Australian Code of Practice on Disinformation and Misinformation
Redbubble Limited
Annual Transparency Report
1 January 2022 - 31 December 2022

Summary

Redbubble is a global artist marketplace dedicated to giving independent artists a meaningful way to sell their creations. Redbubble hosts user-generated content uploaded by artists and provides them with online tools to upload their art and to design and sell products printed with their art to their customers worldwide.

Redbubble recognises the harm that arises from the spread of misinformation and disinformation and is committed to preventing, detecting and removing such harmful information from its marketplace. In doing so, Redbubble strives to balance the fostering of artistic freedom with the goal of preventing the spread of disinformation and misinformation through its platform.

Redbubble has increased its use of scalable technologies, such as duplicate detection, image matching, keyword detection, text-in-image detection, machine learning and artificial intelligence to more accurately and efficiently identify and remove problematic content on its platform. Redbubble has also expanded the misinformation and disinformation topics that it proactively screens for, particularly in the area of harmful medical misinformation, which is detailed in Outcome 1a.

This report provides transparency into the measures that Redbubble takes to prevent, detect and remove disinformation and misinformation on its marketplace.

Commitments under the Code

Outcome 1a: Reducing harm by adopting scalable measures	Opt in
Outcome 1b: Inform users about what content is targeted	Opt in
Outcome 1c: Users can easily report offending content	Opt in
Outcome 1d: Information about reported content available	Opt in

Outcome 1e: Information about recommender engines	Opt in
Objective 2: Disrupt advertising and monetisation incentives for disinformation	Opt in
Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms	Opt in
Objective 4: Empower consumers to make better informed choices of digital content.	Opt in
Objective 5: Improve public awareness of the source of political advertising carried on digital platforms	Opt out
Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research	Opt in
Objective 7: Signatories will publicise the measures they take to combat Disinformation	Opt in

Reporting against commitments

Outcome 1a: Reducing harm by adopting scalable measures

Redbubble prohibits users from uploading harmful disinformation and misinformation to the marketplace in the Redbubble User Agreement, published at https://www.redbubble.com/agreement, and the Redbubble Community and Content Guidelines at: https://help.redbubble.com/hc/en-us/articles/202270929.

Efforts over the last year to combat the spread of disinformation and misinformation include the expansion of current scalable technologies and implementation of new ones.

- Duplicate detection to find previously moderated images that users may re-upload;
- Image matching to detect content that is similar to images that are known to contain disinformation or misinformation;

- Expansion of keyword detection in text-based user-generated fields, such as title, tags and description to detect keywords associated with disinformation and misinformation;
- Text-in-image matching to detect text-based misinformation and disinformation contained within the image itself;
- Machine learning to detect user accounts that are connected to networks of accounts that are known to violate Redbubble policies; and
- Artificial intelligence to detect users' keyword tagging patterns that are associated with disinformation and misinformation.

Every image uploaded and keyword generated on the Redbubble platform must pass through one or more of these scalable detection technologies. Continuous improvements to tooling allow Redbubble to more accurately and efficiently detect and remove misinformation and disinformation from its marketplace. With the optimization of this tooling, works found via automated means increased from 105K in Dec 2021 to 119K in Dec 2022. The inflow of duplicate content decreased by 50% by Dec 2022 and the amount of time from proactive identification to removal of violative works was reduced to a baseline weekly average of less than 24 hours.

The Redbubble Content Safety Team proactively screens the marketplace on a daily basis for potential misinformation and disinformation detected by Redbubble's tools, and the team removes content that it determines violates our User Agreement and Community Guidelines.

When the Content Safety Team makes decisions relating to content that potentially includes misinformation or disinformation, the team considers Redbubble policies, past decisions (to ensure consistency of approach and decision-making) and further research particularly in relation to new or emerging topics. The Content Safety Team's framework for content review is built on clear criteria and a repeatable and scalable workflow. This allows the team to make moderation decisions in an unbiased and consistent manner. Furthermore, policies and decision-making frameworks undergo continuous review and refinement to ensure their ongoing effectiveness.

Redbubble proactively monitors and screens for over 300 different content safety topics on its platform, spanning issues like incitement of violence, racism, disinformation and misinformation.

The following are examples of misinformation and disinformation topics that Redbubble screens for:

- Medical misinformation, such as anti-vaccine propaganda that may encourage the spread of communicable disease;
- Denials of real-world catastrophes, such as the Holocaust; and
- Political misinformation, such as false political conspiracy theories that are linked to real-world harm.

In 2022, updates were made to the type of content targeted in proactive screening with a focus on medical misinformation around vaccines. To update the screening guidelines, the team made use of external sources in determining the boundaries of misinformation, such as independent fact-checking websites. The team compiled additional reference content to create review guidelines with updated training for content review teams and quality assurance of moderation decisions.

In the vaccines area, Redbubble's tools and review guidelines detect content that misleads consumers to believe that:

- Vaccines may make people more susceptible to Covid or other infections;
- Certain treatments that have not been approved by applicable medical regulators are more effective than vaccines; and
- Common household cleaning supplies provide full protection against Covid or other infections.

For example, images identical or similar to the one below are removed when uploaded to the Redbubble platform:

Redbubble also applies restrictions on product types sold by artists over the marketplace. Certain messages and imagery may be acceptable on most product types but not on certain ones where the misinformation and disinformation risk is high. For example, a mask printed with the words "N-95" could mislead the wearer and the public to believe the mask has a certain filtration level.

In 2021, these tools contributed to the removal of 23,123 user-uploaded images that violated Redbubble's User Agreement and Community Guidelines, including content that contains misinformation and disinformation. In 2022, that number increased by 32% to a total of 30,613 images.

Outcome 1b: Inform users about what content is targeted

The communication to users of what constitutes misinformation and disinformation is important in stopping its spread. To this end, Redbubble publishes content rules in its Community Guidelines and in various Redbubble Help Centre articles. In the Redbubble User Agreement, users are required to adhere to the Community Guidelines and represent and warrant that the products they sell are free from misinformation and disinformation. The User Agreement states that Redbubble reserves the right to review and in its sole discretion remove any such content from the website and terminate user accounts.

The Community Guidelines detail Redbubble's rules regarding user behavior and content on its marketplace and are made public in the spirit of open communication with artists and their customers. The relevant prohibition under the Community Guidelines addresses misinformation and disinformation in the following terms:

> "Harmful misinformation is not permitted. We define this as any misleading or false information that harms or significantly threatens public health and safety, or where the intent is to cause fear and suspicion about a topic that can cause real-world harm."

The Community Guidelines are expressly noted to be adaptable and subject to refinement over time as the environment and circumstances change.

The Redbubble User Agreement provides further information to artists proposing to upload content to the marketplace. It provides that an artist uploading content represents and warrants that:

> "The content does not contain material that is harmful, abusive, inflammatory or otherwise objectionable; and
>
> The content is not misleading and deceptive and does not offer or disseminate fraudulent schemes or promotions."

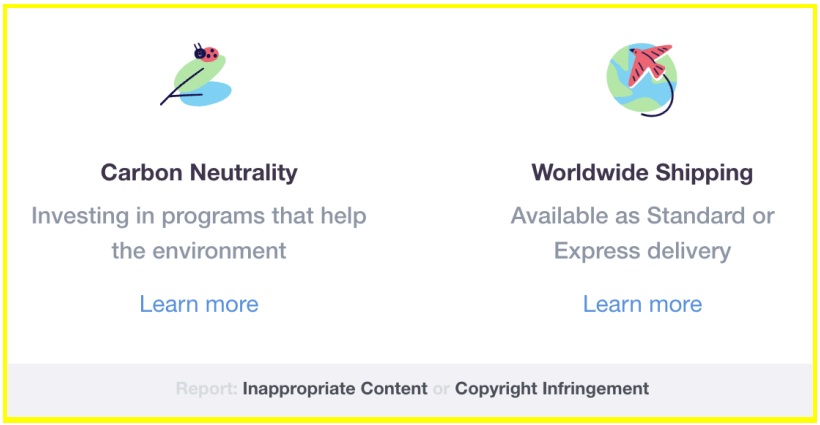
Additional information for users is included in the Redbubble Help Centre's Community Guidelines FAQ located at: https://help.redbubble.com/hc/en-us/sections/4404750122004-Community-Guidelines-FAQ. The Content Safety team adds to this FAQ on a regular basis to provide helpful information in response to commonly asked questions regarding disinformation and misinformation and other content safety topics.

Outcome 1c: Users can easily report offending content

Redbubble requests that all users of our marketplace flag behavior or content that contravenes the Community Guidelines through one of the reporting functions on our site. The User Agreement prompts users to report such content:

> “Please help us by letting us know straight away about any inappropriate Content you see on the Marketplace. You can do this by clicking the "Inappropriate Content" link displayed on each Product listing page.”

Redbubble provides a simple and accessible reporting tool for users. On every product listing page created by sellers on the marketplace, a prominent “Report Content” link is provided:



This link directs users to a web form (shown below) where they can quickly and easily report content that they believe falls outside of Redbubble’s policies, including works that may contain disinformation or misinformation.

The Content Safety team regularly reviews all user reports to assess whether the content or uploading account should be subject to moderation or other actions. This reporting tool

continues to be an important way to detect content that violates Redbubble policies, including disinformation and misinformation.

Outcome 1d: Information about reported content available

In 2022, users submitted over 14,000 reports using the reporting functionality described in Outcome 1c, which includes but is not limited to reports of misinformation or disinformation.

Outcome 1e: Information about recommender engines

Product recommendations to users of the Redbubble marketplace are primarily based on keyword-matching algorithms that connect the text-based title, tags and description generated by the artists with user search and navigation behavior, such as prior keywords the user searched for and keyword similarities between content they clicked on and additional content that may interest them.

Objective 2: Disrupt advertising and monetisation incentives for disinformation.

Redbubble reduces monetisation incentives from artist sales of products on the marketplace by swiftly detecting and removing accounts and content that violates Redbubble policies. These measures are described in more detail throughout this report. If a user violates the Redbubble User Agreement or Community Guidelines, their violative content will be removed and their accounts will be subject to account penalties, up to and including account termination. Networks of connected accounts will also be terminated.

Redbubble also disrupts incentives by using keyword blocking tools that prevent content tagged with terms related to misinformation and disinformation from appearing on offsite marketing platforms where artists promote their products and generate sales. This blocklist covers approximately 8,000 keywords related to content safety topics, including hundreds of terms related to misinformation and disinformation.

Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms.

At the account-level, Redbubble uses third-party account abuse detection software that combines machine learning with a global network of data to detect users who are likely to violate the Redbubble User Agreement. This software uses data points that are customized to the Redbubble marketplace and allow room for adjustments based on emergent trends. This is an effective tool for maintaining content integrity on Redbubble and detecting users who are likely to upload content that perpetuates disinformation or misinformation. For example, this tool has helped detect users who use bots to create networks of multiple

accounts and attempt to upload large amounts of images intending to sell products related to trending topics that may cause public harm.

Objective 4: Empower consumers to make better informed choices of digital content.

The measures and tools discussed throughout this report summarize the ways that Redbubble detects and removes misinformation and disinformation, which mitigates the risk of harmful content and accounts that consumers could be exposed to on its marketplace.

Redbubble also puts artists front and center in its public communications, such as web copy and promotional materials, to ensure that consumers understand that images have been uploaded by independent artists and that the products offered on the marketplace are designed and sold by artists.

The Redbubble User Agreement makes it clear to users that artists are responsible for the content they upload, the products they offer for sale in their shops, and the titles, tags and descriptions they write to describe their products.

Email marketing materials contain the following statement: “All products on the Redbubble marketplace are designed and sold by independent artists”, and every product listing page created by artists contains the words “designed and sold by [artist’s username]”.

Various Help Centre articles published on the Redbubble website also seek to educate consumers on misinformation and disinformation, such as the article on masks located at: https://help.redbubble.com/hc/en-us/articles/360042269512-Masks-General-Information. This article conveys to consumers the importance that Redbubble places on being “transparent with you and to provide detailed information about this product.” The Content Safety team searches for misleading information that artists may associate with their products, especially with respect to masks.

Objective 5: Improve public awareness of the source of political advertising carried on digital platforms.

Redbubble has opted out of Objective 5, because political advertising is not considered to apply to the Redbubble business, and Redbubble does not sell ad space to parties conducting political advertising.

Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.

Redbubble is open to supporting independent research that has the purpose of improving public understanding of disinformation and misinformation. At this time, Redbubble does not provide financial support to third-party strategic research in this area.

Objective 7: Signatories will publicize the measures they take to combat Disinformation.

Redbubble will continue to publish these transparency reports, which are accessible to the public on the DIGI website at digi.org.au/disinformation.