than make Pinterest safer for teens; we also want to support teen mental health. A recent study we conducted with UC Berkeley's Greater Good Science Center suggests 10 minutes a day looking at any inspiring content could help young people guard against stress and toxicity.

We've designed our product to help further this mission. For example, we have no filters on beauty. Beauty filters and changing appearance every time people post 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 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 someone can do to improve their mood if they are feeling anxious, sad or trying to manage difficult emotions[7]. 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 direct access to suicide prevention lifelines.

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 low number of users exposed to harmful content of this nature, as reported in our H2 2022 Transparency Report, 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 the inherent risk rating and controls we have in place, we considered how each of the influencing factors could amplify 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, no matter what 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. The 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. See the [Controls and mitigation efforts](#) section.

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 not addictive, we tune our algorithmic systems to prioritise explicit signals, not just views alone. We use a unique set of metrics that we call "inspired actions." An example of one of those actions is "saves" - 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 nourishing and inspirational content enables us to create a more positive environment that can support the mental health of our users.

---

[7]Currently available in the following EU countries: Ireland, Germany, France, Italy, Spain, Austria, and Sweden.

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 put in place restrictions for certain categories of advertisements, including drugs and paraphernalia, sensitive content (such as excessively violent or profane content), tobacco, alcohol, gambling products and services, and weight loss products and services.

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](#).

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, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence in the EU as *Low*. However, given the potential harm that this risk can pose to users and EU society, we are taking additional steps to further mitigate this risk.

Further mitigation efforts

In addition to the mitigation efforts already listed for the other systemic risks, because of the harm that this risk can pose to users and EU society, we're implementing an additional mitigation specifically targeted at this risk category. We have partnered with the Digital Wellness Lab to develop the Inspired Internet Pledge. On 20 June 2023, we announced Pinterest as the first signatory and are committed to championing the pledge amongst our peers. We commit to taking meaningful, measurable actions to do the following in service of supporting more positive mental and emotional wellbeing outcomes for all people, and especially young people:

- Understand and tune for behaviours that enhance emotional wellbeing;

- Listen to and act on insights from people who have experienced harm online; and

- Share lessons collaboratively across the tech industry.

6. Conclusion

Our first annual DSA Risk Assessment summarises our 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. The work is never done: we continue to invest heavily in measures like machine learning technology to fight policy-violating content on Pinterest and to work with outside experts and organisations to inform our policies and content moderation practices. Our content moderation practices are always evolving to adapt to new behaviours and trends, and to create a more positive place on the internet for everyone.

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 4 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

Probability	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

Inherent risk	Score
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

Rating	Definition	Score
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
Somewhat 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

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