

Microsoft Bing Transparency Report (Regulation (EU) 2022/2065)

In accordance with the requirements of Regulation (EU) 2022/2065 (the EU Digital Services Act or DSA) for Very Large Online Search Engines, Bing provides the following report on content moderation activities engaged in during the period of October 1 – December 31, 2023.

DSA Article 15(1)(a): Government Orders from Member States

- **The number of orders received from Member States’ authorities to act against illegal content provided by recipients of the service,¹ categorized by the type of illegal content concerned and the Member State issuing the order.**

During the relevant period, Bing received zero orders from EU Member States’ authorities to act against illegal content provided by recipients of the service.

- **The number of orders received from Member States’ authorities to provide specific information about one or more individual recipients of the service in the context of illegal content-related incidents, categorized by the type of illegal content concerned and the Member State issuing the order.**

During the relevant period, Bing received zero orders from EU Member States’ authorities requesting specific information about individual recipients of the service.

- **The time needed to inform the authority issuing the order, or any other authority specified in the order, of its receipt, and to give effect to the order.**

As noted above, Bing received zero orders from EU Member States’ authorities requesting specific information about individual recipients of the service during the relevant period.

DSA Article 15(1)(c): Own-Initiative Content Moderation

- **Meaningful and comprehensible information about the content moderation engaged in at the provider’s own initiative, including the use of automated tools, and the measures taken to provide training and assistance to persons in charge of content moderation.**

¹ “Recipients of the service” does not include the owners of websites indexed by the online search engine. DSA Recital 77. Please see [Microsoft’s CSR Reports Hub](#) for additional information on content removals regarding indexed website content.

This section describes activities Bing undertakes to detect and address illegal content or information in violation of Bing’s terms and conditions that is provided by recipients of the service, and the activities Bing’s generative artificial intelligence (AI) features² undertake to address recipient accounts in violation of their terms of use and code of conduct.³

Use of AI-based classifiers on search prompts. Traditional web search begins with a user search query, namely the input (text, voice, or image) a user sends to Bing from the search bar. Similarly, to initiate a search in Bing’s new AI-enhanced search experience, a user submits text, voice, images, and/or other enabled queries as input to the AI model, which then performs the relevant searches of the Bing service and generates a response (or, in the case of Bing Image Creator, generates an image). This is known as a “prompt”. Bing uses classifiers (machine learning models that help sort data into labeled classes or categories) and content filters on user search queries and prompts to help mitigate harm or prevent misuse. Examples include requests for information that could potentially lead to users being unexpectedly exposed to self-harm, violence, graphic content, hateful content, or misleading information. Flags from these classifiers may lead to mitigations, such as not returning generated content to the user, diverting the user to a different topic, or redirecting the user from AI-enhanced search to traditional web search. Bing tracks accuracy metrics for these measures such as precision, recall, error rate, under blocking and over blocking to help monitor the interventions’ effectiveness and help ensure Bing does not unduly limit free access to information. Accuracy metrics are overseen by human reviewers and based on Bing content policies.

Automated content detection – Bing Visual Search. Bing’s Visual Search feature allows users to upload an image and search for similar images or ask Bing Chat questions about the image. As part of Microsoft’s longstanding commitment to preventing the spread of child sexual exploitation and abuse imagery (CSEAI), Bing uses hash matching technologies to detect matches of previously identified CSEAI. In the context of the immediate search, the use of these technologies furthers Bing’s goal to avoid inadvertently surfacing potentially harmful web content to users. More broadly, images that have been used as queries in Bing Visual Search may contribute to training Bing’s image-matching algorithms; by scanning images that users attempt to upload, Bing helps to ensure that CSEAI is not included in the Visual Search training data. Please see below for additional details about these processes.

Bing’s generative AI features (powered by Microsoft Copilot and Microsoft Designer) enforcement actions. Copilot (including Copilot in Bing features) takes actions on recipient accounts where violations of Copilot Terms of Use are found, to either temporarily or permanently limit their access to be the same as an unauthenticated user,⁴ which has the same functionalities but limited conversation turns. Designer (which includes Image Creator from Designer features in Bing) takes actions on recipient accounts where violations of Designer Terms of Use are found, to either temporarily or permanently suspend their account access to Designer.

² Bing’s generative AI features include Microsoft Copilot (including Copilot in Bing features) and Microsoft Designer for Web Image Creator and Brand Kit (Designer). Please see [Copilot Terms of Use](#) and [Designer Terms of Use](#) for additional information.

³ Due to their nature, Bing search and generative AI features do not generally conduct “content moderation” as that term is defined in the Digital Services Act due to the nature of those products, as content is not provided by recipients of the service nor hosted by Bing. Search queries, similar to user prompts, trigger systems that ensure the services work as intended. The outputs of these systems are not provided by a recipient of the service. Nevertheless, we have provided additional descriptions of how these systems operate.

⁴ An unauthenticated user is defined as a user who is not logged into their Microsoft account when accessing Copilot features.

Training and assistance. Human reviewers receive extensive training on our policies including the rationale behind them and how to apply them accurately and consistently. Decisions are periodically checked to ensure the policies are being applied consistently. Ongoing coaching and training are provided to review teams, as legal obligations evolve, new types of harms emerge, or policies otherwise need to adapt. For high-consequence harms, like child sexual exploitation and abuse, specialized teams receive additional focused training. Microsoft provides a program to support the mental and emotional wellbeing of Microsoft employees whose work may bring them into contact with objectional material. This program provides resources such as one on one counseling, monthly education sessions, on demand small group sessions, virtual community of practice gatherings, and access to program manager office hours. Microsoft requires our vendors to provide wellness programs for any vendor employees working with objectional material.

- **The number and type of measures taken that affect the availability, visibility, and accessibility of information provided by the recipients of the service, categorized by the type of illegal content or violation of the terms and conditions, by detection method and by the type of restriction applied.**

During the relevant period, Bing took voluntary actions to detect, block, and report 34,532 items of suspected CSEAI content provided by recipients of the service. These items were identified through the use of automated content detection in Bing Visual Search as described above.

- **The number and type of measures taken that affect recipients' ability to provide information through the service, and other related restrictions of the service, categorized by the type of illegal content or violation of the terms and conditions, by detection method and by the type of restriction applied.**

Bing does not provide capabilities for users to share content or interact with other users on the platform. As such, during the relevant period, Bing did not take measures that affected a recipient's right to share content or interact with other users on the service due to illegal content or violations of terms and conditions. In the case of Bing's generative AI features (which are powered by Microsoft Copilot and Microsoft Designer), there are some scenarios where a user may be suspended or blocked from using the Copilot or Designer services due to violations of the Terms of Use. Note that in Copilot in Bing (previously known as Bing Chat) and in Image Creator from Designer (previously known as Bing Image Creator) there is still no ability for users to upload, post or share content on the service. Account restrictions are not related to actions affecting other users, but instead relate to users attempting to generate content for their own use that violates relevant terms or codes of conduct.

Copilot in Bing is currently unable to separate its reporting from the standalone Copilot service. During the relevant period, Copilot (including Copilot in Bing features) temporarily restricted the recipient's ability to interact with the product (thus limiting the number of turns in a conversation) in 60 instances and permanently restricted the recipient's ability to interact with the product in 17 instances as a result of violations of Copilot Terms of Use and Code of Conduct which prevented the recipient from further exploitation of Copilot safety systems. For each of these instances, the violation was due to "jailbreaks" - i.e., attempts to bypass Copilot safety systems.

Image Creator from Designer in Bing is currently unable to separate its reporting from the standalone Designer service. During the relevant period, Designer (which includes Image Creator from Designer features in Bing) temporarily or permanently restricted access in 45,412 instances (thus limiting the recipient account's access to Image Creator from Designer features in Bing) as a result of violations

of the Microsoft Designer Terms of Use. In the relevant period these actions were taken as a result of users attempting to bypass Designer safety systems.

DSA Article 15(1)(d): Appeals

- **Number of complaints received through the internal complaint-handling system provided for recipients of the service to lodge complaints against decisions whether or not to remove or disable access to information provided by a recipient of the service, restrict visibility of information provided by a recipient of the service, suspend or terminate the provision of the service, in whole or in part, suspend or terminate a recipient’s account, or restrict a recipient’s ability to monetize information provided.**

During the relevant period:

- Bing received zero appeals of the types of decisions described above;
- Copilot (including Copilot in Bing features) received zero appeals of the type of decisions described above;
- Designer (which includes Image Creator from Designer features in Bing) received 15,296 appeals⁵ of the type of decisions described above.

DSA Article 15(1)(e): Automated Content Detection

- **Any use made of automated means for the purpose of content moderation, including a qualitative description, indicators of the accuracy and the possible rate of error of the automated means used in fulfilling those purposes, and any safeguards applied.**

Automated content detection – Bing Visual Search. As described above, Bing relies on the hash matching technologies PhotoDNA and MD5 to detect matches of previously identified CSEAI in images uploaded to the Bing service or an associated website by recipients of the service using the Bing Visual Search feature. This is one element of Microsoft’s overall commitment to prevent the spread of CSEAI, as described more fully in its [Digital Safety Content Report](#) and other public announcements.

Hash-matching technology works by using a mathematical algorithm to create a unique signature (known as a “hash”) for digital images and videos. The hashing technology then compares the hashes generated from content provided by the recipient of the service with hashes of reported (known) CSEAI, in a process called “hash matching”.

A layered approach to detection of CSEAI is applied in this context, combining both hash-matching technology and manual review. Microsoft implements its own hash verification process in which Microsoft-trained analysts review and confirm images associated with hashes provided from non-profits and other industry partners. Microsoft also implements an additional manual review process as an ongoing hash quality check. Reversal rates of the initial content moderation decision (for example, on appeal) are tracked, as a reflection of Microsoft’s application of hash-matching technology.

⁵ The number of appeals for Designer is represented at the global level. A recipient whose account access to Designer features was suspended as a result of violations of the Microsoft Designer Terms of Use may have appealed the suspension multiple times.

DSA Article 42(3): Information on monthly active users

- **Information on average monthly active recipients of the service for each Member State in the European Union.**

Information about the average monthly active users of the Bing service in the European Union is published semi-annually. The most recent information is available [on this page](#) and reports approximately 116 million average monthly users in the EU during the six-month period ending December 31, 2023. The table below details the monthly active users for each EU Member State during this period. Note that these numbers may include overlap in recipients of the service who accessed Bing from multiple Member States during the relevant time period.

EU Member State	Average MAU (million)
Austria	3.1
Belgium	4.1
Bulgaria	0.8
Croatia	0.6
Cyprus	0.2
Czech Republic	3.0
Denmark	2.1
Estonia	0.3
Finland	1.7
France	19.6
Germany	27.0
Greece	1.5
Hungary	1.7
Ireland	2.5
Italy	11.5
Latvia	0.4
Lithuania	0.7
Luxembourg	0.3
Malta	0.2
Netherlands	7.8
Poland	8.7
Portugal	2.9
Romania	1.9
Slovak Republic	1.0
Slovenia	0.5
Spain	10.7
Sweden	3.7

This information was compiled pursuant to the Digital Services Act and thus may differ from other user metrics published by Bing.