



EU Digital Services Act

Trustpilot's Transparency Report

This report is published by Trustpilot A/S in accordance with the transparency reporting requirements under Article 15 of the European Union's Digital Services Act ("DSA"). This report contains information for a reporting period from **17 February 2024 to 16 February 2025**.

Published: 16 May 2025

Summary

Trustpilot began in 2007 with a simple yet powerful idea that is more relevant today than ever — to be the universal symbol of trust, bringing consumers and businesses together through reviews.

Our ambition is to help consumers make the right choices and help businesses to build trust, grow and improve. We have five trust principles which form the foundation for how we build and maintain a platform people can rely on, namely: Neutral, Open, Fair, Transparent, and Relevant. These principles underpin how we operate, guiding our unique approach to how we uphold the integrity of our platform.

As a provider of intermediary services and in accordance with the reporting requirements under the DSA's Article 15, we publish this further report regarding our content moderation efforts in the EU that are regulated by the DSA. The information contained in this report is for the reporting period from 17 February 2024 through 16 February 2025.

In addition to this report, we release an annual [Transparency Report](#) explaining how we safeguard our community of businesses and consumers, and provide global insights and metrics into activity on our platform. To gain a more holistic understanding into the actions we take to protect and promote trust online, we encourage you to explore Trustpilot's [Trust Website](#). We also recommend our most recent [Transparency Report](#), where you can learn more about how we ensure content on our platform is accurate, helpful and safe for our community of businesses and consumers.



Table of Contents

[EU Digital Services Act](#)

[Summary](#)

[Table of Contents](#)

[Section 1: Member State authority requests](#)

[A. Member State removal orders](#)

[B. Member State disclosure orders](#)

[Section 2: Article 16 notices](#)

[A. Trustpilot approach](#)

[B. Article 16 notices submitted by Trusted Flaggers](#)

[C. Table 2.1 - Article 16 notices](#)

[D. Table 2.2 - Removal actions in response to Article 16 notices](#)

[Section 3: Trustpilot's own initiative content moderation](#)

[A. Trustpilot Approach](#)

[B. Trustpilot's Automated Technology](#)

[C. Trustpilot's Content Moderation Team](#)

[Coverage, Training & Quality](#)

[Support & Wellness](#)

[D. Table 3.1 - Trustpilot's Proactive Content Moderation Actions](#)

[Section 4: Complaints \(i.e. appeals\) received through internal complaint handling systems](#)

[A. Table 4.1 - Appeals](#)

[Section 5: Out-of-court settlement body disputes](#)



Section 1: Member State authority requests

Article 15.1(a)

A. Member State removal orders

DSA Article 9 empowers EU courts and government agencies to submit legal requests to remove content on Trustpilot. These “orders to act” are issued on the basis of the applicable Union law or national law in compliance with Union law (removal orders). Trustpilot has created a [special form](#) for Member States to submit removal orders.

During the reporting period, Trustpilot received **zero** removal orders from EU Member States’ authorities issued in accordance with Article 9.

B. Member State disclosure orders

DSA Article 10 empowers EU courts and government agencies to submit legal requests to provide specific information about recipients of Trustpilot’s service. These “orders to provide information” are issued on the basis of the applicable Union or national law in compliance with Union law (disclosure orders). Trustpilot has created a [special form](#) for Member States to submit disclosure orders.

During the reporting period, Trustpilot received **zero** disclosure orders from EU Member States’ authorities issued in accordance with Article 10.

Section 2: Article 16 notices

Article 15.1(b)

A. Trustpilot approach

DSA Article 16 requires Trustpilot to put mechanisms in place to allow any individual or entity to notify us of the presence on our service of specific items of information that the individual or entity considers to be illegal content (Article 16 notices).

Trustpilot’s community of businesses and consumers have always played a role in promoting and protecting trust on the platform. Our business and consumer communities can flag content to us at any time if they believe it breaches our [Guidelines for Reviewers](#) or [Guidelines for Businesses](#) (together, our “Guidelines”). We provide our businesses and consumers with a variety of easy to access mechanisms to report illegal content or other Guideline violations.



Business users can [flag reviews via their Trustpilot Business account](#) to report reviews they consider are illegal or otherwise go against our guidelines. If businesses don't have an account, or need to explain in more detail why they consider the content to be illegal content, they can reach out to us via our [contact form](#). Businesses can also use the contact form to report other content, like usernames or profile pictures.

Consumers can use the [consumer flagging tool](#) to flag reviews that they consider are illegal or otherwise go against our guidelines. Alternatively, if consumers prefer to contact us without an account, or need to explain in more detail why they consider the information in question to be illegal content, they can reach out to us via our [contact form](#). Consumers can also use the contact form to report other content, like business replies to reviews, usernames, or profile pictures.

We assess every reported review, using a combination of automated tooling and our trained content moderation specialists. We also have a Fraud & Investigations team, who consider and act on any particularly complex situations. Trustpilot assesses any reported content against its Guidelines, which take into account illegal content considerations. To the extent content is reported for allegedly violating specific local laws, our content moderation team will liaise with our legal team to identify whether the review should be filtered for violating local law. If a review breaches our Guidelines and is filtered, the review is no longer visible anywhere on the platform.

B. Article 16 notices submitted by Trusted Flaggers

In addition, under DSA Article 22, Trustpilot must ensure that Article 16 notices submitted by Trusted Flaggers (individuals or entities designated as trusted flaggers pursuant to the DSA) are prioritized for review. Trustpilot has created a [special form](#) for Trusted Flaggers to submit Article 16 notices, to ensure prioritisation.

During the reporting period, Trustpilot received **zero** Article 16 notices from Trusted Flaggers issued in accordance with Article 22.

C. Table 2.1 - Article 16 notices

This table reflects the number of reports of EU-based users' content received by Trustpilot during the reporting period. This includes reviews (and other content, such as business replies to reviews) reported via the consumer flagging tool, Business accounts and our contact form, during the reporting period broken down by type of alleged illegal content. This data would also include any notices submitted by Trusted Flaggers but, as noted above, we received zero Article 16 notices from Trusted Flaggers.

This data includes notices that we may later deem to be invalid.



Wherever possible, we have categorised content in line with the Quantitative Template, as set out in Annex I of the implementing regulation on transparency reporting under the DSA. Our internal reporting terminology does not map directly to the Quantitative Template, so we have used our best endeavours to do so.

Type of alleged illegal content	User notices
HIDDEN_ADVERTISEMENT	47,874
DEFAMATION	51,554
OTHER_ILLEGAL_OR_HARMFUL_SPEECH	10,737
TERRORIST_CONTENT	642
INAUTHENTIC_USER_REVIEWS	251,822
OTHER_VIOLENCE	3,831
Not Captured by Other Subcategory ¹	38,043

D. Table 2.2 - Removal actions in response to Article 16 notices

This table reflects the number of reviews (and any other content, such as business replies to reviews) that were filtered in response to reports by EU-based users.

Wherever possible, we have categorised content in line with the Quantitative Template, as set out in Annex I of the implementing regulation on transparency reporting under the DSA. Our internal reporting terminology does not map directly to the Quantitative Template, so we have used our best endeavours to do so.

Type of alleged illegal content	Number of content items filtered
HIDDEN_ADVERTISEMENT	13,192
DEFAMATION	7,576
OTHER_ILLEGAL_OR_HARMFUL_SPEECH	1,675
TERRORIST_CONTENT	70

¹ Trustpilot enables its users to report content which breaches its Guidelines. The bases for which users can report content extends beyond illegal content (and the categories contained in the Quantitative Template). The “Not Captured by Other Subcategory” relates to content filtered for reasons including, for example, containing personally identifiable information or content which relates to another business.



INAUTHENTIC_USER_REVIEWS	94,382
OTHER_VIOLENCE	517
Not Captured by Other Subcategory	18,747

Section 3: Trustpilot's own initiative content moderation

Section 15(1)(c)

A. Trustpilot Approach

Our [Guidelines](#) apply to all and encourage responsible behaviour among our review community. Both consumers and businesses sign up and must adhere to our Guidelines when using the platform. Our Guidelines govern the platform and apply equally to both consumers and businesses.

We make use of a range of safeguarding techniques to protect the Trustpilot platform - Technology, Community, and People. This helps to ensure that content on the Trustpilot platform adheres to our Guidelines, and we maintain the trustworthiness and quality of our online review community.

Our Community's support in protecting the Trustpilot platform is explained at Section 2 above.

B. Trustpilot's Automated Technology

Trustpilot uses advanced technology to analyse content submitted by users, even before the review is live on our platform. For example, the written content of every review submitted to Trustpilot's platform is assessed by our AI-led moderation system, aimed at detecting violations of our Guidelines. This includes identifying reviews containing content that is harmful and illegal, advertising or promotional, or shares sensitive personal information. Where the technology identifies a potential breach of Guidelines, the review content is immediately hidden. Depending on the type of potential breach, the user will be notified of action taken. This includes, for example, moving the review permanently offline, asking the user to edit the problematic content, or sending the review content for manual review.

At Trustpilot, building trust means getting moderation right. We're committed to a balanced approach, carefully weighing the need to filter violating content against the importance of minimizing errors and respecting legitimate expression. Before any automated tool goes live, it undergoes rigorous testing with human oversight. Our deployment is phased, allowing us to closely monitor performance at each step. Once live, we continuously analyze data, including appeals and permitted content, to fine-tune our models. This ongoing vigilance ensures our



moderation remains accurate, fair, and effective, upholding the integrity of our platform for everyone.

C. Trustpilot's Content Moderation Team

In addition to technology, we rely on our people to safeguard the Trustpilot platform. Trustpilot has a team of 90+ employees across its Trust and Trust Tech teams. The teams' skillsets include policy and public affairs experts, content moderation agents and training specialists, fraud and fake review investigators and analysts, lawyers, technology and data experts, and communications experts. All individuals support our focus on safeguarding the Trustpilot platform. In addition, Trustpilot outsources various content moderation tasks to allow for flexibility in managing the volume of review activity.

Our team of content moderation experts - we refer to them as the Content Integrity team - are responsible for investigating and taking action in response to content which breaches our Guidelines. This action is more fully explained in our [Action We Take policy](#), and can include suspending or blocking user access to their accounts, in the event of repeated breaches of our Guidelines.

Coverage, Training & Quality

Our Content Integrity team has a thorough training and enablement plan for any new internal specialist or outsourced partner agent. New joiners undergo a comprehensive onboarding program spanning several weeks. This in-depth training focuses on Trustpilot policies, processes and procedures, and includes regular one-on-one coaching to discuss Quality Assurance results. We also offer focused training sessions on specific topics and themes, incorporate reinforcement through refreshers and quizzes, and utilize a variety of engaging training modalities such as in-person workshops and blended learning programs that effectively apply adult learning methodologies. Our Content Integrity team also has access to a detailed online repository of articles and resources for ongoing consultation.

In addition, our Content Integrity team has a well documented Knowledge Management Process, designed to capture, organise, transfer and maintain knowledge for our Content Integrity specialists and partner agents. There is a Quality Assurance Process, whereby a specified number of tickets are selected (on a weekly basis) by our Quality Management System for review. We also maintain an [appeals process](#), enabling dissatisfied users to challenge content moderation decisions (see more at Section 4 below).

In assessing the number of human resources required to support Trustpilot's content moderation activities, we use historic data (e.g. number of tickets handled, average handle times, and overall review volume) to compute an estimated forecast of hours required.



Support & Wellness

At Trustpilot, we believe that our employees' safety and well-being are paramount. We want our employees to feel empowered and fully supported to take action and make the adjustments they need to thrive. To ensure this, we offer a range of resources designed to support our employees including:

- **Employee Assistance Program:** We provide access to an independent, free, and confidential counselling service available 24/7. This service offers emotional support and short-term counselling, allowing our employees to connect with professional counselors and explore a wealth of resources, including articles, videos, webinars, and podcasts, all aimed at helping them navigate through difficult times.
- **Regular Manager Check-ins:** Consistent meetings with in-line managers provide dedicated time for open communication, support, and guidance.
- **Mentorship Opportunities:** The option to have a mentor offers valuable peer support, guidance, and opportunities for professional and personal growth.
- **WAMH (Wellbeing and Mental Health):** Our employee resource group, WAMH, is dedicated to nurturing the mental health and well-being of all employees. They achieve this through peer-to-peer support, educational initiatives, and proactive advocacy.

D. Table 3.1 - Trustpilot's Proactive Content Moderation Actions

This table reflects the content moderation activities taken at Trustpilot's own initiative. The table does not include data for content moderation actions taken in relation to reviews identified as fake, on the basis that these fall within the scope of the "deceptive high volume commercial content" exception at DSA Article 17(2).

Wherever possible, we have categorised content in line with the Quantitative Template, as set out in Annex I of the implementing regulation on transparency reporting under the DSA. Our internal reporting terminology does not map directly to the Quantitative Template, so we have used our best endeavours to do so.

Content Category	Own Initiative Measures
OTHER_CYBER_VIOLENCE	10
DEFAMATION	3
OTHER_ILLEGAL_HARMFUL_SPEECH	1,018
TERRORIST_CONTENT	28
OTHER_SELF_HARM	27



INCITEMENT_VIOLENCE_HATRED	645
OTHER_VIOLENCE	13
ADULT_SEXUAL_MATERIAL	148
OTHER_VIOLATIONS	27,366

Section 4: Complaints (i.e. appeals) received through internal complaint handling systems

Article 15.1(d)

When we filter a piece of content for violating our Guidelines, or take action against a user account (e.g. restricting or blocking access), we send an email to notify the user of the action taken.

In our emails, we aim to educate users about how and why they breached our Guidelines, to encourage positive future contributions to Trustpilot. In our emails, we also explain that the user can reply to our email if they think we made the wrong decision and we'll review the decision. To the extent that the user remains dissatisfied with our subsequent and final decision, the user has the right to appeal our decision by clicking the "Dispute Our Decision" link in the email. We refer to this as our Decision Dispute Process, which you can read about [here](#). We reverse our original decisions if we determine that our initial assessment was incorrect. We process appeals in a timely, non-discriminatory, diligent and non-arbitrary manner.

A. Table 4.1 - Appeals

Wherever possible, we have categorised content in line with the Quantitative Template, as set out in Annex I of the implementing regulation on transparency reporting under the DSA.

Please note that this data relates to our appeal volume, globally. It is not currently possible to provide a dataset limited to EU-based users only.

Metric	Value
Appeals	20,705
Upheld Decisions	16,221
Reversed Decisions	3,427
Median Time to Reach Decision (Hours)	189
Decisions Omitted	1,057



Section 5: Out-of-court settlement body disputes

Article 24.1(a)

Trustpilot received notice of **zero** disputes submitted to out-of-court dispute settlement bodies in accordance with Article 21 during the reporting period.