

Digital Services Act Transparency report

17 Feb 2024 to 31 Dec 2024



DSA Transparency Report

At Canva, we take content moderation seriously and are committed to fostering a safe and inclusive platform. As a company that provides multiple services under the Canva brand, this transitional report consolidates data across all our services into a single document, ensuring a comprehensive overview of our content moderation efforts.

The figures presented here have been prepared exclusively for the **EU's Digital Services Act ("DSA")** and are not intended for any other purpose.

Section 1 - Report identification

In accordance with Articles 15 and 24 of the DSA, this transparency report provides some information about Canva's moderation efforts in the EU across its services between 17 Feb to 31 Dec 24 and is guided by the section and requirements in the European Commission-issued templates for transparency reports.

| Description | Value |
|---|-------------------|
| Name of the service provider | Canva Pty Ltd |
| Date of the publication of the report | 2025-02-25 |
| Date of the publication of the latest previous report | N/A |
| Starting date of reporting period | 2024-02-17 |
| Ending date of reporting period | 2024-12-31 |

Section 2 - Member States Orders

During the reporting period, we received 0 takedown requests from EU member states pursuant to DSA Article 9, and 15 information requests pursuant to DSA Article 10.

| Description | Total |
|--|---------|
| Number of orders to act against illegal content received | 0 |
| Number of orders to act against illegal content granted/complied with | 0 |
| Number of items moderated on the basis of an order to act against illegal content | 0 |
| Number of orders to provide information received | 15 |
| Number of orders to provide information granted/complied with | 11 |
| Median time [1] to inform the authority of the receipt of the order to act against illegal content | N/A |
| Median time to give effect to the order to act against illegal content | N/A |
| Median time to inform the authority of the receipt of the order to provide information | 1 hour |
| Median time to give effect to the order to provide information | 37 days |

Of the 15 orders to provide information received:

- 11 orders were granted
- 4 orders were not granted (3 orders closed due to no response and 1 order rejected)

The breakdown by member states (for orders to provide information). [2]

| Member State | Total | Granted/Complied With |
|----------------|-------|-----------------------|
| Spain | 8 | 6 |
| Germany | 3 | 3 |
| Belgium | 1 | 1 |
| France | 1 | 0 |
| Czech Republic | 1 | 1 |
| Austria | 1 | 0 |

Section 3 - Notices

Number of notices of illegal content in the EU received during the reporting period.

| Category | Number of notices received | Number of notices received from Trusted Flagger |
|---|----------------------------|---|
| Number of notices received | 17741 | 0 |
| Number of notices reviewed/moderated | 17023 | 0 |
| Median time to take action (hours) | 31 hours | N/A |
| Number of actions taken | 1076 | N/A |
| Number of actions taken on the basis of the law | 0 | N/A |
| Number of actions taken on the basis of the terms and conditions of service | 1076 | N/A |

The breakdown of notices by category of reported illegal content.

| Category of reported illegal content | Number of notices received | Notices received from Trusted Flaggers | Median time to action (hours) | Number of actions taken on the basis of the law | Number of actions taken on the basis of the T+Cs |
|--|----------------------------|--|-------------------------------|---|--|
| Illegal or harmful speech | 2745 | 0 | 11 | 0 | 90 |
| Sexual content | 2151 | 0 | 30 | 0 | 134 |
| Consumer information | 2063 | 0 | 36 | 0 | 39 |
| Animal Welfare | 1842 | 0 | 50 | 0 | 84 |
| Scams and fraud | 1642 | 0 | 13 | 0 | 409 |
| Civic discourse or elections | 1641 | 0 | 7 | 0 | 30 |
| Data protection and privacy violations | 1303 | 0 | 25 | 0 | 82 |
| Violence | 1006 | 0 | 20 | 0 | 46 |
| Risk for public security | 773 | 0 | 7 | 0 | 26 |
| Harmful to minors | 715 | 0 | 33 | 0 | 27 |
| Unsafe and prohibited products | 705 | 0 | 5 | 0 | 28 |
| Non consensual behaviour | 698 | 0 | 17 | 0 | 22 |
| Self harm | 433 | 0 | 52 | 0 | 3 |
| Intellectual property infringements | 85 | 0 | 26 | 0 | 56 |

Section 4 - Own Initiative

Number of content items from EU countries that Canva has proactively moderated on its own initiative during the reporting period.

| Description | Volume |
|---|------------|
| Number of items moderated | 30,379,676 |
| Number of items detected solely using automated means | 30,362,699 |
| Number of restrictions imposed | 19,339 |

The breakdown of own-initiative by category of illegal content.

| Description | Number of restrictions imposed at provider's own initiative | Visibility restriction (removal) | Visibility restriction (age restricted) | Account restriction (termination) | Number of restrictions imposed that was detected solely using automated means |
|--|---|----------------------------------|---|-----------------------------------|---|
| Illegal or Harmful Speech | 695 | 690 | 5 | 0 | 223 |
| Illegal products | 965 | 169 | 785 | 11 | 28 |
| Intellectual property infringements | 9485 | 9483 | 0 | 2 | 22 |
| Negative effects on civic discourse or elections | 22 | 22 | 0 | 0 | 1 |
| Non-consensual behaviour | 13 | 9 | 4 | 0 | 3 |
| Pornography or Sexualised | 4483 | 3759 | 711 | 13 | 915 |
| Protection of Minors | 681 | 561 | 5 | 35 | 132 |
| Risk of Public Security | 84 | 80 | 0 | 4 | 66 |
| Scams or Fraud | 2217 | 1902 | 0 | 315 | 950 |
| Scope of Platform Service | 5 | 1 | 0 | 4 | 0 |
| Self Harm | 24 | 17 | 7 | 0 | 6 |
| Violence | 734 | 353 | 381 | 0 | 16 |

Section 5 - Appeals, recidivisms and automated means

5.1 Appeals

259 appeals were submitted from the EU, with 255 appeals processed and no decisions omitted [3] as of the date of this report.

| Description | Total number | Number of decisions upheld | Number of restrictions newly imposed as a result of an appeal | Number of decisions reversed | Median time (hours) |
|--|--------------|----------------------------|---|------------------------------|---------------------|
| Number of appeals processed | 255 | 135 | 0 | 120 | 131 |
| Appeal based on procedural grounds | 10 | 1 | 0 | 9 | 569 |
| Appeal regarding the interpretation of illegality or incompatibility | 219 | 110 | 0 | 109 | 108 |
| Restriction imposed is claimed to be not diligent, objective, or proportionate | 26 | 24 | 0 | 2 | 122 |

| Description | Visibility restriction (removal) | Visibility restriction (age restricted) | Account restriction (termination) |
|---|----------------------------------|---|-----------------------------------|
| Number of decisions upheld | 69 | 60 | 6 |
| Number of restrictions newly imposed as a result of an appeal | 0 | 0 | 0 |
| Number of decisions reversed | 97 | 18 | 5 |

5.2 Out-of-court dispute settlement bodies

Number of out-of-court dispute settlement submitted during the reporting period.

| Total number of decisions submitted to OOC dispute settlement body | Decisions upheld | Decisions reversed | Median time (days) | Decision omitted | % of outcomes implemented |
|--|------------------|--------------------|--------------------|------------------|---------------------------|
| 0 | N/A | N/A | N/A | N/A | N/A |

5.3 Suspensions/Terminations imposed on repeat offenders

Number of suspensions/terminations imposed during the reporting period.

| Description | Total Number |
|---|--------------|
| Number of terminations enacted for the provision of manifestly illegal content | 387 |
| Number of terminations enacted for the provision of manifestly unfounded notices | 0 |
| Number of terminations enacted for the provision of manifestly unfounded complaints | 0 |
| Number of terminations enacted, originated from self-initiatives | 378 |
| Number of suspensions enacted for manifestly illegal content, unfounded notices or complaints [4] | 0 |

5.4 Use of automated means for content moderation

Number of uses of automated means in moderation during the reporting period.

| Description | Total | Appeals | Notices | Notices by Trusted Flaggers | Own initiative |
|--|-------|---------|---------|-----------------------------|----------------|
| Number of items solely processed by automated means [5] | 2922 | 0 | 2922 | 0 | 0 |
| Number of items partly processed by automated means [6] | 2377 | 0 | 0 | 0 | 2377 |
| Number of items not processed by automated means | 25099 | 255 | 7862 | 0 | 16982 |
| Accuracy rate of the items processed solely by automated means | N/A | N/A | 98% | N/A | N/A |
| Accuracy rate of the items processed partly by automated means | N/A | N/A | N/A | N/A | 98% |
| Possible Error Rate [7] of the automated means applied | N/A | N/A | 2% | N/A | 2% |

Section 6 - Use of Automated Means and Human Resources

6.1 Summary of the content moderation engaged in at Canva's own initiative

Canva aims to provide a safe platform for users to create content. To achieve this, we use automated means to detect potentially harmful content at scale, while human moderators validate automated responses and assess contextual nuances.

6.2 Meaningful and comprehensible information regarding the applied detection method

We use a combination of automated means and human detection methods.

- Automated detection flags potentially unsafe visuals and text based on Canva's Acceptable Use Policy (AUP) and escalate them to human moderators for review and decision-making.
- Human moderators proactively detect unsafe content and make moderation decisions using mechanisms such as targeted keywords sweeps.

6.3 Measures taken to provide training and assistance to persons in charge of content moderation

Human moderators are provided with standards of practices and guidelines to follow for content review and decision-making. During onboarding, they undergo initial training and receive internal guideline documentations on the AUP to help them understand and apply in relation to content moderation. Successful completion of an onboarding assessment is required before they can moderate content in production.

We also periodically update our internal guideline documents, ensuring moderators remain informed of any changes to the AUP. Additionally, higher-tier reviewers regularly evaluate moderators' decisions and provide feedback to ensure high-quality and consistent performance.

6.4 Summary of the use made of automated means for the purpose of content moderation

Canva uses automated means to support content moderation at scale. While automation assists in flagging potentially unsafe content, all restriction-imposed decisions are made by human moderators during this reporting period.

Detection:

- Automated means detect potential unsafe content, which is then reviewed by human moderators for final decision-making.

Decision:

- No restriction-imposed decisions were made solely by automated means during this reporting period. Automated processes are only used to close notices with a "No Violation" status when no violative content is detected in user-submitted reports.

6.5 Qualitative description of the automated means

We use automated means to scan visuals and text in regulated content to detect potential violations of the AUP.

- Visuals: Detects potentially unsafe content, such as pornography or sexualized imagery, that may violate the AUP.
- Text: Detects potentially harmful text, including phishing, scams, or spam, that could breach the AUP.

6.6 Specification of the precise purposes to apply automated means

Purposes of automated means in our content moderation:

- Supporting Detection: Assists moderators in identifying potentially unsafe content at scale, providing leads to enhance proactive human moderation efforts.
- Efficient Resolution: Expedites the handling of high-confidence, non-AUP-violating user reports, enabling faster response times.

6.7 Safeguards applied to the use of automated means

While automated means support our moderation workflow, all restriction-imposed decisions—such as content removal, age restrictions, or account termination—are made by human moderators during this reporting period.

We also monitor the ‘possible error rate’ by monitoring how often actions taken by automated means are later reversed through human manual action and successful appeals.

[1] The Median time is counted from the date the request is received to the date the request is closed.

[2] Any EU member state not listed indicates that 0 orders were received from them during the reporting period.

[3] Decision omitted refers to appeals that were reviewed but not processed with a final decision. Outstanding appeals are not classified as omitted.

[4] Temporary suspension is not a sanction Canva permits for repeat offenders of Canva's Acceptable Use Policy.

[5] Sole automation is only used for notices submitted by notices, where the automation determines the content has not committed a violation. If a violation is detected then those notices proceed to human moderation.

[6] Number of items partly processed by automated means measures the number of content detected and flag potential unsafe content by automated means to human moderators for further review. All decisions to impose restrictions are made entirely by human moderators.

[7] Possible error rate for own-initiative items partly processed by automated means is measured by the rate of restrictions-imposed decisions made by humans on items flagged by automated means that are later reversed as 'No Violation' after successful appeals.