

# **Trustpilot response**

# **Online Safety**

# **Call for evidence**

**13 September 2022**

This document sets out Trustpilot's response to Ofcom's Call for evidence on Online Safety.

## Preliminary question

**Q1. Please provide a description introducing your organisation, service or interest in Online Safety.**

### 1.1 Trustpilot's online services

Trustpilot is an online consumer reviews platform that brings businesses and consumers together to foster trust and inspire collaboration. Our vision is to be a universal symbol of trust. We are free to use, open to everybody and built on transparency.

Trustpilot hosts user-generated content in the form of reviews. These reviews help consumers shop with confidence, and deliver rich insights to help businesses improve the experiences they offer. The more consumers that use our platform and share their own opinions, the richer the insights we offer businesses, and the more opportunities they have to earn the trust of consumers from all around the world.

### 1.2 Our interest in online safety

We welcome the UK government's aim to create safer online experiences and the opportunity to respond to this consultation. As a review platform, Trustpilot has two key interests in Online Safety:

- As a business, we proactively work to keep our platform a safe and trusted space for the consumers and businesses who use it — this is a priority for us and central to all of the work we do.
- As an online reviews platform, we will be within the scope of some of the upcoming changes proposed in the Online Safety Bill. We are therefore working to understand and (once the Bill is finalised) implement these.

### 1.3 Number of consumers and businesses who use Trustpilot, globally and in the UK

Over 167 million reviews have been submitted to Trustpilot, including over 46 million reviews in 2021. In 2021, we recorded a total of 508 million business profile page<sup>1</sup> views globally.

[REDACTED]

[REDACTED]

---

<sup>1</sup> Each “business profile page” on Trustpilot is where we display all of the reviews written about the business, in reverse chronological order. We also show the TrustScore and star rating awarded to the business, information about how and whether the business uses Trustpilot (does it have a paid subscription, reply to its reviews, etc.) contact information such as the business's website address, and any extra information the business may have provided about itself.

- [Redacted]
- [Redacted]
- [Redacted]

[Redacted]

[Redacted]

- [Redacted]
- [Redacted]
- [Redacted]

#### 1.4 Business model and revenue generation

We operate what’s commonly called a ‘freemium’ business model. Consumers have the ability to use Trustpilot at any time completely free of charge. Businesses can also use our basic services free of charge to invite their customers to review them, respond to reviews and report them. Businesses on our free plan can also display selected widgets, get notifications about new reviews posted, access our self-service Help Center and see basic review statistics. In fact, during 2021, around 93% of the 714,000+ business websites reviewed on the platform used it for free.

Trustpilot generates nearly all of its revenue from businesses choosing to subscribe to our additional software services. Companies who use our paid-for model can access extra features such as analytics to monitor their performance and access real-time insights on customer feedback, and a range of options to showcase Trustpilot reviews in their marketing channels. For example, our review insights offering helps businesses explore customer feedback using sentiment trends and insights into specific topics of interest to help them improve.

---

<sup>2</sup> We define this per month as: the business received at least one TrustBox (or widget) impression, or had at least one review written in response to an invitation.

### 1.5 Global and UK revenues

For the 2021 financial year, Trustpilot's global revenue was USD 131.4 million. Our UK revenue for 2021 was USD 53.1 million.<sup>3</sup>

### 1.6 Functionalities of Trustpilot's online service

Trustpilot is an online reviews platform that hosts user-generated content in the form of reviews about businesses. These are publicly accessible reviews that can be read by consumers who visit Trustpilot and search by business name or domain name, or by category. Given the nature of our services, not all of the 'functionalities' listed in the Bill are provided by or relevant to Trustpilot:

User-to-user service functionality	Provided by Trustpilot?
<p><b>Creating a user profile, including an anonymous or pseudonymous profile</b></p>	<p>To read reviews left on Trustpilot, visitors to our site don't need an account. However, to submit a review, a user must create a profile and username and this must be attached to a valid, permanent email address. The username selected for public display can be a pseudonym, but Trustpilot's <a href="#">Guidelines for Reviewers</a> require it to reflect the user's identity and it must not impersonate another person. We work hard to maintain the quality of reviews on Trustpilot and to protect the integrity of our community. This is why every review must be connected to a user profile that has a real person behind it.</p> <p>In early 2022, we introduced the option for consumers using Trustpilot to verify their identity by safely and securely sharing a copy of government-issued photo ID, along with a selfie. This is entirely optional, but it lets reviewers choose whether or not to provide an additional 'trust signal' for those reading the information: consumers who have verified their identity will have their reviews marked as coming from a 'verified reviewer'.</p>
<p><b>Searching within service for user-generated content or other users</b></p>	<p>Trustpilot allows readers of reviews to search for individual businesses to read the reviews written about each business, or to search for top-rated businesses within different categories of products/services. These functions allow readers of reviews to see reviews which have been left by other users for each business.</p> <p>It is not possible to search by reviewer name or username on the site to find a review written by a particular reviewer. It is also not possible to search user-generated content across the Trustpilot website.</p> <p>Where each review is displayed on a business profile, the reviewer's username appears alongside the review. Readers can click through via the reviewer username to read all of the reviews on Trustpilot authored by that particular reviewer.</p>

<sup>3</sup> See Trustpilot's [2021 Annual Report](#).

<p><b>Forwarding content to, or sharing content with, other users of the service</b></p>	<p>This is not a function of our site.</p>
<p><b>Sharing content on other internet services</b></p>	<p>Consumers can manually share individual reviews on Trustpilot by creating a post on Facebook or Twitter. This function is accessible via the "Share" button below each review shown on Trustpilot. [REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>For businesses, Trustpilot has several features that allow businesses the option to share their star rating and showcase reviews. This can be via widgets on their websites, in their own materials, and via posts made to Facebook and Instagram. Businesses sharing to their own social media sites can publish a post once or automatically post updates at regular intervals using a function in the business app. Businesses will usually choose to showcase positive reviews. For transparency to consumers, all of our widgets clearly state which types of reviews have been selected for showcasing.</p>
<p><b>Sending direct messages to or speaking to other users of the service, or interacting with them in another way (for example by playing a game)</b></p>	<p>Sending direct messages between users is not a function of our site, nor can users comment on reviews left by other users. Users are only able to indicate that they find an individual review helpful by clicking on the "Useful" button next to that review.</p> <p>Businesses who have claimed their profile page on Trustpilot are able to leave public responses to reviews of their company (this is the case regardless of whether they are using our free services, or have a paid subscription with Trustpilot). However, Trustpilot's systems do not facilitate private conversations between businesses and reviewers.</p>
<p><b>Expressing a view on content, including, for example, by applying a "like" or "dislike" button or other button of that nature, applying an emoji or symbol of any kind, engaging in yes/no voting or rating or scoring content in any way (including giving star or numerical ratings)</b></p>	<p>Trustpilot users can leave a star rating and write a review of their experience with a business. Companies can then publicly respond to these reviews if they so wish, provided that they have claimed the profile page hosting the reviews. Each response from the reviewed business is displayed on Trustpilot below the review in question.</p> <p>Users can flag reviews if they breach our rules and guidelines.</p> <p>Readers of reviews are able to manually designate individual reviews as "Useful" by pressing on the thumbs up icon shown with each review. However, this doesn't influence the weighting or the order of display of the reviews. Readers will be able to see how many users have considered an individual review useful (if any), and which reviewers found the review useful (reviewer usernames are displayed as a list when the reader uses their</p>

	<p>cursor to hover over the numerical tally). No other buttons or emojis are available for consumers to respond to posts with.</p>
<p><b>Sharing current or historic location information with other users of the service, recording a user's movements, or identifying which other users of the service are nearby</b></p>	<p>The information displayed on Trustpilot is limited to user location by country. Reviews show the username of the reviewer, how many reviews they have written on Trustpilot, and the country of their location. For example, a reviewer located in London will have "GB" displayed next to a location icon with their review.</p> <p>It is not a function of our site to suggest whether other users are nearby.</p>
<p><b>Following or subscribing to particular kinds of content or particular users of the service</b></p>	<p>This is not a function of our site.</p> <p>Each business that has been reviewed on Trustpilot and claimed its profile page will be able to access a business account. This includes the option to set up an email notification that alerts them when they receive new reviews written about their business on Trustpilot. This function is provided in order to help businesses log in to their account and respond to new reviews in a timely manner.</p>
<p><b>Creating lists, collections, archives or directories of content or users of the service</b></p>	<p>It is not a function of our site to create lists, collections, archives or directories of users.</p> <p>Trustpilot has a "Categories" function that allows users to compare businesses with other similar businesses, or discover new businesses by searching for products or services within a particular category.</p> <p>In order for businesses to learn from the feedback they receive in the form of reviews, it is possible to export publicly available review content from Trustpilot via an Excel spreadsheet or similar. This option is designed to allow businesses to analyse the content of reviews written about them on Trustpilot. This function is not available for consumers.</p>
<p><b>Tagging or labelling content present on the service</b></p>	<p>For consumers, this is not a function of our site.</p> <p>Businesses are able to do limited tagging and labelling of reviews written about their services within their business account. Reviews can be tagged via the business account with useful keywords to assist companies in understanding the feedback they are receiving. However, the tagging or labelling is <i>only visible to the business concerned</i> when they log into Trustpilot's business app. It is not visible to and does not in any way affect how reviews are presented to readers of reviews on Trustpilot.</p>
<p><b>Uploading content relating to goods or services</b></p>	<p>The Trustpilot platform lets consumers write individual, text-based reviews about their experiences with businesses. Such reviews sometimes include details about the goods or services that were purchased or used.</p>
<p><b>Applying or changing settings on the service which affect the</b></p>	<p>User-generated content on Trustpilot is limited to individual reviews posted by consumers who have had a genuine experience with a business. Trustpilot users and readers of reviews must search for a particular business</p>

<p><b>presentation of user-generated content on the service</b></p>	<p>by name, domain name, or via category of product/service in order to read the reviews written about it.</p> <p>On each business profile page, reviews are displayed in reverse chronological order, appearing as they're posted, with the most recent reviews at the top. For flexibility, readers of reviews have the option to adjust the settings in order to display reviews by language (for example, they can choose to see only reviews written in Spanish), or by their star rating (for example, only five-star reviews), or they can search by keyword (for example, to show all reviews that include the word "delivery"). [REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>[REDACTED]</p> <p>Readers of reviews who use our "Categories" function will by default see businesses listed in order of highest TrustScore and highest number of reviews. This information is explained on the categories page, where it states that default filter settings show companies as best in each category if they're actively asking for reviews and have received 25+ reviews in the past 12 months. However, readers of reviews can easily adjust the filters to search by city or postcode, number of reviews, time period, or company status.</p>
<p><b>Accessing other internet services through content present on the service (for example through hyperlinks)</b></p>	<p>Reviews posted on Trustpilot cannot contain hyperlinks. This is prohibited by our <a href="#">Guidelines for Reviewers</a> and generally prevented by an upload filter.</p> <p>To assist consumers in understanding which business has been reviewed, each business profile page on Trustpilot includes the relevant website domain name and the business's logo. Readers of reviews can click on either of these to directly access the business or domain that has been reviewed.</p>

## Risk assessment and management

### Q2. Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services?

At Trustpilot, we bring together a community of consumers and businesses across the globe who can share reviews and learn from each other's experience. Trust is at the heart of this, and we take any misuse of our platform, be it via illegal content or content which breaches our guidelines, very seriously. In this section, we first provide background information about how Trustpilot operates as an open platform, which provides context for the types of content posted on our site.

#### 2.1 Trustpilot is an open platform

As an open platform, Trustpilot provides consumers with the freedom to share reviews about their genuine experiences as and when they choose, for free, without waiting for an invitation from a business, provided they create a Trustpilot account and follow our guidelines. Businesses can ask consumers for feedback and respond at any time, for free.

In line with this *open* principle, reviews on Trustpilot are published without interference, other than being run through our technology to filter reviews with a strong likelihood of being fraudulent, and applying an automated language filter to prevent the submission of reviews with profanities or links. In 2022 we also introduced a third-party API system to detect and flag reviews that contain illegal content, such as violent threats, before they are published on our site (see our answer to **Q12**).

Being an open platform, however, doesn't mean that there are no rules, or no safeguards. All consumers and businesses using Trustpilot must follow our guidelines, and these explicitly prohibit posting illegal content (see also our answer to **Q6**).

Alongside our flagging mechanisms (detailed in our answer to **Q8**), Trustpilot's purpose, design and the functionalities it delivers are not conducive to spreading content widely at speed. Reviews hosted on our consumer site are text-based, and as stated above in our answer to **Q1**, forwarding content to, or sharing content with, other users of the service is not a function of our site, nor is sending direct messages to or speaking to other users of the service, or interacting with them in another way (for example, by playing a game). Further, if individual reviews are marked as "useful" then this does not amplify their status or affect their display or ranking in any way. As reviews are also character-limited, extremely long posts are not possible. The presentation of information is controlled by the user who must search for a particular business by name, domain name or via category of product/service in order to read the reviews written about it. Trustpilot does not display a personalised or curated "feed" of review content for users.

A majority of businesses and consumers use the Trustpilot platform constructively, and in the way it was intended. However, there is a small minority who use the platform in a way that doesn't meet our guidelines — and this can involve posting illegal content.

#### 2.2 Types of illegal content





Examples of “illegal content” defined by the Bill and submitted to Trustpilot include hateful speech in response to political or social issues; for example, in 2022 review bombing of a crowd funding platform that took a controversial stance on the Ukraine war. In such cases, our anomaly detection systems flag the unusual increase in review volume and our Fraud & Investigations Team can analyse whether immediate action is needed to remove reviews. We also see some obscene content, or reviews containing violent threats or discrimination made against business staff or owners, typically where customers have had a particularly bad experience. Such reviews are often promptly flagged by the businesses concerned and we remove them. See our answer to **Q8** for an outline of our flagging mechanisms.

Since consumers can, in theory, review any business with a valid domain name on Trustpilot, we sometimes receive reviews about businesses that are engaged in inherently illegal activities. We call these ‘bad-fit businesses’ and have a program for their detection and removal. Illegal content in reviews can also include scam or spam messaging that constitutes fraudulent online activity.

2.3 Trustpilot’s Transparency Reports

As part of our commitment to be a universal symbol of trust, Trustpilot publishes an annual ‘Trust and Transparency Report’ on our website which is downloadable by everyone. Within this report, we detail how we safeguard our platform and protect our users, including both how we handle harmful and illegal content, as well as providing specific data in relation to this. See the [“Trust” section of our site](#), which hosts our [2021 Transparency Report](#) and our [2022 Transparency Report](#).

2.4 Reviews flagged as harmful or illegal

As our [2022 Transparency Report](#) relays, in 2021 consumers across the globe reported 12,177 reviews for being illegal or harmful, whilst 64,678 were flagged by businesses on this basis. This is a small proportion of the 46.7 million reviews written in 2021.

Further, not all of these reports were assessed as *valid* by our expert Content Integrity Team: consumers had an overall flagging accuracy rate of only 16%, while businesses had a flagging accuracy rate of 77.2% across all categories. This is reflected in the (lower) number of reviews removed under each flagging reason.

[Redacted text block]

[Redacted text block]

[Redacted text block]

- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]

We note that the average flagging accuracy rate for consumers (globally), at 16% in 2021, has improved compared to an accuracy rate of only 12.4% in 2020. We believe this improvement is a result of our work to enhance the clarity of our flagging reasons and processes in 2020, including providing more education about what types of content we will remove. However, we note that the generally low level of accuracy also reflects the subjective nature of assessing content within its unique context. We will continue to educate both consumers and businesses to help build their understanding of our flagging reasons and processes.

At Trustpilot, we are committed to preserving the right to freedom of expression, which is integral to our open platform model, and we also acknowledge that freedom of expression is not an absolute right and does not protect all speech. To preserve freedom of expression, we generally set a high bar for the removal of content and explain where we draw the line as clearly and transparently as possible. We let users know via the flagging process that we will not remove content simply because readers of reviews don't like it or because they personally find it offensive. Unfortunately, we still receive a proportion of reported reviews that fall within these latter two categories, and many of these will be assessed as "invalid" reports. In contrast, many of the reports we assess as "valid" and therefore keep offline are deemed to not only breach our guidelines, but are often also likely to be illegal. See **Appendix 1** for examples of reviews flagged for harmful or illegal content on Trustpilot. [We note that none of these include content that is illegal to hold or view, such as child sexual abuse.]

2.5 Reviews that contain advertising material or are flagged as scams

We also have the option for Trustpilot users to flag advertising or promotional messages of any kind that are disguised as reviews.



In 2021, our global figures show that, of 46.7 million reviews written on Trustpilot, 60,236 were flagged by consumers as “advertising or promotional”, and 32,870 were flagged by businesses for that reason. [REDACTED] We class these reviews as “spam”.<sup>4</sup>

A separate group of reviews contain content linked to fraudulent scams. We define these “scam” reviews as non-genuine reviews which aim to exploit the reader by tricking or misleading them into taking some form of action based on the content of the review. For example, they can include information about services for spying or hacking, promotion of investments, money recovery, cryptocurrency, or even provide fake phone numbers purporting to be for the reviewed business’s customer service and prompting the reader to contact them. [REDACTED]

[REDACTED] Globally, we removed just over 1.2 million reviews within this category.

## 2.6 Reviews about businesses who engage in illegal activities: Bad-fit businesses

Other illegal content on Trustpilot can arise from users writing reviews about businesses whose activities are inherently illegal. We have taken increasing measures to remove businesses, or not accept businesses as customers, where we deem those businesses are not suitable for our platform (for example, because they promote hatred or facilitate criminal activities). This includes displaying consumer warnings on profiles, removing profiles that offer illegal or harmful services and preventing Trustpilot’s sales teams from communicating with such businesses.

In our publicly available [Action We Take Policy](#), we explain to users that we’ll take steps to remove and block certain particularly unethical businesses from our platform. We refer to these businesses as being a ‘bad-fit’ for Trustpilot.

A bad-fit business can cause or create harm, and they don’t share our values and core beliefs; it’s for these reasons that we simply don’t want them on Trustpilot. The types of businesses we consider to be bad-fits are often those that are engaged in behaviour that is likely to be illegal, such as those that:

- aid businesses or individuals to manipulate news, reviews, documents and results;
- promote hatred, violence, terrorism, xenophobia or any form of discrimination against any individual or group;
- offer illegal products and services; illegal drugs, prescription drugs sold on the illegal market;
- provide or facilitate the provision of escort services, mail-order brides, prostitution or any form of forced labour or human trafficking;
- offer or produce any sexual abuse or explicit imagery, and any material that presents children or animals in a sexual or illegal manner;
- organise illegal animal fights or sell products made with endangered animal parts;

---

<sup>4</sup> “Spam” reviews include reviews that have been posted multiple times with identical/almost identical content by the same individual, whether based on genuine experience or not and without the obvious intention to exploit consumers; or reviews that consist partly or wholly of content that advertises or promotes a businesses. Examples include promotional codes in reviews, marketing spam or SEO-gaming spam, repeated posting of the same or similar non-genuine review content, or reviews left as a result of a media storm that don’t reflect a genuine experience, etc.



- facilitate criminal activities of any sort, including those carried out by means of computers or the Internet;
- are engaged in financial scams such as pyramid schemes, credit card fraud, mortgage scams; or, otherwise operate illegally; for example, selling fake or unsafe goods and services, or generally mis-sell.

Users of Trustpilot and readers of reviews can report potential bad-fit businesses to us, and we'll look into each report and take action if needed. If we agree that the business is a bad-fit we'll remove them from Trustpilot, and can also ensure that they don't reappear. We also have an automated system that scans our platform and either removes domains (and notifies the businesses concerned) or flags them to our Content Integrity agents for manual investigation.

[REDACTED]  
[REDACTED] We are continually improving our processes for accurately detecting such domains and removing them.

### **Q3. How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?**

#### 3.1 Risk of harm

Given the type of service that Trustpilot is, we consider the risk of harm to individuals in the UK from illegal content presented by our service to be relatively low.

As outlined in the table above (at **1.6**), the purpose, nature and functionalities that the Trustpilot service provides act as a frame for how people use our services and the types of content posted on Trustpilot. A number of these aspects are relevant to the risk of harm to Trustpilot's UK users, and specifically, the *absence* of certain functionalities means that risks associated with those are not present. For example:

- every user profile must be connected to a real, valid and permanent email address, and reviewers can also choose to verify their identity as an extra trust signal
- it is not possible to use Trustpilot as a direct, user-to-user messaging service, nor can businesses use our service to send private messages to reviewers. It is also not a function of our site to allow users to share and forward content with each other on Trustpilot
- readers of reviews choose the information they see — they must actively search for the businesses they wish to read reviews about and are not provided with a curated “feed” of content decided by an algorithm. Readers of reviews cannot search by reviewer name to find reviews written by a particular reviewer, nor can readers glean detailed location information or non-public details about individual reviewers from our interface
- where readers of reviews designate a particular review as “useful”, this will not affect how it is displayed on Trustpilot
- reviews written about businesses are currently limited to text, without the ability to post photos or videos, and review content cannot contain any links, and
- we generate nearly all of our revenue from business subscriptions to extra features, rather than relying on advertising. We only display ads on Trustpilot's US website.

### 3.2 Identifying risks

It is our ambition to earn the trust of both consumers and businesses globally, and to be the platform that ignites trust between the two. Inherently, our work to build trust involves identifying and reducing any risks of harm to our users, while also balancing this with our commitment to protecting people's rights such as freedom of expression, in line with our open platform model. We have invested significant resources in safeguarding our platform — including from illegal content, but as ways to manipulate online content constantly evolve, we must also constantly adapt to stay ahead.

All employees are trained in our mission<sup>5</sup> and core values, which strongly emphasise building trust. While Trustpilot is in the process of moving to a formal and structured risk-based approach, many of our teams are already incorporating risk assessment as a necessary, day-to-day part of their work because it is essential to building and maintaining trust with users.

In how we operate, we provide employees with a framework of foundational principles such as:

- Neutral: Being a neutral platform that allows consumers and businesses to help one another, but we stay independent of both
- Equal treatment: We treat all reviews equally, and require consumers and businesses using Trustpilot to follow our guidelines
- Open: We provide consumers with the freedom to share their genuine experiences as and when they choose, for free, and businesses can ask consumers for feedback and respond at any time, for free
- Transparent: We provide clear and consistent communication about what we do, and why we do it. All businesses have transparency pages (or “Company Activity” pages) showing our community exactly how they engage with reviews on Trustpilot.

The structure of our organisation and the high level of human oversight embedded throughout our processes and procedures provides a continuous feedback loop on the functioning of our platform, and the risks that can arise through its use. Our community of businesses and consumers are also equipped with multiple options to alert us to illegal content and behaviour, including a flagging function, contact form, whistleblower mechanism and the ability to write a review about the Trustpilot service on Trustpilot. We also periodically gather feedback via surveys and customer satisfaction scores.

Trustpilot's Trust & Transparency (T&T) Team is an important source of information about illegal content on Trustpilot (see also our answer to **Q4** for detail on our governance structure), any risks posed to our users, and any changes that need to be made to our service to mitigate these. The T&T Team is positioned as an essential part of Trustpilot, integral to its strategy and relevant to all facets of our work.

T&T includes experienced Content Integrity (CI) agents that handle flagged reviews and respond to reports about harmful or illegal content on a day-to-day basis. They must act in accordance with our internal content policies, processes and procedures.

---

<sup>5</sup> Trustpilot's vision is to be a universal symbol of trust.

Our policy creation process necessarily incorporates data indicating the prevalence and quality of different types of harmful or illegal content on our platform, as well as research on how users respond to it, the different competing fundamental rights at play such as the right to freedom of expression (and how these can be balanced against the right to non-discrimination and consumer protection, among others), and best practices to prevent and address illegal content online. Policy decisions incorporate input from members of management across our Trust & Transparency Team,

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[REDACTED]

3.4 Exacerbating and mitigating risk factors – users, business model, features/functionality

The purpose of Trustpilot is to be useful to consumers. Building trust is integral to our services, which means it is in our interest to *design for trust*, and to minimise the risk of potential harm to our users, whether by deliberate manipulation or inadvertent misuse.

As stated at **1.3**, Trustpilot’s business model is software-as-a-service (SaaS) based, with revenue derived from the sale of subscriptions to businesses. We also receive a very small amount of advertising revenue from ads displayed on US business profile pages,<sup>6</sup> but our SaaS-model ensures that our focus remains on creating useful and trustworthy features and functionalities for businesses and consumers.

[REDACTED]

[REDACTED]

---

<sup>6</sup> As an example, in 2019, Trustpilot’s annual group revenue was USD 81,915,000, whereas our advertising revenue for the same period was [REDACTED]. In 2022, Trustpilot only displays ads on the US version of our website.

[Redacted text block]

[Redacted text block]

[Redacted text block]

**Q4. What are your governance, accountability and decision-making structures for user and platform safety?**

Trustpilot’s governance, accountability and decision-making structure for user and platform safety reflects our organisational focus on building trust and, as an integral part of that, prioritising user safety and maintaining a high level of transparency about how we work.

4.1 Senior management oversight

*Trust & Transparency Committee*

Trustpilot has a Trust and Transparency Committee, which is made up of our Chief Trust Officer as Chair, together with four Non-Executive Board Directors and supported by other Management and senior leaders who are invited to present on their areas of expertise. The Committee was established in March 2021, with the aim to continue to embed trust and transparency throughout the Trustpilot business. The Committee receives regular reports on and is responsible for overseeing the policies and procedures we use to maintain the integrity of Trustpilot’s products and services.

*Senior management level*





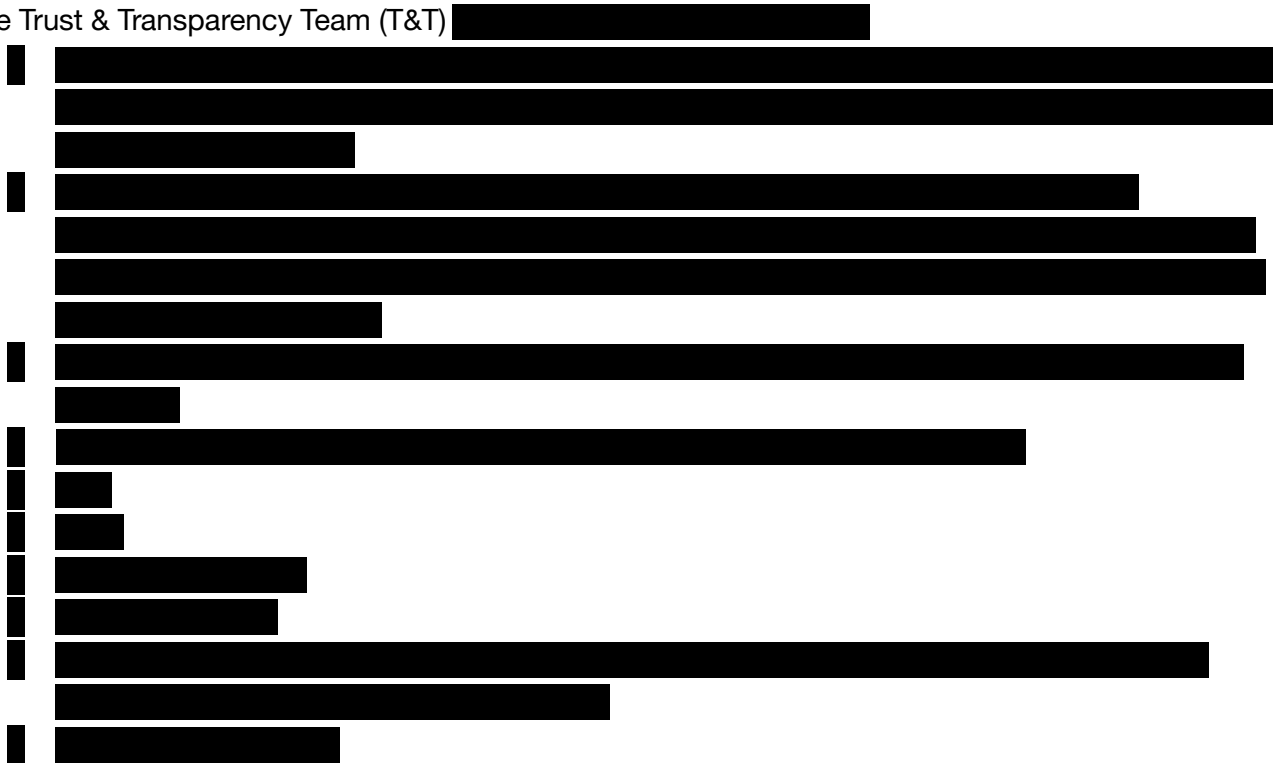
Our Chief Trust Officer participates in our Executive Team or “ELT” group of C-Level managers who help shape the direction of the organisation. This team includes our CEO, Chief Financial Officer, Chief Operating Officer, Chief People Officer, Chief Marketing Officer and Chief Commercial Officer. Having the Chief Trust Officer participate at this level ensures that trust and safety issues are embedded into strategic decisions affecting the entire organisation. It also reflects the importance of trust and safety at Trustpilot and sets the tone for ensuring that these topics are given the consideration they deserve throughout the organisation.

Senior managers (including our VPs and many of their direct reports) also participate in our Global Leadership Group (“GLG”) which is a broad, cross-disciplinary team of leaders spanning all facets of the organisation, who regularly come together to share knowledge and discuss issues, including on trust and transparency-related questions.

*Chief Trust Officer and Trust & Transparency Team*

Trustpilot has an entire “Trust & Transparency Team” dedicated to trust and safety, including user safety on our platform. This cross-disciplinary team encompasses a significant part of our workforce: at the end of December 2021, the team had approximately 88 employees (out of Trustpilot’s 900-strong workforce) including agents, investigators, lawyers, technology experts and communications and training specialists, all supporting Trustpilot’s focus on trust and transparency. In addition, we also outsource various content integrity tasks to allow for flexibility in managing the volume of review activity. The T&T team is located across multiple offices including in the UK, EU and the US. It is led by our Chief Trust Officer, who reports directly to Trustpilot’s CEO.

The Trust & Transparency Team (T&T)



*Content Integrity*



Every reported review is considered by our trained Content Integrity specialists. The first response time to all flagged reviews, in all markets globally, is currently within 48 hours. They are supported by a Fraud & Investigations Team, who consider and act on any particularly complex situations. Both teams have access to powerful tools to help them examine review patterns for further anomalies.

[Redacted]

[Redacted]

Decisions on flagged or reported reviews must follow our processes, and if a business or reviewer thinks that the CI team has made a mistake and disputes a decision, this will be escalated for review by specialists in our Customer Resolutions Team (a sub-team within Content Integrity), via our formalised Decisions Dispute Process.

[Redacted]

*Quality Assurance & Knowledge*

[Redacted]

Our quality assurance process drives the performance of our internal and partner CI agents. Quality specialists regularly evaluate performance elements to identify gaps that are then addressed through agent coaching. Evaluation results are also subject to root-cause and trend analysis and reported to teams for corrective actions via our quality system.

[Redacted]

[Redacted]

#### 4.2 Multi-disciplinary collaboration

T&T works closely with other teams across the business. This includes R&D and our product teams designing the product, data scientists and analysts monitoring how it is running, and customer-facing teams who handle queries and concerns from consumers and businesses using the platform. (See also information about our [Redacted] above at **3.4**).

[Redacted]

Overall, T&T is a main contributor to strategic and project goals, including identifying platform improvement tasks to enhance user safety and satisfaction. As an organisational collaborator, T&T provides a strong system of checks and balances for changes to our features and functionalities proposed by our R&D teams.

### **Terms of service and policy statements**

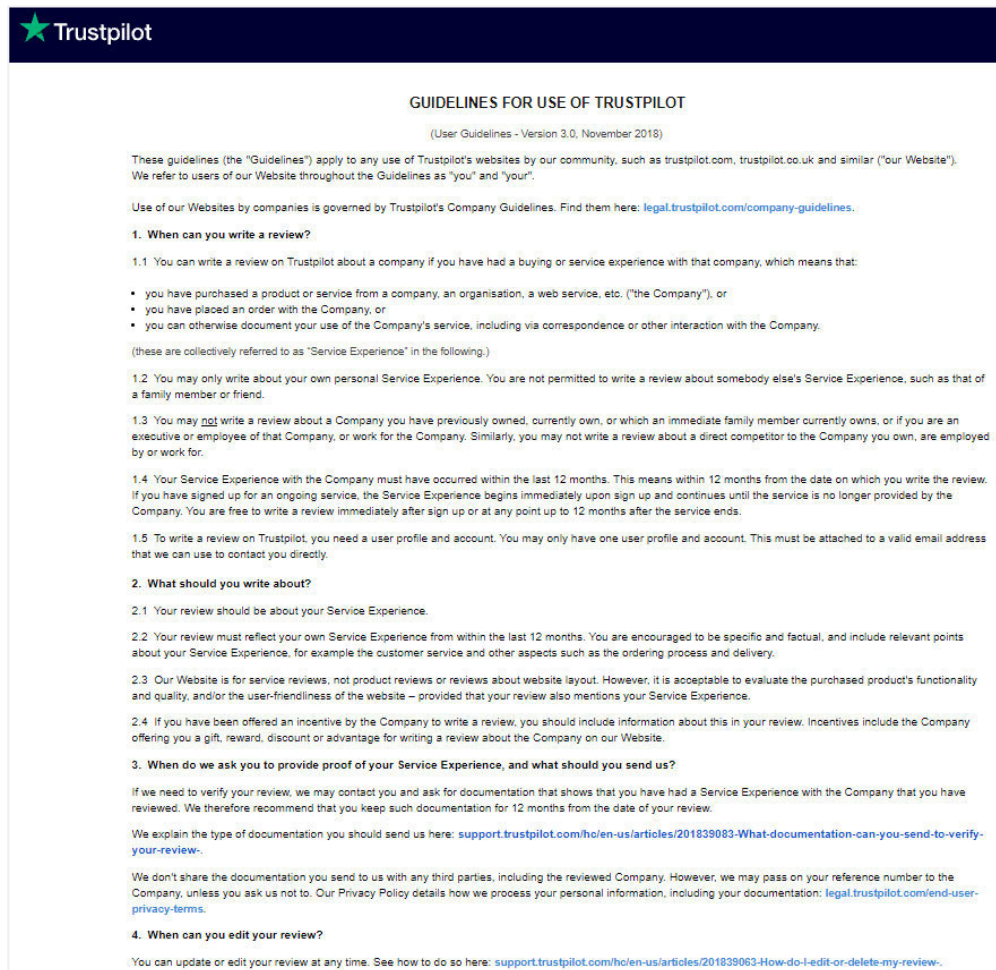
#### **Q5. What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?**

In building trust on our platform, it's vital that both users and businesses can understand our terms of service and public policy statements. To this end, we have done considerable work to make these clear and accessible for users, and are always looking at ways to improve them.

#### 5.1 Clear language

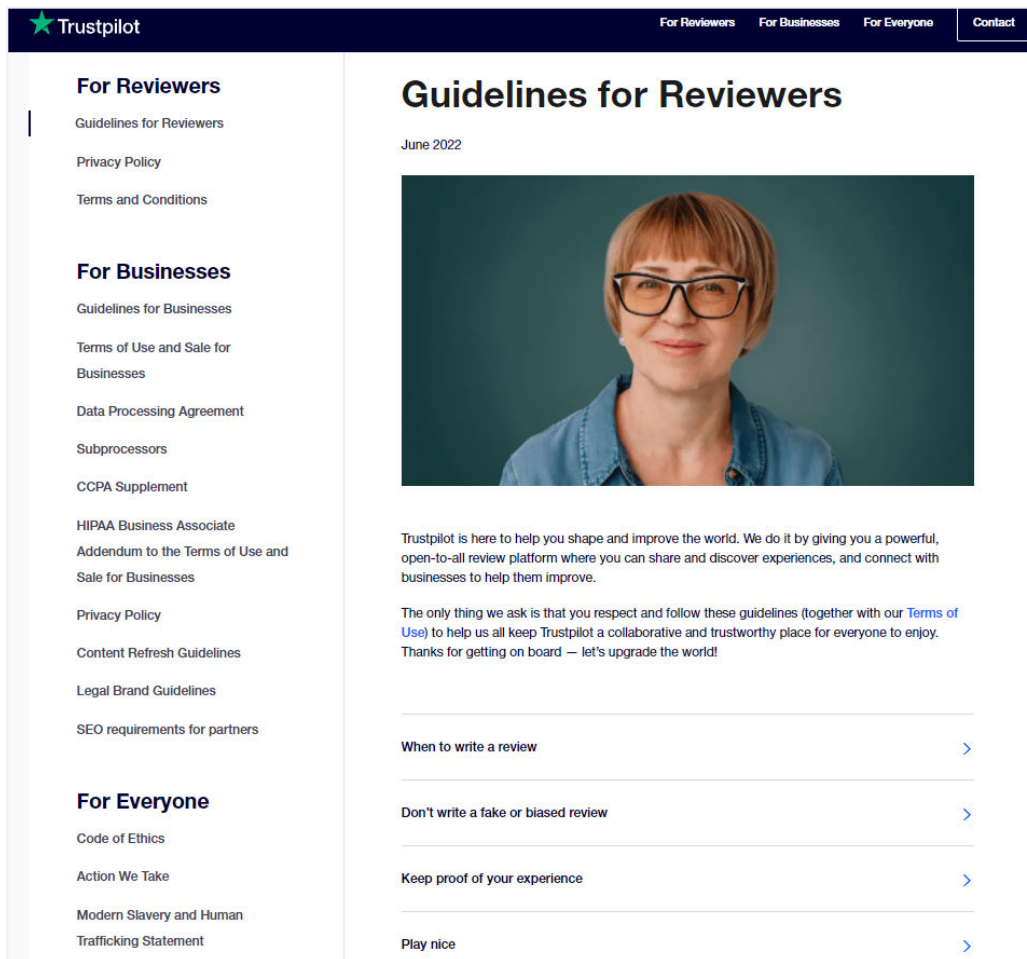
In our experience, terms and policy statements tend to receive more engagement with users where they are written in plain language without legal or technical jargon, where they are presented as useful educational material available in different formats, and where they are made easily accessible at relevant parts of the user journey.

To this end, in 2019/2020 we undertook a multi-disciplinary task within Trustpilot to simplify our guidelines and rewrite them in plain language, as well as present them in a user-friendly format. As of 2019, our guidelines were text-heavy, written in small font and contained some legal jargon (see screenshot of our User Guidelines at **Figure 5.1**, below). Our aim in rewriting the guidelines was to improve both our sets of guidelines for businesses and for consumers to be friendly, easily understandable, engaging and shorter. The scope of our review encompassed looking at the text and tone, design/layout, presentation and location/hosting of the guidelines. Research inputs included evidence on which parts of our guidelines were most often overlooked or breached. We focused in particular on making sure these overlooked elements were as clearly articulated as possible.



**Figure 5.1 (above):** Screenshot of Trustpilot's guidelines for reviewers in March 2019, as hosted on Trustpilot.com.

In May 2020, we launched the resulting text and design (see screenshot at Figure 5.2). This was presented as an accordion-style page with bite-sized pieces of content that pop out under each heading when it is clicked on, allowing the reader to select which parts of the guidelines they want more detail on.



**Figure 5.2 (above):** Screenshot of current Guidelines for Reviewers on Trustpilot’s website.

In the first half of 2022, we have updated our [Terms of Use and Sale for Businesses](#) to provide more clarity about what constitutes misuse of our services, and also revised our [Terms and Conditions](#) for reviewers which include rules about what is permitted versus prohibited is using the platform.

### 5.2 Accessible during the user journey

Overall, we strive to find the right balance between providing digestible, readable and useful information versus inserting too much complexity into the user journey, which will ultimately lead to information overload or confusion on the part of our users. We continue to increase our understanding of how users experience Trustpilot to help reduce inadvertent misuse in particular, and to guide correct use. This means taking a broader view than our terms and looking at the bigger picture — the overall user journey and the information about terms and policies provided at each step of the way.

Alongside improving the tone, look and feel of our guidelines to help users understand and engage with them in 2019/2020 and updating our terms (as outlined above), we have also investigated where in the user journey information about our rules would be most relevant or helpful. This is because we know that placing relevant information along the user journey so that people can access it *at the point at which they need it most* is more effective in practice than asking people to read terms without having a specific focus or practical use. We have also observed that bite-sized pieces of information tend to be more digestible than large inputs. That is, transparency isn’t the same as



clarity, since especially online, too much information tends to overload people and they don't read it, or are unable to take it in.

One key focus area is the information available to users while they are writing reviews on Trustpilot. For example, in recent years we have made changes to the wording and to the layout of our review form to reinforce terms and content policies. While the old review submission form asked reviewers to "Voice your opinion", the new form reinforced that *reviews should be based on a recent personal experience with the relevant business*, and therefore state "Rate your experience" and "Tell us about your recent experience".


The ghost text in the review box was changed from "Share your honest experience [...]", to broaden the focus from just honesty to providing "honest, helpful and constructive" feedback. Honesty is an important element in reviews, but it is not the only requirement set out in our terms and policies, and (except for limited cases of defamatory material) it cannot justify content that otherwise breaches our terms.


We also added next to the review form a link to our Guidelines for Reviewers, and a link to a more user-friendly article setting out tips for [how to write a useful review](#) that reinforces key parts of the guidelines. This article has since become one of our most popular Help Center articles.


[REDACTED]  
[REDACTED] our Help Center and its educational articles is a format that tends to get a high level of user engagement. This is especially so where links to such information are placed at relevant parts along the user journey. Therefore, we also use this resource to highlight key points from our terms, at the point at which it becomes relevant. All of our Help Center articles are designed to be easy to read and digest, and as much as possible written in layman's terms.

We have also added a declaration to be agreed to by reviewers before submitting their review, and this includes a link to information about who is eligible to leave a review (in case users want to check before posting), and also reminds them that, as set out in our guidelines, incentivised reviews are not permitted.

**Voice your opinion! Review Yodel now.**

**Rating**  
 2 stars: Poor – an inadequate experience with a lot of friction.

**Your review**  
Share your honest experience, and help others make better choices.  
  


**Title of your review**  
 

**Enter Tracking Number**  
Your Tracking Number helps Yodel identify you, making it easier for them to reply to your review.

**What is Trustpilot?**  
Trustpilot is a review community. We help consumers everywhere find companies they can trust.

**Share your experiences**  
and be a part of our community. It couldn't be easier - with your Trustpilot account you can write, edit and share your reviews from one convenient location.

**How do I get my Tracking Number?**  
We want to ensure that all customers are able to write a review, without exception. If you do not have your Tracking Number, you can send a request to: [socialmediaescalations@yodel.gnatta.com](mailto:socialmediaescalations@yodel.gnatta.com).

**Figure 5.3 (above):** Screenshot of our old review submission form circa 2018, and prior to making changes for clarity.

Trustpilot

**YODEL** Yodel  
yodel.co.uk

Rate your **recent experience**

★ ★ ★ ★ ★

Tell us about **your experience**

[Read our Guidelines for Reviewers](#)

This is where you write your review. Explain what happened, and **leave out offensive words**. Keep your feedback **honest, helpful, and constructive**.

[How to write a useful review](#)

**Date of experience** ⓘ  
dd/mm/yyyy

**Give your review a title**  
Write the title of your review here.

**Enter Tracking Number (optional)**  
Include your Tracking Number so it's easier for Yodel to identify you and reply to your review.  
Enter Tracking Number here.

I confirm this review is **about my own genuine experience**. I am **eligible to leave this review**, and have not been offered any **incentive or payment to leave a review** for this company.

**Submit review**

**Figure 5.4 (above):** Screenshot of Trustpilot’s current review submission form, with wording that reinforces our terms and policies highlighted in yellow.

To reinforce our terms and policies around flagging, we added educational information within the flagging process (shown below as a series of screenshots at **Figure 8.1**).

We are also engaged in ongoing work to incorporate more education into enforcement activities for misuse of Trustpilot to help consumers and businesses understand the rules and how to use the platform correctly. We are aware that some of the misuse of our platform arises from a lack of understanding about what is permitted by our terms and policies, or how the platform works.







**Q6. How do your terms of service or public policy statements treat illegal content? How are these terms of service maintained and how much resource is dedicated to this?**

Trustpilot has a zero-tolerance approach to illegal content and this is communicated clearly to both our users and businesses.

6.1 Communicating about illegal content and breaches of our policies

We explain how we treat illegal content in our terms of use, our Action We Take policy, user guidelines, Help Center articles, and the messaging we provide to users at various points in the user journey, such as within our notice-and-action (flagging) process. Our rules specifically prohibit illegal content and activities.

Trustpilot's [Guidelines for Reviewers](#) and [Guidelines for Businesses](#) have the same text prohibiting illegal activity:

*"...Play nice*

*We expect you to be a respectful contributor to our platform. So play nice, don't be a jerk. Don't post anything harmful, hateful, discriminatory, defamatory or obscene. And don't lie, bully, blackmail, make threats or do anything illegal..."*

Our [Reviewer Terms and Conditions](#) specifically prohibit uploading illegal user-generated content at point 2.6:

*"Registered Users must not, and must not allow any third party to, publish UGC on the website which:*

*...*

*is unlawful,*

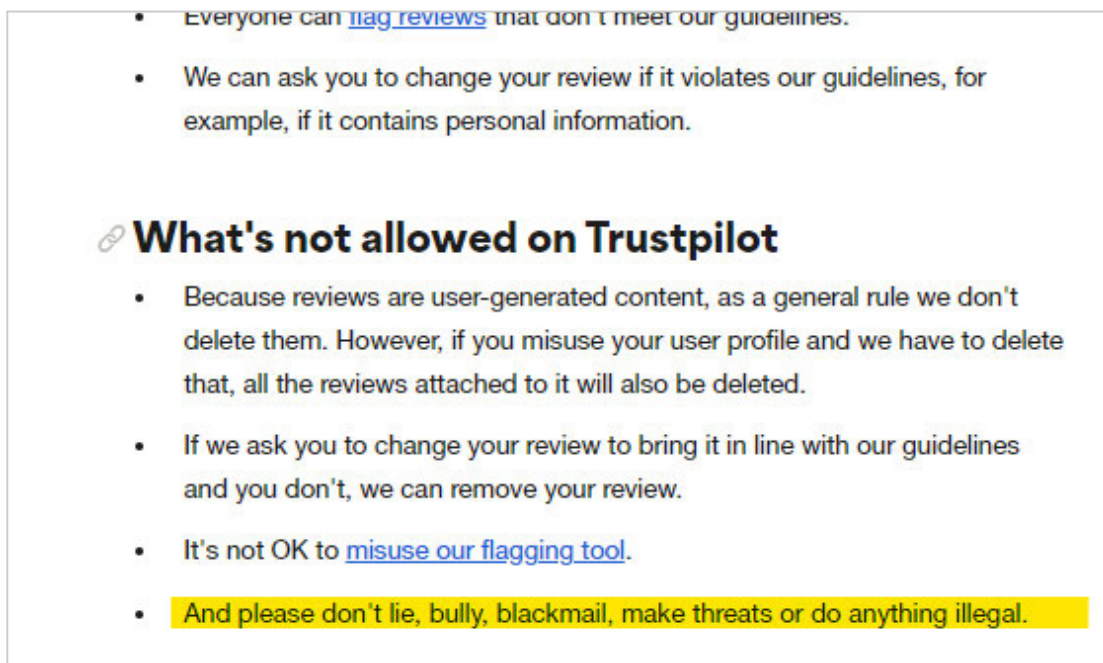
*..."*

Our [Action We Take policy](#) describes the different kinds of misuse and misbehaviour we encounter on the platform, the action we take to prevent it, and the consequences for anyone who breaches our guidelines, including for illegal activity. This policy clearly states what action we will take against content such as fake reviews, misleading reviews that are collected in an unfair way, reviews that are reported for breaching our guidelines (including for being harmful or illegal), which businesses we remove as “bad-fit businesses” and what we will do to counter misuse or abuse of our service, including misuse of our reporting tools, brand, or abuse of our online community.

We also have [information on our website](#), in the section entitled “Combating fake reviews”, that outlines the proactive measures we take to identify and remove fabricated reviews.

Additionally, a Help Center article summarises our Guidelines for Reviewers, called “[Quick guide to Trustpilot's Guidelines for Reviewers](#)”. This article includes tips for staying within the rules and four bullets on what’s not allowed, including “do[ing] anything illegal”:

**Figure 6.1 (below):** Partial screenshot of “[Quick guide to Trustpilot's Guidelines for Reviewers](#).”



For businesses, a corresponding Help Center article summarising the guidelines into skimmable bullet points is “[Quick Guide to Trustpilot’s Guidelines for Businesses](#).”

Other relevant Help Center articles may include:

- [How consumers can flag reviews that breach our guidelines](#) - sets out the four categories of reasons for which a review can be flagged by a consumer, and instructions on how to do so, including an overview of the process.
- [How do we handle misuse of the review flagging tool for consumers?](#) - sets out what happens if a consumer flags an online review, provides examples of what we consider misuse of our flagging function by consumers, and lists possible sanctions for misuse.
- [Why was a review removed?](#) - helps businesses understand the reasons why a review may no longer be visible on their business profile page.

## ★ Trustpilot

- [Trustpilot's fraud detection software](#) - explains how our software works to detect fake reviews, when reviews will be removed, how consumers can appeal if they disagree, and how Trustpilot's employees manually oversee and work in partnership with our software.
- [For which reasons can businesses flag service reviews?](#) - sets out in one place all of the reasons for which businesses can flag/report reviews, with tips to help them comply with our guidelines and avoid misuse of the Trustpilot website.
- [What happens if businesses misuse the review flagging tool?](#) - explains what Trustpilot considers misuse of the flagging/reporting function by businesses, and the disciplinary action that will be taken against businesses engaging in misuse.

### 6.2 Maintenance

We devote a considerable amount of resources to communicating about our terms, policies and practices, and why we remove certain content from our platform. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Changes to our terms also need to be incorporated into our processes, and where relevant, the email communications we send out to users about reported reviews. This is coordinated by our Content Integrity Team, with input from relevant stakeholders such as UX writers and legal experts. Our Localisation Team helps with translations and Content Integrity reviews the translations.

As set out at **5.1**, we recently updated our [Terms and Conditions](#) for reviewers, including to make them more understandable to our users, and to ensure we explicitly include all of the relevant information. [REDACTED]

[REDACTED] While we have acted on such content to date, this will be the first time we have conveyed this as explicitly to users.

As stated above, in order to create value for readers of reviews and businesses who use our free and paid services, we need to make sure we're hosting useful review content. This means we want to help users use Trustpilot correctly, and communicating the rules of our platform (including on illegal content that is prohibited) is in our interest.

### 6.3 Reaching and educating different audiences

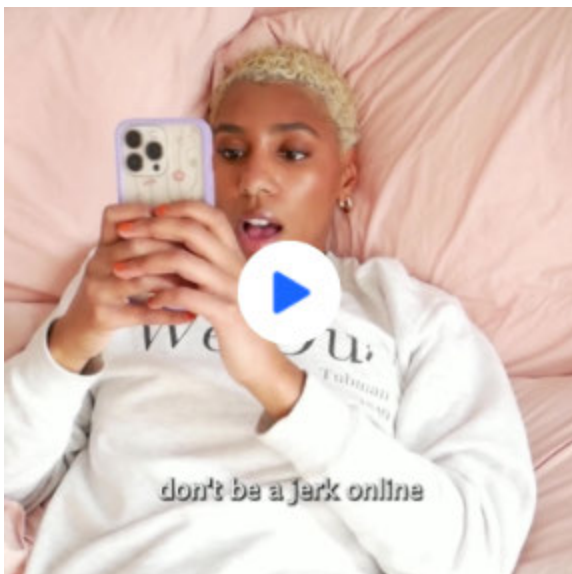
Data we collect on flagging (published in our Transparency Report) and input from our Content Integrity agents (who handle flagged reviews) suggests that we need to take both a reactive and proactive approach to illegal content online. While it is clear that some users have read and understood our terms, others seem to have little idea about what type of content needs to be removed because it is illegal online, or in breach of our terms.

Once reviews are posted, we can reactively remove those that contain illegal content. We observe that some users are more receptive than others to information about their review being removed, or requests from us to edit problematic content. However, we also know that some users will benefit more from up-front information about what is compliant and delivered *before* their reviews are written, rather than information about already-posted reviews that are non-compliant and flagged for removal.

As one example of proactive messaging, we ran a campaign earlier in 2022, called “Helping Hands”, to encourage helpful and useful content on Trustpilot. Our message was a reminder to everyone that their ‘Helping Hands’ can do wonderful things when communicating online, so they should: “...Stop. Take a moment. Use them constructively and make your message count!”

This type of communication aligns to our values of being ‘positively human’, but also around helping people trust each other. We know the power it has when consumers provide constructive feedback to businesses and businesses respond well and act on it.

The approach involved a multitude of different elements, including global research into understanding people’s grudges and whether people could forgive and forget their grievances or regretted acting too hastily. Our lead video content featured ‘Hanz (the hand guru)’: “...before you post online, pause and take a breather; relax.” This was followed up with a drumbeat of regular content in varying languages. We tapped into the world of influencers and creators online to help us take the core message and develop content their audiences love. For example, Hebs — a US influencer who writes songs — encapsulates the core message in her own unique way, in a song “Please don’t be a jerk online”, reminding people that their words can have damage.



**Figure 6.2 (above):** Screenshot of one of our campaign video posts, with influencer Hebs singing “Please don’t be a jerk online.”

In this way, we consider what types of messages will be most effective in order to reach and educate users and help them comply with the rules.

On accessibility more generally, please see Trustpilot's [Accessibility Statement](#).

## Reporting and complaints

### **Q7. What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?**

#### 7.1 Reporting or complaints routes

To flag an individual review for breaching our guidelines via the flagging icon (see **Figure 8.1**, below), users must have a registered account and must be logged in to their Trustpilot profile.

Non-registered users can use the contact form in our Help Center to let us know about a problem.

We ask for an email address as part of the process.

We do not specifically distinguish between children and adults in terms of reporting or complaints mechanisms on our services, but we strive to make the process intuitive and all of the communications as plain language and easy to understand as possible.

#### 7.2 Informing users about reporting and complaints mechanisms

Online services should provide multiple avenues to report non-compliant or illegal content, and should communicate information about this to users in a range of formats. Trustpilot has a flagging mechanism for reviews that breach our guidelines (outlined below at **8.1**), and a contact form for anyone wanting to report a problem. As explained above, we deliver information about this at various places along the user journey, and in different formats such as the more traditional places like our terms, but also in easy-to-read Help Center articles. Overall, users seem to be aware of our reporting mechanisms, and given the volume of reports we receive, they are able to use these to flag reviews.

All of our communications on reporting and complaints mechanisms are designed to be as plain language as possible. At Trustpilot, we involve UX writers as part of the process of making the information on our platform interface user-friendly and understandable. However, it is an ongoing process to refine and improve the educational and guiding information to help people use our tools correctly and effectively. As stated above and in our [2022 Transparency Report](#), we believe we have already seen improved accuracy in the validity rates of reviews reported to us for breaches of our guidelines, and this is likely to have been influenced by our changes to improve education at relevant touchpoints.

Increasing digital literacy and education generally around what is illegal on the internet will likely also be beneficial to users of Trustpilot and others participating online. This will also help to set people's expectations about what types of content warrant removal, versus those that don't. For global platforms operating across multiple jurisdictions and cultures, helping people understand where different services draw the line, and why, is challenging.

**Q8. If your service has reporting or flagging mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?**

8.1 Options for flagging reviews

We provide several options for our reviewer and business communities to flag a review to us at any time if they believe it breaches our guidelines. These flagging options are designed to cover the different types of illegal or harmful content we see in text-based reviews.



The reporting flow, flagging mechanisms and education surrounding these are continuously being reviewed, updated and improved by applying learnings about how consumers and businesses experience these and how effective they are in validly reporting content to us that breaches our guidelines or may be illegal.

We refer to reviews flagged to us as “reported reviews”, and each of these reviews is assessed by a member of our team of experienced Content Integrity specialists. We treat reports of illegal content in reviews particularly seriously: these reviews are immediately hidden from view pending assessment, and we ensure that we prioritise handling of the tickets. Where reports are valid, these reviews stay offline unless or until they are edited to address the problematic content.

Occasionally, the nature of a review suggests that the reviewer has no interest in complying with Trustpilot’s rules and providing useful feedback to a business. In this case, we remove the review, let the reviewer know we have done so and why, remind them that Trustpilot is not the place for this type of content since we want everyone to feel welcome on our site, and link to our guidelines and top tips for writing a review in case they want to change their approach and write a new review.

8.2 Reporting reasons

Reviews can be flagged by consumers (using the flag icon shown alongside every review — see below at **Figure 8.1**) or by businesses from their business account. Consumers and businesses must have a registered account and must be logged in to do so.

Reviews can be reported for one of the following reasons/where they contain:

- harmful or illegal content,
- personal information,
- advertising or promotional content, or
- are not based on a genuine experience (for example, they are spam), or
- are written about a different business than the one listed (only businesses can report for this reason).

Under the category “harmful or illegal content”, reviews can be reported for containing:

# ★ Trustpilot

- hate speech or discrimination,
- terrorism,
- threats or violence, or
- obscenity.

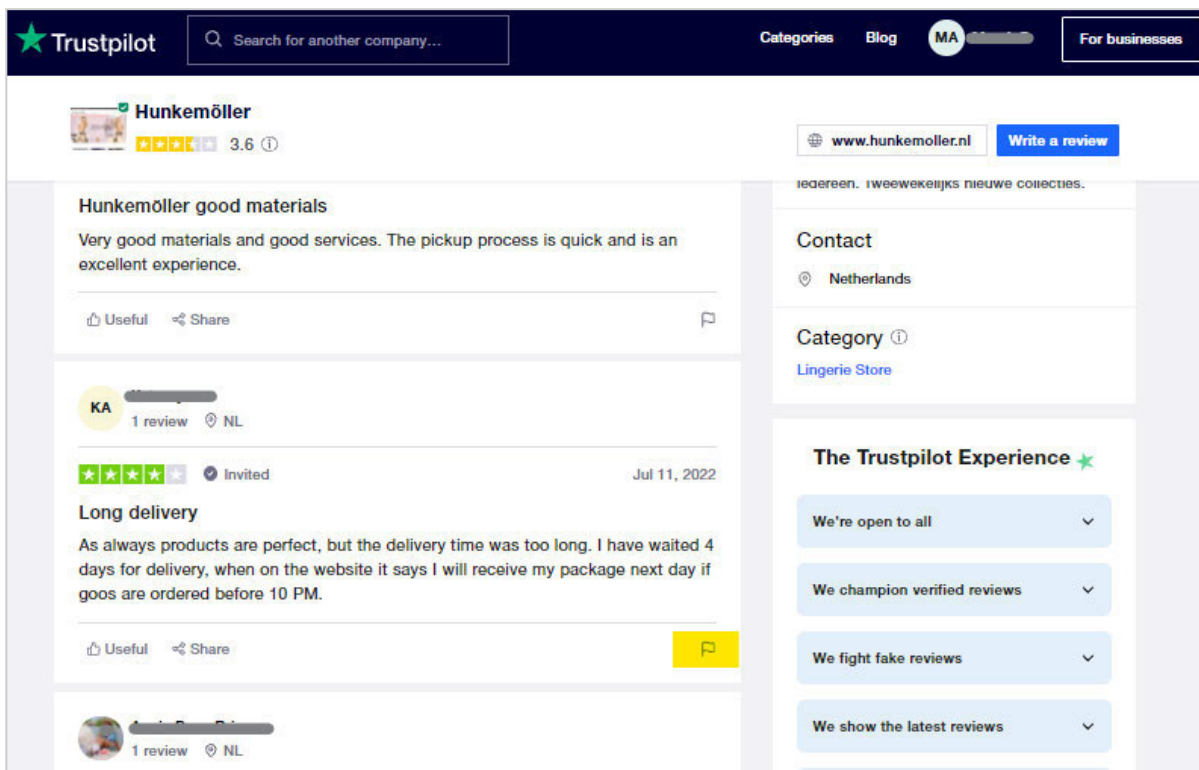
Additionally, businesses can report reviews for defamatory content. In line with our commitment to freedom of expression, we set a high bar for removal of content under this reason.

We explain the reasons for which reviews can be flagged in two Help Center articles, [for businesses](#) and [for consumers](#).

## 8.3 The flagging flow for consumers

To flag an individual review, consumers must click on the flag icon, highlighted yellow in the screenshot. This is placed near every displayed review in order to be user-friendly and accessible.

**Figure 8.1 (below):** The flagging process for consumers (via a registered account)



Consumers must confirm that they are not acting on behalf of the business who is reviewed – business users should flag individual reviews via their business account. (This ensures we deliver the right information to businesses during the flagging process, and also preserves the accuracy of the information we provide to consumers about how each business uses Trustpilot, and how many reviews they flag.)

**Do you think there's a problem with this review?** ✕

You can use this flagging process if you're a consumer. [Read more.](#)

If you're from [www.hunkemoller.nl](http://www.hunkemoller.nl) and want to flag this review, please use your [business account](#).

I'm not from [www.hunkemoller.nl](http://www.hunkemoller.nl)

**Next**

**Want to flag this review?** ✕

**Please choose a reason**

- Harmful or illegal
- Personal information
- Advertising or promotional
- Not based on a genuine experience

**Back** **Next**

To flag the review, the consumer must select a reason from five different options. For harmful or illegal content, they will be asked to choose a sub-reason from four options:

**Want to flag this review?** ✕

**It is harmful or illegal because it contains:**

- Hate speech or discrimination
- Terrorism
- Threats or violence
- Obscenity

**Back** **Next**

After users have selected a reporting sub-reason, we provide some brief information about what types of content we remove, or do not remove. This sets expectations and is intended to prevent consumers from flagging reviews simply because they disagree with the statements made in a review or find it critical.



**Want to flag this review for threats or violence?** ✕

**We can remove content that has, for example:**

- Calls to target people or businesses with abuse or harassment online
- Serious threats or threats of violence
- Behavior that urges offline action such as physical harassment or destruction of property

**We don't remove content just because:**

- You dislike or disagree with it
- It criticizes someone

Users are then asked to highlight the relevant section of the text that breaches the rules. We introduced this feature so that our Content Integrity agents can quickly focus on the problematic text. (Note: In the below example, the text that is highlighted would not be assessed as valid “threats or violence.” The example is used to illustrate the process only.)

**Want to flag this review for threats or violence?** ✕

Select the **threats or violence**

**Long delivery**

As always products are perfect, but the delivery time was too long. I have waited 4 days for delivery, when on the website it says I will receive my package next day if **goos are ordered** before 10 PM.

Before submitting their flagging notice, users must confirm that their submission is true and correct, and provide their email address. We have added this step to ensure a level of accountability for flagging, and to limit the number of unfounded reports we receive.

**Want to flag this review for threats or violence?** ✕

I confirm that the information I've provided here is true and correct.

My Email

**Thanks for flagging this review.** ✕

Feedback like yours helps us keep our review platform trustworthy.

#### 8.4 Flagging costs



Tickets flagged for any of our harmful or illegal sub-reasons are hidden immediately from view and prioritised to be handled as quickly as possible by our Content Integrity agents, and within the Service Level Agreements (SLAs) specified. The first response time to all flagged reviews, in all markets globally, is currently within 48 hours.

#### 8.5 Trusted flaggers

We do not currently work with designated trusted flaggers. However, anyone can let us know about a review or problematic content on our website.

As stated above, it is the *type* of report that determines the urgency with which the ticket is handled. Our triage system prioritises reviews that are flagged to us as harmful or illegal, since these are deemed to be most serious and are temporarily hidden pending assessment.

#### 8.6 Complaints and whistleblowers

We also have an additional “whistleblower” function for people to confidentially report any problems to us — and this can include reporting illegal content to us. Consumers, businesses or others can select this option when they access our [contact form](#) through our Help Center.

Users can either navigate to our Help Center via the header (for registered users, under the username, select “Help”), or via the Trustpilot footer (click “Support Center”).

**Figure 8.2 (below):** The flagging process for consumers (via a registered account): How to access our Help Center from a profile page displaying reviews about a business.

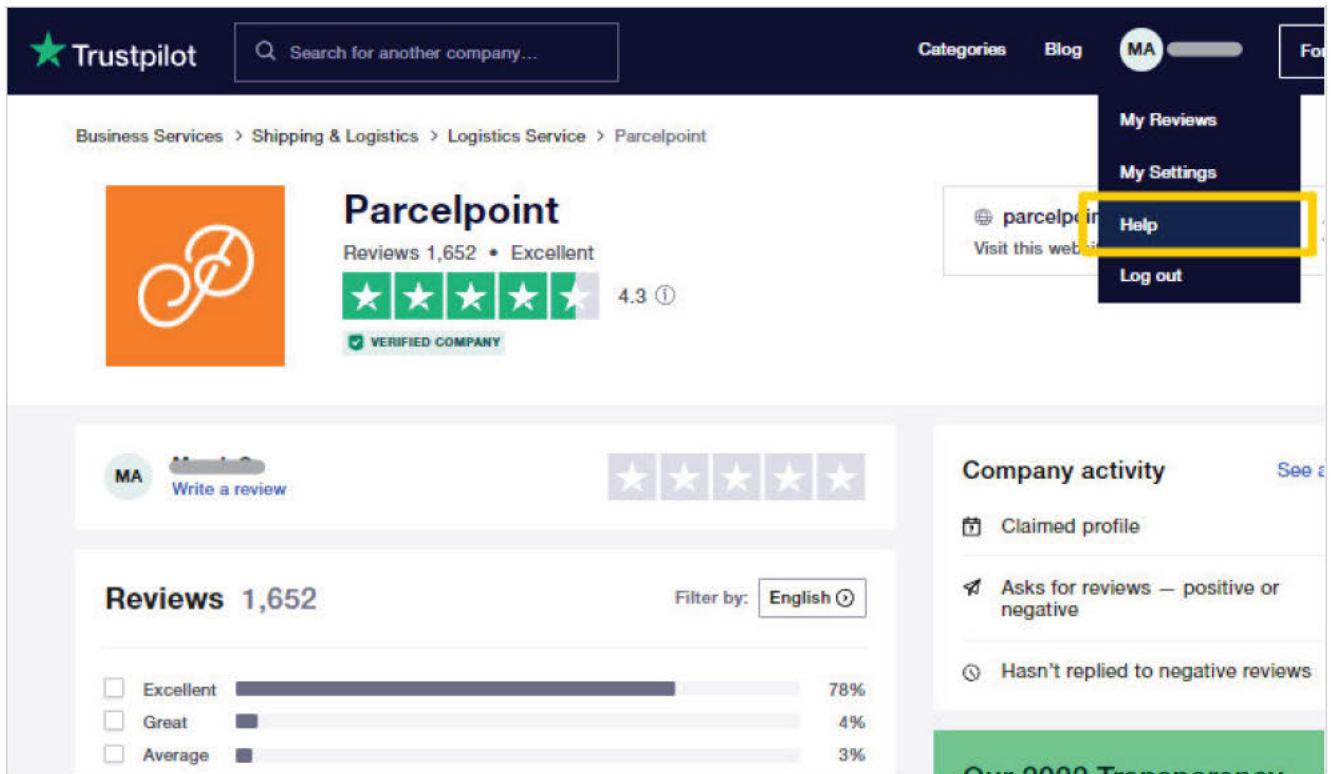
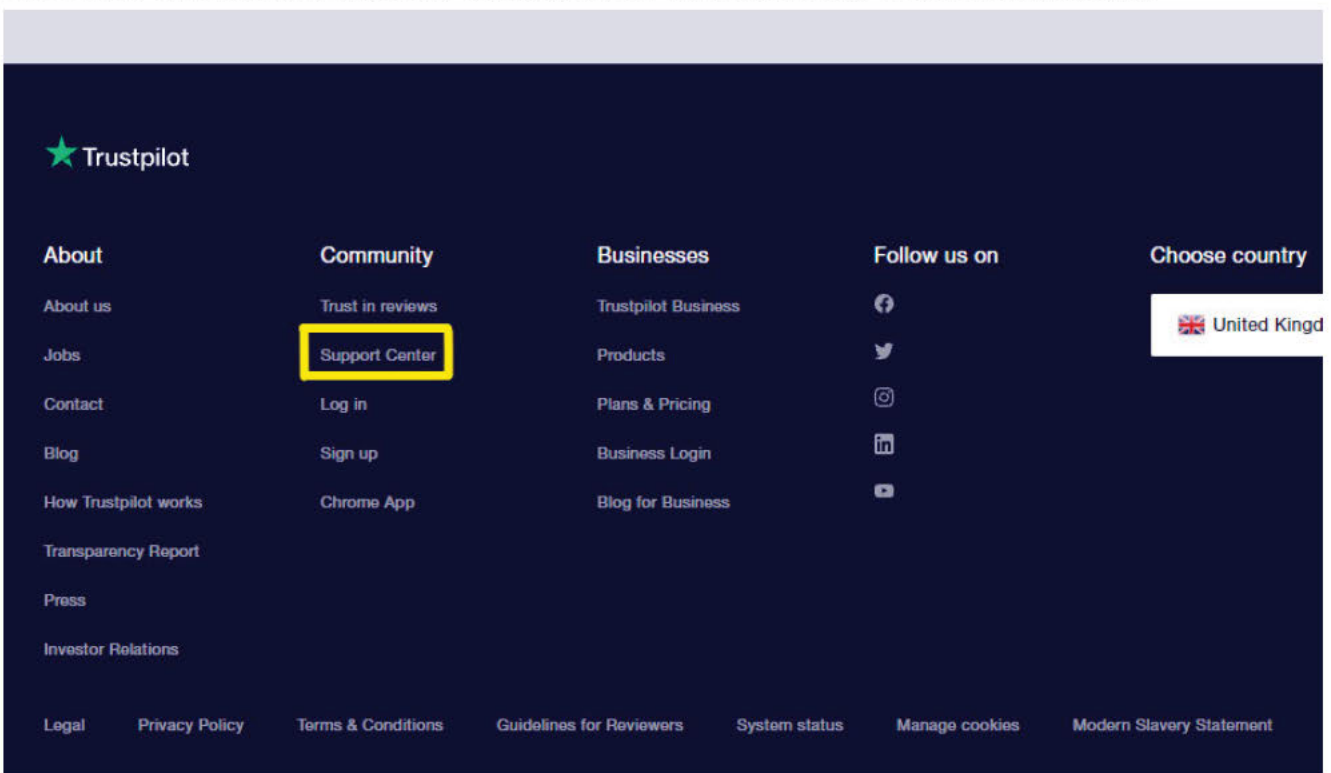


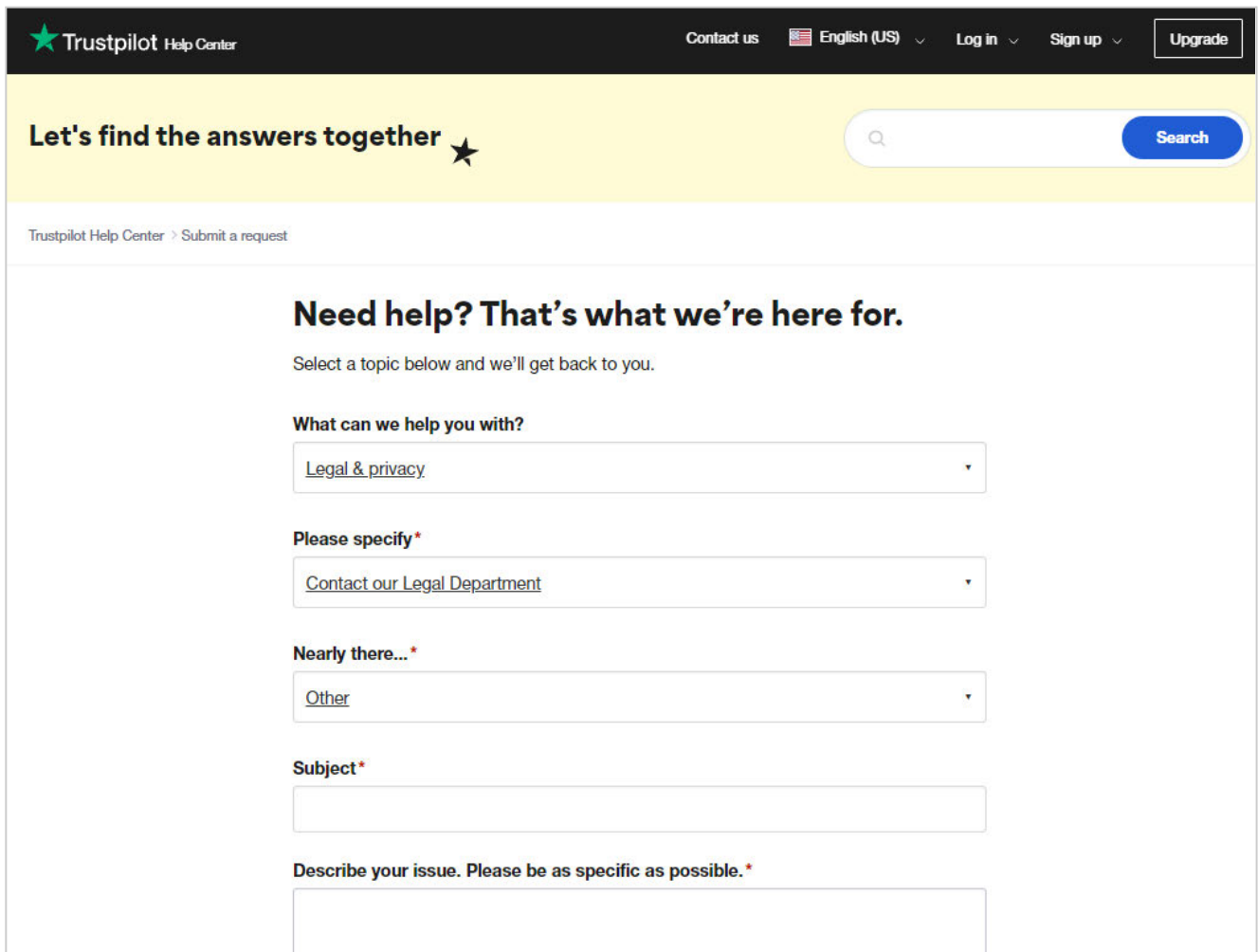
Figure 8.3 (below): Access to Trustpilot’s Help Center from the footer section of Trustpilot’s homepage. Our Help Center was previously called our “Support Center” and this wording will be updated shortly.



To arrive at the form, users click on “Contact us”. It is not necessary to have a registered account with Trustpilot, or to be logged in, when using the form. Anyone can use the form to let us know about a problem.

Our contact form has several broad options to choose between, and this ensures we prioritise the query or complaint correctly and send it to the right team to address. The choices are: Reviews, Technical support, Legal & privacy, Whistleblower, or Sales & pricing. These topics are selected on the basis of data about the kinds of submissions Trustpilot has received historically.

Users can choose from the drop-down list of pre-approved topics, or choose “Other” where none are suitable.



The screenshot shows the Trustpilot Help Center interface. At the top, there is a navigation bar with the Trustpilot logo, 'Help Center', and links for 'Contact us', 'English (US)', 'Log in', 'Sign up', and 'Upgrade'. Below this is a yellow banner with the text 'Let's find the answers together' and a search bar. The main content area is titled 'Trustpilot Help Center > Submit a request' and features a heading 'Need help? That's what we're here for.' followed by the instruction 'Select a topic below and we'll get back to you.' The form consists of several fields: a dropdown menu for 'What can we help you with?' (selected: 'Legal & privacy'), a dropdown menu for 'Please specify\*' (selected: 'Contact our Legal Department'), a dropdown menu for 'Nearly there...\*' (selected: 'Other'), a text input field for 'Subject\*', and a larger text input field for 'Describe your issue. Please be as specific as possible.\*'.

It is possible for non-registered users to flag reviews via this form. This is not the preferred method for flagging since it relies on the user correctly identifying the problematic review by copy pasting the URL, but it does allow an additional route for non-registered users. Submissions via this form will create a ticket for our Content Integrity Team to handle.

Trustpilot Help Center [Submit a request](#)

**Let's find the answers together** ★

Search [Search]

**Need help? That's what we're here for.**

Select a topic below and we'll get back to you.

**What can we help you with?**

Reviews

**Help us direct you to the right person \***

I am a reviewer

**Select a topic \***

Flag a problematic review

**Please specify \***

How do I flag a problematic review?

**Subject \***

The review has terrorist content

All of the flagging reports and submissions made via our content form are handled [REDACTED] and we therefore have a written record that is saved and stored in accordance with our privacy and data retention policies.

See our answer to **Q9** for details about our Decision Dispute Process, which allows users to appeal decisions made by our Content Integrity Team.

See above, at **2.4**, for data about the validity rate of flagged reviews. We believe that well-placed information increases the effectiveness of our flagging mechanism by helping users understand when to flag.

**Q9. If your service has a complaints mechanism in place, how are these processes designed and maintained?**

9.1 Complaints about a decision

As explained at **Q8**, users can complain about illegal content on our services by:

- using the reporting mechanism to flag individual reviews, (for registered users only) or



- getting in touch via our Contact Form to let us know about a problem (for registered or non-registered users).

In this answer to **Q9**, we focus on our mechanism for appealing decisions made by the Content Integrity Team *subsequent to* and following on from one of the above processes.

## 9.2 Trustpilot's Decision Dispute Process

Our Content Integrity Team works hard to ensure that we treat both consumers and businesses fairly. However, we recognise that our team is only human, mistakes can happen, and that it can be frustrating to people who have spent their time writing a review, or sharing information with us that they feel we have overlooked in some way. Taking these concerns on board and with a quality-oriented and customer service driven approach in mind, we recently introduced a Decision Dispute Process (DDP) to provide a formal, specific and dedicated route for consumers and businesses to let us know where they feel we've missed something in our initial decision.

Our aim is to ensure that every individual who has an interaction with our Content Integrity Team walks away from that interaction feeling listened to, understood and respected, and with a clear understanding of not just the decision we reached itself, but why we've made that decision. This is central to ensuring that we have the trust of consumers and businesses using our platform. The feedback we receive via this process also provides us with an opportunity to learn and improve.

Our DDP is outlined in our Help Center article, "[How can I dispute a decision made by the Content Integrity Team?](#)" Every complaint submitted via the Decision Dispute Process form is handled by a dedicated team of specialists, to make sure reviewers and businesses are heard and have confidence in the decisions we're making.

An electronic form needs to be filled out to dispute any decision made by the Content Integrity Team within the past six months, including where a review was removed, if a user account has been blocked, the outcome of a flagged review, receipt of a warning or cease and desist letter, or where Trustpilot adds a Consumer Warning or Consumer Alert to a business's profile. The form is easily accessible via our website, and is included as a link in our email messaging from Content Integrity agents where the initial decision is conveyed.

Complainants who submit a dispute immediately receive an email to acknowledge that their information has been received, which also allocates them a ticket number for future reference. A specialist agent from our Customer Resolutions Team (a sub-group within Content Integrity) investigates each case in line with our guidelines to determine whether the right decision was made. They'll respond to the complainant with a final outcome within two working days. If that's not possible, or more information is needed, complainants will be kept informed about the likely resolution time.

## 9.3 Designing DDP

The design of the DDP process incorporated input from a wide range of teams across Trustpilot, including those with particular input on how consumers and businesses experience the Trustpilot

platform, and the challenges they face around decisions made. The aim was to design and deliver a DDP that is fair and accessible to all of our users — both business and consumers.

Multi-disciplinary workshops were carried out to understand the key parameters and requirements for the concept; to gather input on the practical aspects of the process, including when, how complaints should be submitted, what should be included in the process, who should be involved, and where the DDP will work to meet user needs; and develop solutions with assistance from our engineers, communications experts, privacy specialists, customer success teams, and many more.

#### 9.4 The DDP form

See below for two screenshots of our current DDP form. These show different options.

The screenshot shows the Trustpilot Help Center interface. At the top, there is a navigation bar with the Trustpilot logo, 'Help Center', and links for 'Contact us', 'English (US)', 'Log in', 'Sign up', and an 'Upgrade' button. Below the navigation bar is a yellow banner with the text 'Let's find the answers together' and a search bar. The main content area is titled 'Dispute a decision made by the Content Integrity Team'. Below the title is a paragraph explaining the process: 'If you'd like to dispute a decision made by the Content Integrity Team, please fill out this form and our Customer Resolution Team will review your dispute and respond within 2 working days.' The form consists of several fields: 'Your email address\*', 'Your name\*', 'Are you a reviewer or a business?\*' (a dropdown menu), 'Please give us as much information as possible about why you are disputing this decision\*' (a large text area), and 'Attachments — Upload a screenshot of the issue you're having, or the site content you'd like help with.' (a file upload area).



## Dispute a decision made by the Content Integrity Team

If you'd like to dispute a decision made by the Content Integrity Team, please fill out this form and our Customer Resolution Team will review your dispute and respond within 2 working days.

Your email address \*

Your name \*

Are you a reviewer or a business? \*

Your Trustpilot username \*

What are you disputing? \*

Please enter your existing ticket reference number. This will help us handle your request quickly.

Title of your review

Date of your review

The process distinguishes between different categories of users according to what the problem is. The aim is to direct the query to the right queue.

Complaints can be made about a broad range of topics, from freedom of expression to privacy issues. Trustpilot does not have a functionality to de-prioritise reviews. But in cases when content is taken down proactively by technology, such as our fraud detection software, users will often receive an email about the action and will be invited to respond directly to our Content Integrity Team if they believe an error has occurred. Any disputed decisions will be assessed by a Content Integrity agent. If users disagree with the subsequent decision made by the Content Integrity agent, then this can be appealed via our DDP.

### 9.5 Costs and volume

The cost involved in maintaining this process is

[REDACTED] we therefore have a written record that is saved and stored in accordance with our privacy and data retention policies.

**Q10. What action does your service take in response to reports or complaints?**

See above, at our answers to questions **Q8** and **Q9**.

We take action to assess all flagged reviews, and investigate and address all complaints or reports sent to us via our contact form. We address all appeals against decisions made by our Content Integrity team in accordance with our Decision Dispute Process (DDP) and within the designated SLAs. [REDACTED]

10.1 Overall volume of reported reviews

Globally, in 2021, 532,000 reviews were flagged to us by businesses across all of our reporting reasons, and 110,000 were flagged by consumers across all reporting reasons. See above at **2.4** for figures on reviews flagged to us “harmful or illegal”.

Reviews flagged to us as harmful or illegal are handled by Content Integrity agents in accordance with our processes. Where there is an imminent risk of harm to someone, such as a credible threat of violence, a threatened act of terrorism, or suicide-related content, our process is to alert the relevant authorities. We also cooperate with local or national authorities as necessary and in accordance with our legal obligations, to assist with law enforcement activities that include criminal action against persons who have posted illegal content on Trustpilot.

10.2 Reports from consumers

Of the reviews flagged to us by consumers in 2021, the majority were flagged for “Advertising or promotional” content (which includes spam), or as “Not based on a genuine experience” because they were suspected to be fake and therefore misleading. Only 12,177 were flagged as “harmful or illegal” using the sub-reasons described at **2.4**, and a further 8,814 were flagged for including someone’s personal information in breach of privacy laws.

All of the flagged reviews were assessed by our Content Integrity Team, assisted by our technology and tools. Our records show that 18,065 of the reports were deemed as accurate. Therefore, those reviews were removed, or for harmful or illegal reports (which were already temporarily offline) the reviews stayed offline. This means consumers had a flagging accuracy rate of around 16%, which is an improvement on figures for 2020 but also reflects the subjective nature of reporting, the need for further education about what is considered a breach of our guidelines, and the fact that the general public (unassisted by technology) are not always accurate in spotting whether or not reviews are fake.



As stated above, reviews reported to us for harmful or illegal content are immediately hidden from our platform and prioritised for assessment in accordance with our processes. Where we assess that such content should stay offline, we usually notify the reviewer by email. Cases where we do not are where we consider that the content is a deliberate misuse of our systems and informing the person who has posted it will assist them in further attempts to game our system.

### 10.3 Reports from businesses

Of the reviews flagged to us by businesses globally in 2021, a majority — 431,011 — were flagged for the reason “Not based on a genuine experience”; 64,678 were flagged for “Harmful or illegal” content; 32,870 for “Advertising or promotional” content (including spam); 14,362 were written about a different business; 6,018 were reported for containing someone’s personal information; and a small minority were reported for other problems. This is reflected in our latest [Transparency Report](#).

[REDACTED], which provides us with a written record. Of the reported reviews, we deemed 77.2% as accurate, and either moved those reviews offline or kept such reviews offline (for harmful or illegal content) and notified the reviewers concerned.

As stated above, our SLAs for a first response to tickets involving reported reviews across all markets are 48 hours. We regularly monitor response times and adjust the number of Content Integrity agents handling tickets in response to changes in volume.

### 10.4 Misuse of our flagging tool

Our systems are set up to detect misuse, including malicious or potentially unfounded reporting of reviews. For consumers who misuse our tools, such as by repeatedly flagging the same review, or flagging reviews that don’t breach our guidelines, our usual process is to send a warning email to the user to inform them about the problem(s) and ask them to stop flagging reviews incorrectly. If they persist, or the behaviour is particularly severe and engagement with them is ineffective, we might block access to their account. However, this is only used on rare occasions. We set out our process in a Help Center article “[How do we handle misuse of the review flagging tool for consumers](#)”, and explain the consequences in our [Action We Take policy](#).

For businesses, our systems are designed so that every reported review must be given due consideration and so that it is not possible to flag a large number of reviews in a short space of time. We consider misuse by businesses to include flagging only negative reviews, or repeatedly flagging reviews that in fact comply with our guidelines (for example, to remove compliant one-star or two-star reviews and boost a business’s TrustScore rating).

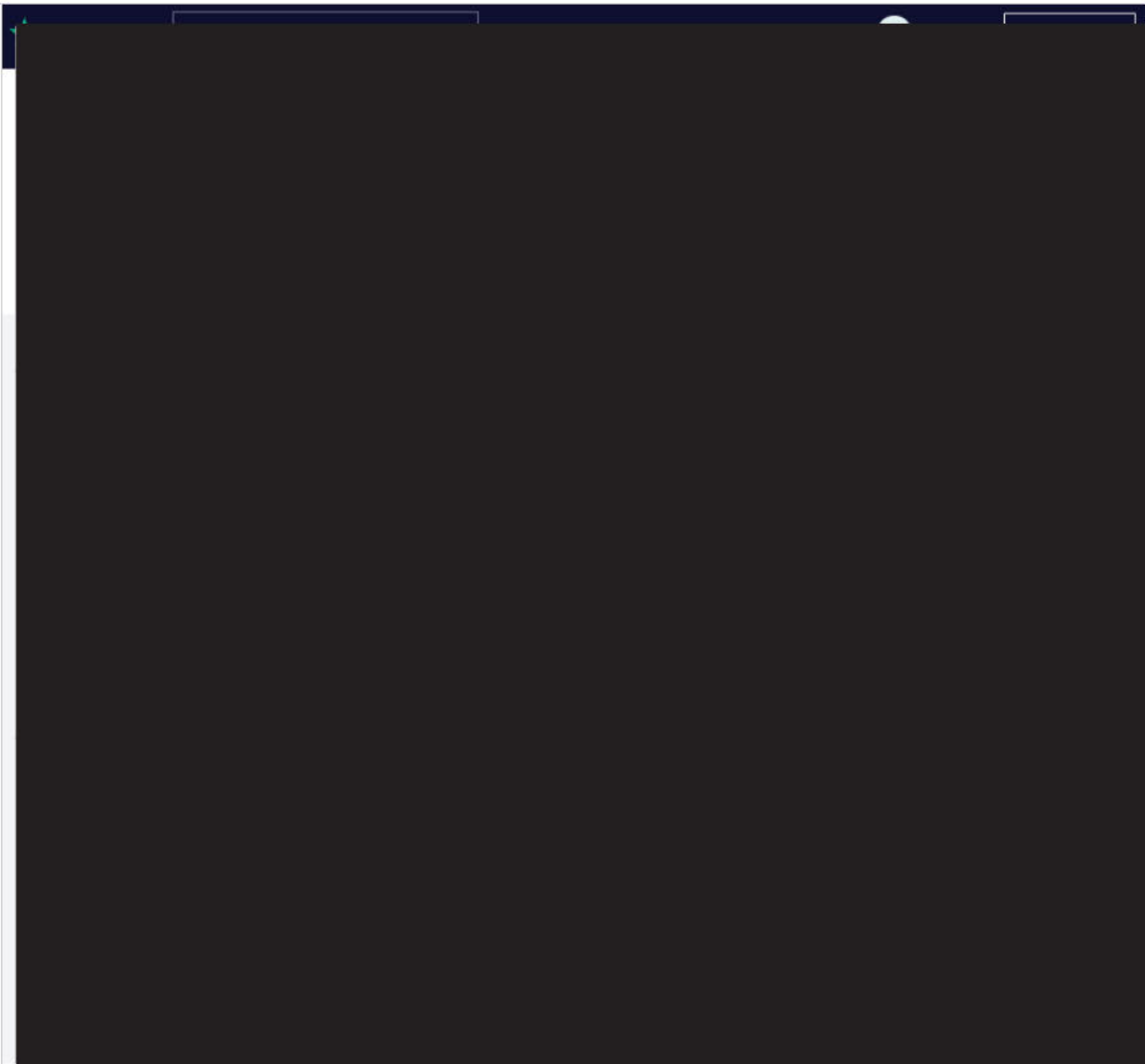
Misuse of our tool is detected by systems and triggers a warning in the business app to discourage the behaviour and ask them to stop. In cases where the behaviour persists, we can take further action such as placing a Consumer Warning on the business’s Trustpilot profile page, limiting access to their business account and, if necessary, taking disciplinary action including terminating any contract they have with us. We explain this in our Help Center article “[What happens if businesses misuse the review flagging tool?](#)” and our [Action We Take policy](#). In 2021, we issued 121,048 warnings to businesses for misuse of the Trustpilot platform. Many of these will have been for



flagging reviews incorrectly or for biased review collection. But we also know that not *all* businesses who receive warnings intentionally breach our guidelines — in fact, many warnings arise as a result of an error or misunderstanding — therefore they indicate an additional point at which to educate businesses and prevent ongoing attempts at misuse of the platform.

To ensure consumers have all the information they need, we display transparent information about the business's flagging history on Trustpilot. This provides an extra incentive for businesses to use our flagging tools correctly. Information is accessible via a link on each business profile page and readers of reviews can easily click through to see this data.

**Figure 10.1 (below):** Readers of reviews can click “See all” on a business's profile page to see how that business uses Trustpilot, and how many reviews it has flagged within a one-year period.



Readers of reviews can click through to a new page via the “See all” link in the “Company activity” section to see how each business uses Trustpilot, including whether or not they have flagged any reviews within the last 12 months.

We provide this information in an effort to be as transparent as possible for consumers reading reviews, and to ensure that they have an overall picture of the reviews a business has received, even if some of those have been removed for breaches of our guidelines.



The above screenshot shows that the business [REDACTED] has flagged two one-star reviews to Trustpilot for breaching our guidelines within the last 12 months, and these reviews have been removed from Trustpilot. (Note: This example was selected at random to illustrate the features of our platform).

## Moderation

**Q11. Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?**

There is considerable complexity in balancing people's different competing rights, while still meeting responsibilities at scale in line with requirements across different jurisdictions. Our efforts to refine and improve our moderation processes and our policies and processes informing them are ongoing, and we incorporate and take into account user feedback in this area.



As stated above, we consider that education of consumers and businesses is an important part of preserving user safety and preventing some types of illegal content being posted online. In practice, we observe that there is a lack of understanding about what is illegal and what rules or standards should be applied to assess content. Providing transparent information about our rules and processes is key to helping everyone understand when we will take action and why.

We believe that there is potential for greater use of automated systems to help with moderation. However, the training, development and fine-tuning of such systems to ensure a high level of accuracy that limits the risk of over-removal or bias is considerable and time-consuming. Many of Trustpilot’s current moderation systems are primarily human-lead and supported by our automated tools.

Unfortunately, the current discourse around high-volume use of automated moderation tools on the largest platforms reinforces both consumers and businesses sometimes having a suspicious approach to decisions to remove content. This can lead users to assume that robots rather than humans have assessed their content, even if this is not the case. Engaging with users in a useful way to change their behaviour can be more difficult if they assume that they are having a conversation with an automated system instead of a human. To try and counter this misconception, and provide transparency to users, we include clear statements in messages triggered by our automated systems to indicate that these are decisions made by robots, whereas our Content Integrity agents sign off email messages written to users with their first names to indicate that a human has handled their query.

We are currently working on a number of projects to both investigate how we can create and use accurate technology to assist us in scaling, and how we can better differentiate these technologies from our human-lead activities by instilling a positive customer experience that is helpful, educational, and encourages confidence in our processes.

**Q12. What automated moderation systems do you have in place around illegal content?**

We have an automated filter that screens for profanities and obscene language at the point at which a review is submitted. [REDACTED]

12.1 Hash-matching

Trustpilot reviews hosted on our consumer site are text-based and character limited, so opportunities for using a review to distribute illegal content are limited as compared with other platforms. We do not use hash-matching to identify illegal content in reviews.

[REDACTED]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted text]

[Redacted]

[Redacted]

**Q13. How do you use human moderators to identify and assess illegal content?**

As stated above, we use human moderators to assess content that has been flagged to us by consumers or businesses, or by our systems. This includes content flagged as harmful or illegal.

[Redacted]

[Redacted]

See also our answers at 4.1, above.

**Actioning content and sanctioning users**

**Q14. How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?**

Our [Action We Take policy](#) details when Trustpilot will restrict access to user accounts, and when we will remove content from our platform.



**Q15. In what instances is illegal content removed from your service?**

Our [Action We Take policy](#) details when Trustpilot will restrict access to user accounts, and when we will remove content from our platform. See also our answers at **6.1** and **8.1** above.

As set out in our Help Center articles, such as “[For which reasons can businesses flag service reviews](#),” we can remove content for being harmful or illegal for the following reasons:

- Hate speech or discrimination: content that attacks or uses derogatory or discriminatory language with reference to a person or a group on the basis of protected personal characteristics, e.g. based on their religion, ethnicity, nationality, race, color, descent, gender.
- Terrorism: content that intends to radicalize people, spreads harmful propaganda, mobilizes support for terrorist activities, or relates to violent extremism.
- Threats or violence: content that contains serious threats or threats of violence.
- Obscenity: content that contains graphic depictions of acts that are violent, sexual, or gory.
- Defamation: content that is likely to cause serious harm to someone’s reputation or serious financial loss to a business.

We don't remove content just because readers of reviews dislike or disagree with it, it criticises a business, or includes swear words. Most of the content we remove for the above reasons is not only a breach of our guidelines, but will also be illegal under UK laws.

We also remove personal information such as employee names, phone numbers, residential addresses, and email addresses (in accordance with privacy laws).

We also remove content if it is fake and/or misleading: for example, the reviewer wasn't eligible to write a review about a business because they are a current employee or a competitor, or have been paid to write a fake review. This content will often, but not always, be illegal.

**Q16. Do you use other tools to reduce the visibility and impact of illegal content?**

No. As explained, we use an automated language filter to prevent obscene language being posted,

[REDACTED]

[REDACTED] Aside from that, we assess reviews reported to us for illegal content, and reports that are assessed as valid result in the relevant reviews staying offline.

**Q17. What other sanctions or disincentives do you employ against users who post illegal content?**

17.1 Sanctions

Our [Action We Take policy](#) details when Trustpilot will restrict access to user accounts. Users are informed about sanctions unless this information will aid them in gaming our systems.

Given the nature of our platform, many of our sanctions target businesses, and reflect our focus on protecting the platform against fake reviews. These include manual and automated warnings, formal



notices against businesses creating fake or misleading reviews, terminating a paid subscription with a business, placing Consumer Alerts and Consumer Warnings on business profile pages. However, we also block user accounts where a reviewer repeatedly breaches our guidelines, such as where they post illegal content, such as blatantly racist remarks, or clearly threaten businesses. In such cases, we can temporarily suspend or block access to the Trustpilot user account, and the reviewer is always informed about our action. Instances where we have used this sanction are rare. [REDACTED]

[REDACTED]

17.2 Safeguards to ensure users are sanctioned consistently and fairly

Trustpilot’s foundational principles help embed fairness into our culture and operations. The principles include:

- Neutral: Trustpilot is a neutral platform that allows consumers and businesses to help one another, but we stay independent of both
- Equal treatment: we treat all reviews equally, and requires consumers and businesses using Trustpilot to follow our guidelines.

The above principles are reflected throughout our policies and processes, which must be followed by our Content Integrity and Fraud & Investigation Teams. For example, our Content Integrity agents apply the same processes for all flagged reviews. This is reinforced in our Help Center article [“Trustpilot’s Content Integrity Team”](#):

*“We apply our rules equally and consistently.”*

[REDACTED]

As set out in our Help Center article, [“The journey of reviews on Trustpilot”](#):

*“...Trustpilot treats all flagged reviews equally, regardless of whether they’re positive or negative.”*

[REDACTED]

For misuse or abuse of our platform, robust processes set out data-based thresholds (where relevant) to help agents determine when to apply sanctions such as to warn a user, or block access to their account.

[REDACTED]

**What safeguards do you use to ensure users are sanctioned consistently and fairly?**

**Q18. Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?**

As stated above, we believe it is important to provide users with relevant information as they need it throughout the user journey. Educational information and reminders about the appropriate tone to use online are important measures to help prevent people posting harmful and illegal content.

**Q19. To what extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?**

As outlined in the table above (at **1.6**), the purpose, nature and functionalities that the Trustpilot service provides act as a frame for how people use our services and the types of content posted on Trustpilot. The *absence* of certain functionalities means that their associated risks are not present. However, we also design features to minimise the risk of illegal content.

We also believe in taking a proactive, preventative approach to minimise the risk of users posting illegal content. In line with this, one recent initiative is the introduction of our optional facility for reviewers to verify their identity by safely and securely sharing a copy of government-issued photo ID, along with a selfie. The process is explained in our [Help Center article here](#). [REDACTED]

**Q20. How do you support the safety and wellbeing of your users as regards illegal content?**

See also our answers to **Q8**. The support we provide reflects the type of illegal content that we see on Trustpilot. As stated above, where we are made aware of reviews that contain suicide-related content, we follow processes developed in collaboration with Samaritans.org to direct users to appropriate supporting materials.

**Q21. How do you mitigate any risks posed by the design of algorithms that support the function of your service (e.g. search engines, or social and content recommender systems), with reference to illegal content specifically?**

See our answer to **Q3**.



The nature of our site and its functionalities mean that not all of the algorithms central to the function of our service are likely to pose particular risks in terms of illegal content. However, these aspects are assessed, where relevant, in the ordinary course of design and implementation. For example, one of the key algorithms is our “TrustScore”, which is the [overall measurement of reviewer satisfaction](#), represented on a scale from 1 to 5. This does not promote particular review content — it is simply a score calculated for every business that is reviewed on Trustpilot, and revised daily or as new reviews come in.

A business that has a high TrustScore is more likely than a business with a low TrustScore to be seen by users who make use of our “Categories” function, which is our main recommender or search system. However, anyone reading about businesses using this feature can easily adjust the search parameters.

Our Categories function groups different businesses based on industry, product or service and makes them searchable according to adjustable parameters. The category names and definitions are in accordance with the Google Business Profile category system and businesses cannot suggest new categories. Users can control which parameters are used to filter businesses in each search, but the current default will display companies with the highest TrustScores first, where businesses only qualify if they’re actively asking for reviews and have received at least 25 reviews in the past 12 months. In practice, businesses with high TrustScores are likely to have more five-star reviews containing positive content and less one-star or two-star reviews (with negative content and which are more likely to include the types of illegal content covered above). Therefore, as an inherent part of the Categories design, it is unlikely that our main recommender system will *increase* the likelihood of leading users to illegal content such as hate speech, terrorism, threats of violence, harassment etc. In any event, as explained above at our answer to **Q3**, our design process effectively builds in a consideration of risks. We also strive to provide transparent information around how our algorithms work, let users know why they are seeing certain results and empower them to change the parameters.

All of our systems are subject to testing and monitoring, and considerable human oversight to investigate and address any problems. Businesses, consumers or others can let us know about problems via our Contact Us form in our Help Center, or they can provide feedback about where we should improve our services by reviewing Trustpilot on our site.

## Child protection

**Q22. What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?**



Trustpilot is not intended for use by children. At 1.2, our [Terms & Conditions](#) for reviewers prohibit children from creating a registered account:

*“1.2 You must be 18 years of age or older in order to become a Registered User.”*

We allow users the option of securely verifying their identity, where the process of verifying typically requires using identification documents most commonly held by adults. However, verification does not focus specifically on age.

**Q23. Can you identify factors which might indicate that a service is likely to attract child users?**

These are likely to vary across age groups, but generally visual content such as videos and photos, and functionalities that include content sharing and private messaging are likely to attract children, as evidenced by the popularity of YouTube, TikTok, Instagram, Facebook and Snapchat.

In contrast, children are less likely than adults to want to read text-based reviews of businesses on Trustpilot, and generally, review content is not likely to be particularly appealing to younger individuals. The content of our website is targeted to a mature audience who are making purchases online. Our interface reflects this, from the colour choice and presentation to its focus on text rather than visual elements.

**Q24. Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?**

Currently, Trustpilot does not use age assurance or age verification tools.



Trustpilot is not intended for use by children. At 1.2, our [Terms & Conditions](#) for reviewers prohibit children from creating a registered account:

*"1.2 You must be 18 years of age or older in order to become a Registered User."*

**Q25. If it is not possible for children to access your service, or a part of it, how do you ensure this?**

N/A

**Q26. What information do you have about the age of your users?**

We take into account overall, aggregated user demographics in designing new features, functionalities or upgrading existing parts of the service, since this can also impact how users understand and use the service. However, we do not gather and hold information about the age of individual users in a structured way.   


We do not share information about user age with third parties.

**Transparency**

**Q27. For purposes of transparency, what type of information is useful/not useful? Why?**

At Trustpilot, we publish annual transparency reports. We believe that transparency about how platforms operate provides benefits to everyone and in particular, is critical in order to maintain trust with the consumers and businesses that use them. However, different services have different ways of operating. Therefore, the data from each service should be considered within the context of how that platform operates. In order to use standardised information and metrics as a form of comparison across services, these different contexts will need to be acknowledged: even small differences in business models, processes and definitions, and how data is recorded and counted (among other things) can affect whether or not figures are truly comparable across platforms.

As one example, Trustpilot is an open platform where reviews can be posted at any time, and consumers do not have to receive an invitation from a business. Many of our review platform competitors have adopted a 'closed' approach where companies collect reviews solely through invites, and can 'unpublish' those reviews they don't like, either after the fact or by pre-screening. It is unlikely to be clear to consumers how they can compare figures from open platform models to those from more closed models in a useful way. Similarly, the type of content posted and the functionalities provided by each service will influence the figures reported about illegal content. For example, a social media platform has a different purpose to an online reviews platform and, by its nature, will receive different types of user-generated content and different types of illegal content. A private messaging service will be used by consumers in a way that contrasts with both social media and online reviews.

We take the view that even platforms that can appear to be similar, such as two online review platforms, will have significant differences in the systems they use and how data is treated, which could distort any purported comparison. Therefore, any publication of data that consolidates and aggregates data across platforms must be treated carefully.

## Other

**Q28. Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?**

As stated at 7.1, we believe that increasing digital literacy and education around what is illegal on the internet will be beneficial to users of online platforms and services. In practice, we observe that there are considerable variations in people's expectations about what types of content are illegal and will be removed. This is partially reflected in the low validity rate for reports we receive from consumers about (harmful or) illegal content: overall, much of the content flagged to us for these reporting reasons does not breach our guidelines and is unlikely to be illegal.

However, for the content posted that *is* illegal, we take the view that preventative and proactive steps can be relatively effective in guiding people towards staying within the law, and there are further opportunities to encourage people to engage in more constructive and useful conversations online instead of posting content that is potentially harmful to other users.

## Appendix 1 - Examples of flagged reviews on Trustpilot

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] further invalid report of hate speech or discrimination by a UK business (#7777269):

[REDACTED]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]



[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

