



# Bing Systemic Risk Assessment Report

August 2023

# Contents

<b>Executive Summary</b> .....	<b>3</b>
<b>Introduction</b> .....	<b>5</b>
Timeline/Scope.....	5
Structure of Document.....	5
<b>Background</b> .....	<b>6</b>
Nature of Bing as a platform.....	6
Bing’s values and commitments .....	6
Impact to risk profile and key risk considerations.....	8
Bing’s approach to and investments in protecting consumers and their fundamental rights online .....	10
Search algorithms and recommender systems: .....	10
Content Moderation systems: .....	10
Applicable Terms and Conditions and their Enforcement: .....	11
Systems for Selecting and Presenting Advertisements: .....	11
Data-related practices: .....	12
<b>Methodology</b> .....	<b>13</b>
Overview of Process .....	13
Assessment Tools.....	14
Report Review Process and Approvals.....	15
<b>Bing Service at a Glance</b> .....	<b>16</b>
DTSP Assessment Level 3 .....	16
High level overview of key Bing surfaces, features, and risks.....	17
<b>Summary of Results</b> .....	<b>22</b>
Summary of Findings: Key Risks and Mitigations .....	22
The dissemination of illegal content through the service.....	22
The actual or foreseeable negative effects for the exercise of fundamental rights.....	25
Key Takeaways.....	51
Key strengths .....	51
Looking forward .....	53
<b>Conclusion</b> .....	<b>54</b>

# Executive Summary

This Risk Assessment Report, which covers all features of the Bing search services, including core search features, enhanced search features, advertising, and new generative AI features, is designed to meet Bing's obligations as a Very Large Online Search Engine under Article 34 of the EU's Digital Services Act (DSA). It provides a summary analysis of the systemic risks present on Bing as of August 2023, key risk mitigations and their efficacy, and areas for continued improvement.

As described in more detail in the following pages, Microsoft Bing is an online search engine, which indexes third-party web content for consumption by its users. In this context, key risks on the platform are primarily related to how Bing's algorithms rank and display this third-party content to users such that users are not misled or harmed by the content returned in their search results. Bing recognizes that content returned in search results could negatively impact each of the 11 enumerated systemic risks in the DSA, including that search results may include illegal materials, or materials that could negatively impact human dignity, privacy, nondiscrimination, consumer protection, rights of the child, democratic processes, public safety, individual physical or mental health, or other important interests. At the same time, given that a primary mechanism for research today is accessing information online, and that search engines are the primary tool users have to discover new material online, over-moderation of content in search could have a significant negative impact on the right to access information/freedom of expression. Bing must carefully balance these competing fundamental rights and interests as it works to ensure that its algorithms return the most high-quality content available that is relevant to the user's queries, working to avoid causing harm to users without unduly limiting their ability to access answers to the questions they seek. In some cases, different features may require different interventions based on functionality and user expectations.

While Bing's risk mitigations may on occasion involve removal of content from search results, where legal or policy considerations warrant removal, in many cases Bing has determined that other mitigations, such as targeted ranking interventions, or additional digital literacy features such as Answers pointing to high authority sources, trustworthiness signals, or content provenance indicators, are more effective. Bing regularly reviews the efficacy of its mitigations using metrics to identify additional areas for improvement and works with internal and external subject matter experts in key policy areas to identify new threat vectors or improved mechanisms to help users' ability to engage safely with content in search results (e.g., new approaches to digital literacy messaging). Bing also takes steps to limit the risk of harmful or illegal content appearing in paid search results via published and consistently enforced terms and conditions and content policies, supplemented by internal enforcement teams and processes for responding to reported issues.

Bing's new generative AI features, the purpose of which is to provide a next-generation search experience for users to find the web content they are seeking more efficiently, including through more sophisticated questions and interactions with the service, is built on longstanding safety systems in search, supplemented by additional protections for new risks related to AI like conversational drift, hallucinations, and jailbreaking. Because Bing's generative AI features largely mirror its search

functionality, the systemic risks inherent in Bing's generative AI features largely mirror risks in traditional search: there is a likelihood that Bing's algorithms and AI systems return third-party content to users that is illegal, misleading, or harmful, or that Bing's safety systems overblock responses and negatively impact access to information or other rights. Bing has partnered closely with Microsoft's Responsible AI team to proactively address these harms and has been transparent about its approach in [The New Bing: Our approach to Responsible AI](#). Bing continues to evolve these features based on user and external stakeholder feedback.

Since Bing does not allow users to post or share content within the service, enforcement of Bing's terms and conditions or content moderation of user content is not a major vector for systemic risks on the platform. However, in the limited scenarios where Bing may take action against a user for their activities on Bing, such as where users violate Bing's terms in its generative AI features by attempting to generate harmful content, or where users attempt to use known child sexual exploitation and abuse imagery (CSEAI) as a visual search query, Bing has established processes to take action as needed and ensures users' rights to contest those decisions are protected.

Bing does collect and store user data, and recognizes it has an obligation to uphold data subject rights and protect users from data-related harms. Bing takes steps to limit risk of harm to users' personal data via its data-related practices in search and in ads, including compliance with relevant privacy laws such as the EU Global Data Privacy Regulation (GDPR) and with Microsoft privacy and security standards.

As described in more detail below, in light of the purpose and functionality of the service and implemented risk mitigations, Bing has adequately addressed the systemic risks enumerated in the DSA. However, as with any service, there is continued room for improvement, particularly as generative AI features and other novel technologies continue to evolve. As a result of this Risk Assessment process, Bing has identified specific areas to focus on in the next period and is implementing work plans to achieve these improvements over the next period.

# Introduction

This Risk Assessment Report describes how the Bing search services have identified, analyzed, and mitigated key systemic risks on the platform, as well as where there is additional room for improvement. This document is not meant to include comprehensive detail on all risks and mitigations, but instead provides a summary. Additional details are available in supporting materials.

## Timeline/Scope

This initial Risk Assessment Report covers Bing's search services as they exist as of August 2023. Bing search services include the service offered on Bing.com, mobile apps, and similar interfaces, and all relevant features, including "core search" features, enhanced search features like answers and suggestions, advertising, and generative AI features.

## Structure of Document

This Risk Assessment Report covers the following:

- **Background:** Background on the Bing search services, providing a summary of the nature of Bing as a search platform how the functionality and design of Bing impacts its risk profile and key risk considerations, and how Bing addresses these concerns through its values, commitments, and investments in safety systems.
- **Methodology:** An overview of the process Bing used for creating this Risk Assessment.
- **Service at a Glance:** An overview of the Bing service, its key surfaces that could give rise to systemic risks, and the risk profile of each.
- **Summary of Results:** A summary of the key risks and mitigations identified in each of the enumerated systemic risk areas, including which areas are vectors of highest and lowest risk, along with plans for further improvements.

# Background

## Nature of Bing as a platform

As an online search engine, the primary objective of Bing is to discover, understand, and organize the Internet's content to offer the most relevant and high-quality results available in response to user queries. Bing supplements this core functionality with additional features designed to help users find answers to their questions more quickly, such as enhanced search features like answers, search suggestions, and narrowed search verticals like shopping, travel, image, or video. In line with Microsoft policies and principles around responsible AI, privacy, digital safety, information integrity, and other critical social issues, Bing has developed a holistic digital safety ecosystem that encompasses programmatic safety systems involving ranking improvements, content filtering, operational monitoring, abuse detection, transparency, digital literacy features, user controls, and reporting functionality, among other protections, to provide a safe search experience for its users throughout the product. For more detail, please see [How Bing Delivers Search Results](#).

In 2023, Bing launched a preview of a new set of generative AI features powered by large language model (LLM) technology. These new features combine the power of LLMs with Bing's sophisticated search algorithms<sup>1</sup> to enable users to ask more complex questions and receive more complex responses, and also enable users to spark creativity through creation of stories, poems, songs, images, or other content. These features build on Bing's long-established safety systems, with additional protections imposed to address new possible harms related to these novel AI features. While the LLMs used by Bing were developed by a technology partner, OpenAI, OpenAI does not have direct involvement in the operation of the Bing service; Bing's models are hosted and managed independently of OpenAI, and the service is layered with additional features and protections, such as a bespoke metaprompt and additional safety filters that make Bing's implementation of these LLMs unique. More detail on how Bing approached the development of these new features with safety in mind is available in [The New Bing: Our Approach to Responsible AI](#).

As a free platform, Bing is monetized almost entirely through search advertising. When a user conducts a Bing search, they are shown both organic (unpaid) results and, where relevant, one or more ads. In search, these advertisements are contextually relevant to the user's search queries (rather than targeted based on personalized profiling), and often can help users find the content they are seeking more quickly. Advertisements on Bing are governed by detailed [terms and conditions](#) and [content policies](#), which are consistently enforced.

## Bing's values and commitments

Online search platforms like Bing provide significant value to users by indexing the trillions of pages of ever-changing content on the internet so that users can, with a simple query input, find the most relevant and authoritative materials that are responsive to the topic they are researching.

<sup>1</sup> More information on how Bing developed the new Bing features is available here: [\(4\) Building the New Bing | LinkedIn](#)

Given that most research today is conducted online, search engines play a vital role in promoting the fundamental right to access information/freedom of expression, and to support media pluralism. It is integral to Bing's design principles to ensure that Bing does not unduly restrict users' ability to access the information they seek. At the same time, Bing recognizes that other fundamental rights and social interests, such as privacy, safety, upholding democracy, public health, and national security, are also vital to our users and to a healthy society, and Bing must balance these interests and rights to maximize benefit to users while minimizing harms. Although strategies may differ, Bing applies these principles and values consistently across subject matters and markets to balance the interests of access to information and free expression with the risks of exposing users to harmful or illegal content.

In order to provide its users with a high-quality, effective and safe search service that appropriately balances rights and access to free information with addressing systemic risks to online safety, the Bing search algorithm and digital safety programs follow the below "Trustworthy Search" principles. These principles guide the product design, experience, search algorithms, and mitigation measures that Bing adopts to address digital safety issues and potential risks or harms arising from the platform.

### **Bing's Trustworthy Search Principles:**



**We aim to provide credible and authoritative results relevant to user queries.**

- We work to provide the highest quality, authoritative content relevant to users' search terms.
- Our goal is to always provide fair, balanced, and comprehensive content. When there are multiple credible perspectives, we try to display them in informative ways.
- When there is no authoritative source, our goal is to avoid promoting bias or potentially misleading information.
- We respect user intent. When a user expresses a clear intent to access specific information, we provide relevant results even if they are less credible, while (as described in more detail below) working to ensure that users are not misled by such search results.



**We promote free and open access to information within the bounds of the law and with respect for local law and other fundamental rights, such as privacy and public safety.**

- We provide open access to as much of the web as possible, but in limited cases we may undertake certain interventions (such as removal of a website or downranking) such as where the content violates local law, or Microsoft's policies.
- When limiting access to content, we strive to ensure our actions are narrowly tailored, so we do not unduly restrict important interests such as the freedom of expression, open access to information, and media pluralism, and we provide transparency regarding our actions.



## We take steps to protect users from harmful and unexpected offensive content.

- We recognize that there are many reasons why someone may want to research or review harmful or controversial content, but also recognize the importance of ensuring users are not inadvertently misled by such content.
- For certain types of content where we identified search results that may include harmful or misleading information, we may provide supplemental information, such as warnings and public service announcements, to inform users about potential risks. We give users control over the type of content they encounter in Bing through features such as SafeSearch and Family Safety.
- Absent a clear intent to access specific content we assume an intent to find high authority results.



## We are transparent about our principles and practices, as well as our decisions and actions.

- We provide users with information about our principles regarding ranking and relevance, and our moderation policies.
- When we limit access to content, where relevant we provide notice to users that content was removed.
- We publish regular transparency reports providing information about the complaints we receive and the actions taken.

## Impact to risk profile and key risk considerations

Bing considers its key risk considerations to be the aspects of the service that are higher vectors for systemic harms than others. Bing's core functionality is to answer users' questions, primarily by connecting them with high quality third-party web content relevant to their queries. In search, users can have many valid reasons for wanting to seek out (legal) content that could be problematic or harmful if encountered in other contexts. As one of the few services that can connect people with as-yet-undiscovered content on the World Wide Web, Bing plays an important role in upholding the fundamental right of free expression and access to information. In this context, one of the highest areas of risk is in ensuring that Bing's algorithmic systems can connect users with the content they are seeking, while ensuring that its content policies and practices regarding this third-party content are sufficiently robust such that users are not misled or harmed by search results.

Bing also collects data from its users, including search terms and related data like location and language preferences that help provide more useful search results, and recognizes its obligation to comply with applicable data protection laws and Microsoft policies to ensure appropriate stewardship of that data and minimize risks of harms related to misuse of personal data. However, given that Bing does not host user content, allow for messaging between users, or allow users to publish content on the platform, and does not engage in extensive personalization or recommendation activities based



on its users' personal information (search results are pulled based on explicit user intent in the form of a search query rather than inferred from behavior), data-related practices in Bing tend to be a lower vector for systemic harms than may be present on other platforms.

Because Bing generally does not allow users to post or share their own content on the platform, unlike in social media, there is limited (if any) risk related to user behavior toward other users on the platform, and limited (if any) risk on the platform that user-generated content (or improper enforcement of rules related to user generated content) give rise to significant systemic harms. Bing does not generally prohibit users from entering search terms that indicate an intent to find harmful, offensive, or potentially illegal content – users can have valid research reasons for seeking out problematic content in search, even content that could be harmful or even illegal in other contexts. This means that while Bing does have established terms and conditions that apply to its users, and does enforce those terms where needed – such as in the case of a user attempting to use illegal CSEAI as a search prompt in visual search, or (as described in the next paragraph) users who violate limitations on generative AI prompts or acceptable use of generative outputs – user behavior on the platform is not a significant vector for harms on search, and enforcement of Bing's terms and conditions plays a relatively minor role in mitigating systemic harms in Bing.

Bing's new generative AI features (Bing Chat and Image Creator) were designed to be a next-generation approach to search: using the power of natural language to make it easier for users to find answers to their research questions, and answer more sophisticated questions, with additional ability to use the tools to inspire new creativity through generating stories, songs, images, or similar outputs. In these new search features, the risks are similar as in traditional search: ensuring the algorithms and AI systems underlying the service are designed to return the content users are seeking, enhanced by content moderation policies and systems designed to ensure users are not harmed by content they seek in search results. Users are still not allowed to post or share content on the platform, however Bing recognizes that there is increased risk that content generated using these tools could be published on third-party platforms in harmful ways, and has implemented terms and conditions, and related enforcement processes, to ensure its tools are used in appropriate ways, and that enforcement actions are fair.

As an ad-supported service, Bing recognizes that there are similar risks in ensuring that the sponsored content appearing on the platform meets Bing's standards for content delivery, in terms of its content policies and enforcement, its ad ranking practices, and data use.

## Bing's approach to and investments in protecting consumers and their fundamental rights online

**Below is a summary of key aspects of the service affecting risk:**

### Search algorithms and recommender systems:

- Complex algorithms generate Bing search results by matching the user's search query with third-party webpages in Bing's index. The majority of content displayed by Bing is in response to an explicit user query rather than based on recommendations based on implied user behavior across the platform. In some scenarios (with appropriate controls and risk assessments), Bing may suggest search topics for a user based on their search history or trending topics, such as via search suggestions, and homepage content on the Images and Videos tabs. Bing designs its ranking and recommendation algorithms to align with core product principles that prioritize high quality, relevant content, and to ensure that users are not offended, harmed, or misled by problematic material in search results. Bing designs and continually improves its algorithms to provide the most comprehensive, relevant, and valuable collection of search results available, including a dedicated "defensive search" team focused on remedying algorithmic failures in high-risk topic areas, such as topics where there are heightened risks to a user's physical, emotional, or financial safety, topics related to current events prone to data void exploitation, and topics that promote hate or advocate violence against individuals or groups.
- Bing's generative AI chat feature is built on LLM technology created by OpenAI and enhanced for use in Bing using Bing's search ranking and relevance algorithms, a unique metaprompt, classifiers, and other safety filters. Bing also offers Bing Image Creator, which is powered by an LLM designed to generate images, enhanced by Bing's classifiers and other safety filters. Bing's implementation of this technology is hosted and operated entirely by Microsoft without direct involvement from its technology provider OpenAI. Bing has worked to provide transparency about how it designed and tested its generative AI features with responsible AI in mind via blog posts and this document: [The New Bing: Our Approach To Responsible AI](#).

### Content Moderation systems:

- Microsoft respects freedom of expression. At the same time, in accordance with Microsoft policies and principles around responsible AI, privacy, digital safety, information integrity, and other critical issues, Bing has developed a safety system including content filtering, operational monitoring, and abuse detection to provide a safe search experience for our users. If Microsoft receives requests to remove content from individuals, businesses, and governments, in limited cases, where quality, safety, user demand, relevant laws, and/or public policy concerns exist, Bing might remove results, inform users of certain risks through public service announcements or warnings, or provide users with options for tailoring their content. Bing limits removal of search results to a narrow set of circumstances and conditions to avoid restricting Bing users' access to relevant information.

- When preventing access to content is warranted, for legal or policy reasons, Bing works to uphold the fundamental right to free expression/access to information by removing content as narrowly as possible. While Bing employs some automated content detection that identifies with a high degree of accuracy crawled content that should be excluded from the index, such as spam and child sexual exploitation and abuse imagery, in most cases automated content detection is not feasible to use for content removal decisions, as most content removal decisions are highly context dependent. As a result, Bing’s content removal practices for the search index generally involve human review. Bing’s support teams are provided with training and oversight in order to ensure removal practices align with Bing’s principles and legal obligations, and have escalation paths for difficult issues, including consultation with local legal experts where needed.
- Bing publishes information regarding the third-party content it removes from search results in biannual [Content Removal Reports](#).

### Applicable Terms and Conditions and their Enforcement:

- Bing’s content moderation activities are largely focused on the third-party website content that is linked to from search results; Bing doesn’t control the operation or design of the indexed websites and has no ability to control what those websites publish via terms and conditions. Bing’s policies regarding third-party web content are described in [How Bing Delivers Search Results](#) and the [Bing Webmaster Guidelines](#).
- Bing’s general terms and conditions governing user behavior are the [Microsoft Services Agreement](#), with supplemental terms for generative AI features. Unlike in services where user generated content is core to the service, moderating user content or behavior on Bing is not core to Bing’s safety story because users do not post or share their own content on the service. Bing generally does not take action against users for their queries (users often have valid reasons to seek out content in search that could be problematic in other contexts). Bing does have and enforce its terms and conditions where necessary, such as taking action against user accounts where authenticated users attempt to use known CSEAI as a search prompt in visual search.
- Users of Bing’s generative AI services are subject to [supplemental terms and conditions and a Code of Conduct](#) governing use of the service, including appeal rights. Users may be banned from the service for attempting to use the service in ways that violate these terms, including users who enter prompts that are designed to bypass Bing’s safety systems or create content that violates the Code of Conduct.

### Systems for Selecting and Presenting Advertisements:

- Ads on Bing are contextually relevant to the search query provided – for example, a search for “flights to Paris” will return ads for airlines. Contextual ads are not targeted based on browsing history or interaction with other websites. Advertising customers are governed by [Terms and Conditions](#) and [Content Policies](#), which are regularly reviewed and updated to ensure they address key areas of concern, and enforced consistently.

### Data-related practices:

- Bing collects some personal data from users to provide and improve the search services, such as user queries, language preferences, and location. Bing also indexes content from the world wide web, which on occasion contains personal data. Bing has an extensive privacy program to ensure it treats personal data in accordance with applicable policies and laws, including impact assessments, transparency, controls, moderation, and data security.

# Methodology

Microsoft Bing began establishing its Article 41 compliance function and planning for this risk assessment in Fall 2022. A Risk Assessment team was formed, consisting of employees from the Bing Compliance Office, Bing Engineering team, Bing Legal team, and the Microsoft Digital Safety Office, all of which play a role in the compliance function. Support and input were provided by policy, compliance and legal teams across the company and via discussions with external experts, including through participation in workshops and events hosted by the Digital Trust and Safety Partnership and Global Network Initiative (GNI), and the Commission's DSA Stakeholder event.

## Overview of Process

### STAGE 1

---

**Risk Definition and Discovery.** As the initial step, the Risk Assessment team worked to identify possible systemic risks on the Bing platform, as defined by reference to Article 34 of the DSA, as well as the [Safe Assessment framework published by the Digital Trust and Safety Partnership \(DTSP\), of which Microsoft is a founding member](#). In this step the Risk Assessment team engaged key product stakeholders and performed initial information discovery on the service's practices in Spring 2023, to gain an initial understanding of the operational and risk landscape in which Microsoft Bing is operating. This stage included data-gathering through workshops, discussions, and document review to complete the Safe Assessment questionnaire. This initial information discovery provided the baseline understanding that drove downstream assessment activities.

### STAGE 2

---

**Risk Analysis.** Grounded in the information and artifacts gathered during the "Discover" stage, in spring of 2023 the Risk Assessment team identified and prioritized specific best practices and risk areas to focus on during the assessment where risk is highest on Bing. The Risk Assessment team relied on a cross-company slate of experts to help analyze the areas of risk on Bing using the DTSP framework and an analysis of the probability and severity of the systemic risks enumerated in the DSA, as well as consultations with external experts, including DSA-specific forums convened by DTSP and the Global Network Initiative, and the DSA Stakeholder event hosted by the Commission.

## STAGE 3

---

**Risk Mitigation.** As a next step, the Risk Assessment team assessed the efficacy of Bing’s risk mitigations in light of the 35 DTSP Best Practices and the enumerated systemic risks in the DSA. In this stage the Risk Assessment team gathered additional documentation of risk mitigations, held additional workshops with stakeholders for deeper discussions regarding practices, processes, and tools, and assessed the efficacy of Bing’s safety systems, applying a Maturity Rating (scale of 1-5) for each of the 35 Best Practices, as well as a residual risk score (following mitigations) for each of the enumerated systemic risks. This analysis is based on knowledge of consensus-based approaches to safety, grounded in industry best practices, feedback from external stakeholders such as civil society and academics, worldwide regulations, and public commitments.

## STAGE 4

---

**Substantiation.** As the next stage, the Risk Assessment team analyzed the testing of the risk mitigation practices identified and assessed in above stages. Appropriate testing mechanisms ensure that the safety practices including people, processes and technologies are working and effective. The Risk Assessment team collected information regarding and reviewed the efficacy of Bing’s internal metrics reporting to understand how the testing results are feeding back into safety practice improvement and how the effectiveness is being evaluated, culminating in the production of this Report in summer 2023.

### Assessment Tools

**Safe Assessment.** The Bing Risk Assessment team first followed the 5-step [DTSP Safe Assessment methodology](#) to analyze and evaluate the service’s current risk landscape and the maturity level of the safety program through the lens of the 35 DTSP Best Practices, assigning a maturity rating for Bing’s systems in each Best Practice, and identifying areas of highest risk. Focusing on the most potentially significant issues allows us to tailor the analysis to the specific service and category.

**Assessment of Systemic Risks.** The Risk Assessment team also analyzed the primary features of Bing’s service through the lens of the possible systemic risks, as enumerated in the Digital Services Act by reference to the Charter of Fundamental Rights of the European Union (2000/C 364/01):

<b>Illegal Content Risks</b>	Risk of dissemination of illegal content through the service.
<b>Risks to Users’ Fundamental Rights</b>	Actual or foreseeable effects on users’ exercising of fundamental human rights, including: <ul style="list-style-type: none"><li>• Rights to human dignity,</li><li>• Respect for private and family life,</li><li>• Rights to the protection of personal data,</li></ul>

	<ul style="list-style-type: none"> <li>• Rights to freedom of expression and information, including the freedom and pluralism of the media,</li> <li>• Rights to non-discrimination,</li> <li>• Rights of minor children,</li> <li>• Rights to consumer protection, and</li> <li>• Freedom from discrimination.</li> </ul>
<b>Risks to Civic Integrity and Public Security</b>	Actual or foreseeable effects on civic discourse and electoral processes, and public security.
<b>Societal Risks</b>	Actual or foreseeable effects in relation to societal issues such as gender-based violence, protection of public health and protection of minors, and serious negative consequences to a person’s physical and mental well-being.

In identifying how these systemic risks arise in the Bing product, the Bing Risk Assessment team collaborated with a slate of subject-matter experts across Microsoft, who are able to draw from their experience and feedback from their relationships with third-party stakeholders. These reviews took into account how these risks are affected by (a) the design of Bing’s recommender systems and any other relevant algorithmic systems; (b) its content moderation systems; (c) applicable terms and conditions and their enforcement; (d) systems for selecting and presenting advertisements; and (e) data related practices.

Based on these analyses, the Risk Assessment team worked to systematically quantify the inherent risk by looking to the *probability* of the risk (based on transparency reports and other metrics) and the *severity* of the risk to users or society (decoupled from the product, severity looks at the complexity of the risk, the gravity of the risk, and the scale of the risk), and the efficacy of Bing’s mitigations, in order to determine what residual risk remains and whether additional mitigations are needed.

### Report Review Process and Approvals

As the final step in the process, the Risk Assessment team compiled this Report, which summarizes the findings.

The Report as well as the supporting documentation developed by the Risk Assessment team, were circulated and reviewed across the rest of the Article 41 compliance function. The function within Microsoft is multilayered, and includes the following:

- The Microsoft Digital Safety Office, a centralized team within the Privacy, Safety, and Regulatory Affairs department at Microsoft, which provides independent coordination and oversight of digital safety activities across Microsoft products and services, including the Risk Assessment. The

Article 41 head of the compliance function, the General Manager and Associate General Counsel for Content Regulation and Governance, and an Article 41 compliance officer sit within this team.

- The Bing Compliance team, which is responsible for ensuring implementation of DSA requirements and supporting audits. This team is accountable for ensuring Risk Assessments are completed and maintained, and that mitigation plans are progressing. The head of this team is an Article 41 compliance officer.
- The Bing Engineering team, which executes on compliance activities and operations. This team is responsible for: operationalizing Microsoft Digital Safety standards; maintaining the process for handling and responding to complaints, requests, and inquiries regarding digital safety and responsible artificial intelligence; providing information for the Risk Assessment, and enacting mitigations.
- The Bing Legal team provides support for the Bing Engineering team in developing product requirements for DSA compliance and risk mitigations.

Members of leadership for each of the aforementioned four groups have reviewed and approved the contents of this Risk Assessment report.

## Bing Service at a Glance

### DTSP Assessment Level 3

We have leveraged the tailoring approach outlined in the DTSP Safe Framework to establish an assessment level for evaluating risk across Bing services. This DTSP Safe Framework takes into account the distinct nuances and risks for the organization and product utilizing a set of common criteria.

These components include:

1. The organizational size and scale of the service;
2. The product or digital service impact; and
3. The business landscape considerations.

Bing evaluated each of the three components using the attributes outlined in the DTSP tailoring framework for calculating risk levels and ultimately the appropriate category for performing ongoing assessments. Following this analysis, Bing concluded as follows:

- The organizational size and scale risk rating is based on global revenue from the previous year as well as the total number of employees. Based on recent fiscal year data, Bing's organization size and scale would be considered "high" risk under the Safe Assessment tailoring framework.
- Risk levels are further defined by user volume and an evaluation of key risk drivers for the Bing service, including: the product's purpose, its intended audience (age ranges/user type), its



features and the markets where it's offered. Bing's search service includes a number of the key risk factors called out in the Safe Assessment questionnaire, including that Bing is a consumer product offered to all ages, that Bing is offered in well more than the DTSP benchmark of 20 languages and 30 countries, offers some curated recommendations to users (e.g., search suggestions) and has released in the past year novel generative AI features. In summary, based on Bing's user volume and the number of key risk drivers, Bing would be considered "high" risk under the Safe tailoring framework.

- Finally, Bing considered the business landscape in which Bing services are operating, namely: market expansion, rapid product changes, new M&A activity, joint venture, or partnership, prior assessments/audit, user growth trajectory, rapid social or political changes. Considering the business landscape applicable to Bing services in conjunction with the launch of Bing's new generative AI features, and the general expansion of AI across the industry, Bing should continue to be considered "high" impact under the Safe tailoring framework.

Taking into account the attributes for evaluating risk across Bing's organizational size and scale; the product or digital service impact; and the business landscape considerations Bing determined it should undertake a **DTSP level 3 Assessment**.

### High level overview of key Bing surfaces, features, and risks.

Category	Microsoft Bing
<b>Service Classification</b>	Microsoft Bing is an <b>online search engine</b> service available on mobile devices, PCs, and web browsers.
<b>User Base</b>	<ul style="list-style-type: none"> <li>• Bing's average monthly active user (MAU) base in the EU is 119 million. A "monthly active user" on Bing is defined as an end user with or without a Microsoft Account that takes an intentional action on Bing.com or in the Bing app, such as entering a search query into Bing.</li> <li>• Bing has both authenticated and unauthenticated users. With the exception of optional features that require authentication to obtain additional benefits, like Microsoft Rewards or Rebates, and full access to Bing's generative AI features, all Bing features are available to both authenticated and unauthenticated users. Thus, Bing does not distinguish between authenticated and unauthenticated users in its user base calculations.</li> <li>• Most of Bing's services are available to users of all ages (authenticated users under the age of consent in their local region require parental consent to use the service). Certain features, including Bing's full generative AI features, are not available to minor users under the age of consent in the local market (minimum age 13 globally).</li> <li>• Bing's average monthly active users is not an accurate proxy for the reach or impact of Bing, as a sizable portion of Bing users enter sparingly</li> </ul>

	<p>few queries per month. Many users are “light users” that query Bing only a few days per month. These are, for example, people who primarily use other search engines (or social media in lieu of search engines) but, who once or twice a month enter queries into a Bing interface, such as through the Windows Start menu or Microsoft Edge browser.</p>
<p><b>Market and Geographic Reach</b></p>	<ul style="list-style-type: none"> <li>• Bing offers over 200 market versions and is offered in over 100 different languages.</li> <li>• The product is available globally as permitted; it follows U.S. laws and blocks the service where export restrictions require (e.g., North Korea). New features may be released in limited markets initially. Index and user data are stored in North America.</li> <li>• Per <a href="#">StatCounter</a>, as of 9 August 2023, Bing’s global search engine market was around 3% (Google, the largest search company, has approximately 92% of the search market). Search engine market share in the EU parallels these global market shares.</li> </ul>
<p><b>User Interactions</b></p>	<p>Bing users do not have the ability to interact with one another within the Bing ecosystem.</p>
<p><b>Summary of Key Bing Surfaces, Features and Risks</b></p>	<p><b>Bing Core Search</b></p> <p>The core functionality of Bing is to enable users to find third-party web content in response to their search queries. To do this, Bing crawls the web to build an index of pages (or URLs) to display as a set of search results relevant to a user-initiated search or action. The content of these pages may include images, videos, documents, and other items. Bing does not host the content on the websites that appears in search results, but at most caches third-party pages in order to more quickly deliver search results to users. Bing has no control over the operation, design, or contents of the materials on third-party websites. As long as a website makes content available on the internet and to search engine crawlers, the content will generally be available through Bing and other search engines. Bing does not allow users to create, upload, or share their own content on the service as part of core search – rather it allows users to find and consume information.<sup>2</sup> Below is a summary of the key systemic risks that arise in Bing core search features. The steps Bing has taken to mitigate these risks are described in detail below in the “Summary of Results” section.</p> <ul style="list-style-type: none"> <li>• There is a risk that Bing over-limits access to content (via safety systems or other moderation activities) such that it negatively impacts the fundamental right to access information/free expression.</li> </ul>

<sup>2</sup> Bing also offers API services, allowing third parties to add search capabilities to their services powered by Bing, and offerings for enterprises such as Microsoft Search in Bing and Bing Chat for Enterprise. These offerings generally adhere to the same safety standards as Bing itself. Bing APIs are offered to third parties via standard contractual terms.

- It is possible that third-party content appearing in search results could be illegal, unsafe, misleading, or otherwise offensive in ways that negatively impact the fundamental rights like nondiscrimination, privacy, or human dignity of individual victims. Such materials could also negatively impact other important public interests such as secure democratic processes, public health or safety.
- There is risk of users being unexpectedly exposed to third-party content that negatively impacts their own fundamental rights, such as if content is hateful or discriminatory, or content that could cause negative effects on their physical or mental health or cause them to fall victim to scams or fraud.
- Minor users could be exposed to content that is not in their best interests, such as adult content.
- User queries or other data retained by Microsoft for providing and improving the product could be used in ways not aligned with legal requirements (e.g., GDPR) or user expectations, or could be subject to data breach.
- There are risks that bad actors could exploit [data voids](#) to provide inaccurate or misleading information undermining social institutions on third-party websites, which may appear in search results for related search terms.
- There is also risk that bad actors or information sources may attempt to manipulate the Bing algorithms so that lower authority content appears. Taking advantage of or manipulating Bing algorithms could amplify the spread of illegal or problematic content.

### **Bing Enhanced Search Features**

In addition to the core search services, Bing has a number of enhanced features designed to help users find the content they are looking for more quickly, such as answers (including indicators of content provenance, fact checks, and public service announcements), topic-specific verticals (e.g., image, video, shopping, or maps results), and search suggestions. In some cases these features may enhance the likelihood of some systemic risks, which are outlined below. How Bing mitigates these risks is detailed in the Summary of Results section.

- Unlike the core search product where users seek access to content through search queries, and in response have access to pages of links in search results to help in their research, in enhanced search features like answers users are more likely to consider the limited set of results provided in answers to be authoritative. Such answers could incorrectly

point to low authority, harmful, misleading, or offensive material that could lead to harm to users or social institutions.

- Bing’s safety systems may over-block the appearance of answers in ways that could be biased towards particular viewpoints or limit users’ ability to access information.
- Specific search verticals such as News, which limit the type of content available in that vertical based on particular criteria/standards for inclusion, could limit access to relevant content in ways that introduce bias into the system, prevent websites from connecting with potential customers, or undermine media pluralism. Content provenance indicators could similarly introduce bias or undermine media pluralism.
- Search suggestions or similar recommended content could lead users to harmful content. Conversely, unduly limiting the appearance of search suggestions in an attempt to protect users could undermine free expression/access to information.

### **Bing Generative AI Features**

Bing’s new generative AI search features are designed to provide a next-generation search experience, combining the power of LLM technology with the power of search. Bing Chat is an AI-enhanced conversational search experience that uses AI to allow for more sophisticated search queries and interactions, and to inspire creative works such as poems, jokes, or letters. Bing Image Creator is a similar tool that enables users to generate images for personal use based on text prompts. Because Bing’s generative AI features build on Bing’s search systems, many of the risks are the same as outlined in core search. However in some cases these features may enhance the likelihood of some systemic risks, which are outlined below. How Bing mitigates these risks is detailed in the Summary of Results section.

- Despite in-product disclaimers, FAQs, and explainers, there is risk that users, who are not yet entirely familiar with the abilities and limitations of LLM technology, could put inordinate trust in outputs from chat features and not verify the content in third-party links the way they would in standard search, which could exacerbate harms to fundamental rights or key social interests that may result from content appearing in search results.
- There is risk that generative AI features will incorrectly summarize web content, or generate content, that is inaccurate, harmful, misleading, privacy-impacting, or potentially illegal (e.g., defamation or copyright infringement).
- Bing’s generative AI features are in part powered by LLMs developed by its technology partner, Open AI. While Bing operates the service

entirely independently of OpenAI, and Bing has added layers of unique features, including combining with search algorithms, adding a bespoke metaprompt, as well as adding its own safety filters/classifiers that make the Bing service unique, Bing relies on OpenAI to create and update the core LLMs, which may in some cases trigger discrepancies in results between traditional search results and chat results that may require additional interventions.

- Bing does not provide a means for users to disseminate generated content privately or publicly within the service, but a user could copy the generated outputs and share them on third-party platforms in ways that are designed to deceive, mislead or harm recipients.

### Advertising

Bing search is available to users free of charge; it is monetized almost entirely through search advertising. When a user conducts a search on Bing, the user is shown both organic (unpaid) results and, when relevant, ads. In search, ads shown to users are relevant to the query/prompt typed. Advertisements on Bing are governed by the [Microsoft Advertising Terms and Conditions](#) and [Content Policy](#). Many of the risks in Bing's advertising features are the same as outlined in core search. However in some cases these features may enhance the likelihood of some systemic risks, which are outlined below. How Bing mitigates these risks is detailed in the Summary of Results section.

- Despite terms and conditions prohibiting such ads, and regular enforcement of ad policies through proactive means and in response to reactive reporting, there is a risk that advertisements delivered to users may include, deceptive, harmful or potential illegal content that could lead to harm to users' fundamental rights or important societal interests. Minor users could be particularly susceptible to harmful ad messaging.
- Inconsistent enforcement of ad policies could effectively result in discrimination against advertising partners.
- Audience targeting categories offered to advertisers could enable discriminatory display of advertising, in some cases limiting certain users' ability to obtain products and services at a favorable rate.
- Failing to comply with legal requirements or policies regarding personal data use or failing to respect user choices or providing inadequate controls, could undermine personal data protection.

# Summary of Results

## Summary of Findings: Key Risks and Mitigations

The following section summarizes the output of the Risk Assessment team's analysis of each systemic risk identified in Article 35 of the DSA, including a summary of how the risk presents on the platform, a calculation of the inherent risk (based on probability x severity), relevant mitigations, residual risk score, and additional actions needed, if any.

### The dissemination of illegal content through the service (CSEAI, hate speech, other criminal offenses, conduct of illegal activities like the sale of prohibited products or illegally traded animals)

#### Summary of risks

---

Bing users generally cannot share their own content with the broader user base on the service, so users cannot disseminate illegal materials through the service. However, third-party website content linked in search results may on occasion include illegal materials. Bing's ranking algorithms are designed to avoid ranking low quality content (such as illegal materials) high in search results but there may be occasions where algorithmic sorting is not sufficient to prevent the display of illegal materials, such as in a data void, where users are expressly seeking such content, or where websites or users try to intentionally manipulate search results. Bing's primary mechanism for moderating content in search results is reactive, based on legal removal orders from governments, trusted flaggers, or affected parties. While Bing's reporting and reactive removal processes are robust, some reports of illegal materials may be mishandled. Bing does not generally use automated content detection technology to remove content except in the case of PhotoDNA to prevent known CSEAI from entering the index, and improper search engine optimization (spam). While PhotoDNA is extremely effective, and Bing supplements this automated detection with additional data from external partners such as the Internet Watch Foundation, Bing's ability to remove such content is limited to the information available. New content that has not yet been identified may still appear on Bing. Similarly, Bing's advertising partner, Microsoft Advertising, has robust terms and conditions and content policies prohibiting ads promoting illegal materials, and reactive reporting and removal processes, but these processes are not infallible. In Bing's generative AI services, user prompts may violate Bing's terms of use or attempt to bypass technical restrictions in an effort to create illegal materials. Bing works to identify and take action against user actions that violate its terms, but given that materials created through its generative AI services would necessarily be distributed on third-party sites (due to a lack of ability to share on the Bing site), Bing may not always be able to detect problematic behavior. Bing's data-related practices are not a risk to dissemination of illegal content.

## Inherent risk score

---

The distribution of illegal materials online is a social issue of significant severity and complexity and encountering illegal materials online can lead to harm to individual users and to victims of crimes. Per Bing's content removal reports, and in light of the size of the Bing index, far less than 1% of the index includes illegal materials (the vast majority of which are reports of copyright infringement). The inherent risk related to dissemination of illegal materials via Bing, absent mitigations, is **moderate**.

## Key mitigations

---

- Bing invests significant time and resources into ensuring its crawlers and algorithms prioritize high quality content to avoid inadvertently returning illegal materials to users. Bing works to ensure its interventions are effective in all languages and regions in which Bing offers the service. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated Objectives and Key Results (OKRs), and through ingestion of user and stakeholder feedback. **See Best Practices PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Bing also adds information to help ensure users are well informed and not misled or harmed by materials appearing in search results, such as public service announcements on queries likely to lead to illicit materials, warnings on sites identified as likely to contain illegal materials, such as pharmaceuticals, prevents autosuggestions likely to lead to illegal materials, and SafeSearch by default prevents the display of adult imagery or gore. Bing works to ensure such additional interventions are available across the markets and languages in which Bing is offered. **See Best Practice PD1, PD2, PD3, PD4, PD5, PD6, PD7, PD8, PG1, PG3, PG5, PE 1.1, PE1.3, PE6.1, PE3, PT5, PE8, PE9, PI1, and P12.**
- For other types of illegal content, Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal materials from governments, users, or others, meeting required legal timeframes for content removal where applicable. Support teams are trained on applicable legal requirements and provided escalation paths where needed for complex issues, including to local legal support as needed. The Bing team also relies on Microsoft's extensive team of subject matter experts on global regulatory issues to ensure awareness of new obligations that may arise in markets where it operates. Bing engages in regular tests and audits of its system to ensure its ability to respond. **See Best Practice PD1, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, P12, P13, P14, P15, PT1, PT2, and PT3.**
- Bing users or third-party webmasters who are impacted by content moderation decisions are provided notice and redress opportunities. **See Best Practice PT1, PT2, PT3, and PT5.**
- Bing proactively uses hash-matching technologies (including PhotoDNA and MD5) to detect matches to known CSEAI, to avoid it from appearing in the index using PhotoDNA and via threat intelligence provided by third party expert partners. **See Best Practice PD1, PD2, PD5, PG5, PG6, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, P6.1, PI2, P13, P14, P15, PT1, PT2, and PT3.**

- 
- Bing and cross-Microsoft SMEs engage with external stakeholders in areas of key policy priorities, such as terrorist and violent extremist content, CSEAI, and copyright infringement, to ensure that our internal policies, practices, and standards are addressing their key concerns. [See Best Practice PD4, PD5, PD6, PD7, PD8, PG1, PG5, PG6, PE1.1., PE1.2, PE6.1, PE6.3, PE7, PE8, PE9, PI2, and PI4.](#)
  - Bing is transparent about its policies and actions with respect to user and webmaster content, and provides these disclosures in all relevant EU member state languages. [See Best Practice PG1, PG3, PT1, PT2, PT3, PT4, and PT5.](#)
  - These same mitigations apply to Bing’s generative AI features, with additional enhanced safety features such as classifiers, filters, and a bespoke metaprompt that further limit the likelihood of illegal content appearing in Bing Chat/Image Creator features. [See Best Practice PD1, PD2, PD4, PD5, PD6, PD7, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, PI2, PI3, PI4, PI5, PT1, PT2, PT3, PT4, and PT5.](#)
  - Microsoft Advertising, which powers ads on Bing, has clear and regularly enforced content policies and practices that prevent advertisements negatively impacting consumer protection. The [Microsoft Advertising Policies](#) set out the requirements for ad content including criteria upon which ad content will be removed. Microsoft requires our advertisers and partners to comply with our policies throughout their use of our services. The Microsoft Advertising Policies prohibit advertising content that is misleading, deceptive, and fraudulent. Microsoft Advertising also has a set of Relevance and Quality Policies to manage the relevancy and quality of the advertisements that it serves through its advertising network. These policies deter advertisers from luring users onto sites using questionable or misleading tactics. [See Best Practice PD1, PD2, PG1, PG2, PG3, and PG6.](#)
  - Bing users may report advertising for takedown through the “feedback” form found on the Bing’s pages. Additionally, ads that may be promoting illegal content or that may appear to violate the Microsoft Advertising terms or policies may be submitted through one of these forms: Low quality ad submission & escalation or Intellectual Property Concern Form. [See Best Practice PD8, PD9, PG1, PG2, PG3, PG4, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE5, PE6, PE6.1, PE6.2, PE6.3, PE7, PT1, PT2, and PT3.](#)
  - Microsoft Advertising is constantly improving its internal systems and process to ensure the advertising content served is legal, clear, truthful, and accurate. Advertising employs a robust filtration system to detect robotic traffic and other harmful cyber-attacks. This system uses various algorithmic systems to detect and neutralize illegal, invalid, or malicious online traffic which may arise from or result in click fraud, phishing, malware, or account compromise. Microsoft Advertising has several teams of security engineers, support agents, and traffic quality professionals dedicated to continually developing and improving this traffic filtration and network monitoring system. Microsoft Advertising’s support teams work closely with its advertisers to review complaints of suspicious online activity, and they work across internal teams to verify data accuracy and integrity. [See Best Practice PD1, PD2, PD3, PD4, PD6, PD7, and PE4.](#)



## Severity post-mitigations: Low

---

Bing's mitigations are designed to minimize the risk that illegal materials will be disseminated via Bing while balancing other fundamental rights and interests such as free expression/access to information. Bing's policies and practices supporting these mitigations are sufficiently mature to address these risks.

## Additional mitigations needed

---

Search engine results are constantly changing because the webpages within an index may be added, removed, or modified. In addition, the nature of search itself—especially with the rise of generative AI—is evolving. Bing must continue to invest in proactive approaches to reducing the harms created by the dissemination of illegal materials online, and to continue to invest in the systems it uses to identify and remove such contents where applicable. Bing should do this by taking a leadership role in the industry's approach to integrating safety and risk reduction into the future of search, as well as expanding mechanisms for gathering and seeking out feedback from users, industry partners, and researchers, as well as expanding Bing's ability to share data with researchers to identify additional solutions. Bing should continue to identify additional investments that can specifically improve its systems across all EU countries and languages. The deliverable for this exercise is to provide additional transparency reporting to the public, further define internal accountabilities, and ensure highly detailed documentation of policies and procedures.

## The actual or foreseeable negative effects for the exercise of fundamental rights, including:

### A. Rights to human dignity

---

**Summary of risks:** The right to human dignity is broad and can be negatively affected by a wide variety of legal and illegal materials and actions, including for example through hateful, discriminatory or stereotyping content, content with impacts to user privacy, exploitative content, misinformation, or reputation-harming content. As Bing works to index as much of the World Wide Web as possible and seeks to answer any question users may ask, some material in the index could negatively impact the right to human dignity and could be returned to users in search results or in generative AI features. Bing's algorithms are designed to minimize the risk that such content appears high in search results, however in some cases algorithms may not function as intended, a user may expressly be seeking harmful content, or results may have been subjected to intentional manipulations, including disinformation or spam campaigns, and such content may be displayed to a user. Bing's algorithms and safety systems, including in auto-suggestions and in generative AI features, may also inadvertently be biased towards certain viewpoints or persons, reflect stereotypes, lead users to harmful content, or inadvertently prevent users from encountering material they seek that is important to engaging in civic life or exercising other rights. Bing's generative AI features could be subjected to abuse by malicious actors who violate safety systems in order to create harmful materials

---

that are distributed on third-party platforms. Advertisements could be displayed to users in discriminatory ways, or encourage users to purchase harmful products or services.

**Inherent risk score:** The right to human dignity is broad; a social issue of significant severity and complexity. Encountering materials that negatively impact the right to human dignity in online search engines can lead to harm to individual users and to third party victims and undermine social order. However, per Bing’s content removal reports, and in light of the size of the Bing index, far less than 1% of the index includes materials subject to removal under Bing’s policies and legal obligations (the vast majority of which are reports of copyright infringement). Bing also tracks the number of queries subjected to defensive interventions out of the total number of queries submitted to the service, which are an additional indicator of the frequency with which users engage with potentially problematic content across the systemic risk areas: these queries constitute less than 1% of the total number of queries in the search services. The inherent risk related to the right to human dignity via Bing, absent mitigations, is **moderate**.

**Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring its crawlers and algorithms prioritize high quality content to avoid inadvertently returning low quality or harmful content to users. Bing’s ranking principles consider low quality content to include “pages that call for violence, name-calling, offensive statements, or use derogatory language to make a point are generally considered low quality.” Bing works to ensure its processes are equally effective across all languages and markets in which Bing is offered. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs, and ingests feedback from users and via social listening systems to continue to make improvements. [See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.](#)
- Bing has a defensive search team that is dedicated to identifying and remedying, via targeted algorithmic interventions, high impact issues in search results, such as misinformation, hateful speech, and other problematic content that could negatively impact human dignity. Bing regularly reviews the efficacy of its interventions to ensure that they are performing as expected and not inadvertently introducing additional bias or other harm. Bing works to ensure its defensive interventions are effective across languages and markets where Bing is offered. [See Best Practice PD1, PD2, PD4, PG1, PE1, PE8, PI1, P13.](#)
- Bing also adds information to help ensure users are well informed and not misled or harmed by materials appearing in search results, such as answers pointing users to authoritative sources when search results lack high authority content, prevents autosuggestions likely to lead to controversial or harmful topics, and defaults SafeSearch features to prevent the display of adult content or gore. Bing works to ensure these features are available in relevant languages and markets across the EU. Bing uses established metrics to review the efficacy of its interventions and identify areas for improvements. [See Best Practice PD1, PD2, PD3, PD4, PD5, PD6, PD7, PD8, PG1, PG3, PG5, PE 1.1, PE1.3, PE6.1, PE3, PT5, PE8, PE9, PI1, and P12.](#)

- 
- Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal materials and other problematic content (including harmful or offensive content, exposed personal information, CSEAI, nonconsensual intimate imagery (NCII), etc.) from governments, users, or other stakeholders, meeting required legal timeframes for content removal, where applicable. Support teams are trained on applicable requirements and policies and provided escalation paths where needed for complex issues, including to local legal experts. The Bing team also relies on Microsoft’s extensive team of subject matter experts and external experts to identify and quickly address emerging concerns. Bing engages in regular tests and audits of its system to ensure its ability to respond. [See Best Practice PD1, PD3, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, P12, P13, P14, P15, PT1, PT2, and PT3.](#)
  - Bing takes steps to prevent the creation of suggestions that could undermine the right to human dignity and provides users with easy mechanisms to report problematic suggestions for removal. [See Best Practice PD1, PD8, PD9.](#)
  - Bing proactively uses hash-matching technologies (including PhotoDNA and MD5) to detect matches to known CSEAI, to avoid it from appearing in the index using PhotoDNA and via threat intelligence provided by third party expert partners. [See Best Practice PD1, PD2, PD5, PG5, PG6, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, P6.1, PI2, P13, P14, P15, PT1, PT2, and PT3.](#)
  - Bing and cross-Microsoft SMEs engage with external stakeholders in areas of key policy priorities, such as terrorist and violent extremist content, information integrity, responsible AI, CSEAI, and copyright and trademark infringement, to ensure that our internal policies, practices, and standards are addressing key concerns. [See Best Practice PD1, PD2, PG1, PG3, PG5, PG6.](#)
  - With respect to risks related to misinformation and disinformation, in addition to its internal threat identification and safety systems, Bing is a signatory to the EU’s Code of Practice on Disinformation, which requires biannual reporting and compliance with anti-disinformation measures to improve Bing’s ability to fight misinformation across the EU, including commitments to providing users with indicators of content provenance, fact checks, and researcher access to data. Bing also regularly collaborates with other signatories through working groups to address emerging issues and threats as well as best practices in combatting disinformation. [See Best Practice PG5, PE9, PT1, PT4.](#)
  - Bing relies on Microsoft’s extensive cross-company compliance infrastructure as part of its new feature review process; Bing uses these cross-company policies and standards to proactively review new products and features to ensure they meet the bar for Microsoft’s policy commitments and legal responsibilities in key areas such as privacy, accessibility, digital safety, and responsible AI. [See Best Practice PD2, PD3, PD4, PD5, PD6.](#)
  - Bing is transparent about its policies and actions with respect to user and webmaster content, and makes these documents available in all member state languages. [See Best Practice PT1, PT2, PT3, PT4, PT5.](#)
  - These same mitigations apply to Bing’s generative AI features, which are further supported by enhanced safety features such as classifiers, filters, and bespoke metaprompts that further limit

---

the likelihood of harmful content appearing in Bing Chat/Image Creator features. Bing has engaged in extensive responsible AI reviews assessing generative AI features in order to ensure outputs are not biased, discriminatory, or likely to generate harmful content. Bing is continually working to ensure that its generative features do not over-block outputs so that users are able to access the information they seek and strives to be transparent about its learnings and how it has evolved the product over time to address the concerns it has identified. Bing's work as a signatory on the COP on Disinformation also includes addressing misinformation risks in generative AI features. [See Best Practice PD1, PG1, PG3, PE3, PE6.](#)

- Microsoft Advertising, who powers ads on Bing, has clear and regularly enforced content policies and practices that prevent advertisements negatively impacting human dignity. The Microsoft Advertising Policies set out the requirements for ad content including criteria upon which ad content will be removed. Microsoft requires our advertisers and partners to comply with our policies throughout their use of our services. Microsoft Advertising also has a set of Relevance and Quality Policies to manage the relevancy and quality of the advertisements that it serves through its advertising network. These policies deter advertisers from luring users onto sites using questionable or misleading tactics. [See Best Practice PG1.](#)
- Bing users may report advertising for takedown through the "feedback" form found on the Bing's pages, as well as specific content reporting channels. [See Best Practice PD1, PD8, PG2, PG4.](#)
- Microsoft Advertising is constantly improving its internal systems and process to ensure the advertising content served is legal, clear, truthful, and accurate. Advertising employs a robust filtration system to detect robotic traffic and other harmful cyber-attacks. Microsoft Advertising has several teams of security engineers, support agents, and traffic quality professionals dedicated to continually developing and improving this traffic filtration and network monitoring system. Microsoft Advertising's support teams work closely with its advertisers to review complaints around suspicious online activity, and they work across internal teams to verify data accuracy and integrity. [See Best Practice PD1, PD2, PD3, PD4, PD6, PD7, and PE4.](#)

**Residual risk score:** **Low.** Bing's mitigations are designed to minimize the risk that materials negatively impacting human dignity will be disseminated via Bing while balancing other fundamental rights and interests such as free expression/access to information. Bing's policies and practices supporting these mitigations are sufficiently mature to address these risks.

**Additional mitigations needed:** Search engine results are constantly changing because the webpages within an index may be added, removed, or modified. In addition, the nature of search itself—especially with the rise of generative artificial intelligence—is evolving. Bing must continue to invest in proactive approaches to reducing the harms to human dignity that arise from materials displayed in search, and continue to invest in its reactive reporting and removal infrastructure. Bing should do this by taking a leadership role in the industry's approach to integrating safety and risk reduction into the future of search, as well as expanding mechanisms for gathering and seeking out feedback from users, civil society, experts, industry partners, and researchers, including expanding our ability to share data with researchers. Bing should continue to seek out additional partnerships with external experts who can help supplement Bing's knowledge of local areas of significant concern. Bing

---

should continue to identify additional investments that can specifically improve its systems across all EU countries and languages. The deliverable for this exercise is to provide additional transparency reporting to the public, further define internal accountabilities, and ensure highly detailed documentation of policies and procedures.

## B. Respect for private and family life

---

**Summary of risks:** Persons whose personal information is available on public webpages may appear in search results, or in the output of generative AI features, in ways that undermine respect for private and family life. For example, the search index may enable easy retrieval of information about an individual that falls within their private sphere and that the individual did not intend or consent to make available publicly, such as non-consensual intimate imagery (NCII) (also known as “revenge porn”), or extremely sensitive personal information that could create risks of identity thefts such as credit card numbers, medical records, etc. Information about individuals contained in news websites or other web content could reveal personal information that is excessive, irrelevant, outdated, or incorrect. This could be exacerbated by ranking decisions that cause such content to be ranked high in search results. Bing’s generative AI features could be abused to create harmful content about individuals or misuse an individual’s image or likeness. Autosuggestions or generative AI outputs could also reinforce baseless or malicious accusations about a user that appear on third-party websites, or AI features could inaccurately interpret source data to provide incorrect or misleading information about an individual. Search history data, if inadvertently disclosed, could reveal an individual’s beliefs, relationships, sexuality, medical issues, or other private information.

**Inherent risk score:** The right to family and private life is an individual and social issue of significant severity and complexity, and encountering materials that negatively impact the right to family and private life in online search engines can lead to harm to individual users (such as limited employment prospects or emotional distress) and to third party victims (such as victims of stalking or harassment). However, per Bing’s content removal reports, and in light of the size of the Bing index, far less than 1% of the index includes materials subject to removal under Bing’s policies and practices regarding inappropriately published personal information including removals under the Right to Be Forgotten (RTBF). The inherent risk related to the right to private and family life via Bing, absent mitigations, is **moderate**.

### Summary of key risk mitigations:

- Bing has a robust privacy and security infrastructure that includes full-time professional compliance managers, requires pre-launch feature reviews and mitigations, adherence to strict standards for data handling and security by internal employees and vendors, training for all employees, and completion of Data Protection Impact Assessments to ensure appropriate risk mitigations. [See Best Practice PD2, PD3, PD5, PD6, and PE1.](#)
- Users of Bing are given controls over collection and use of personal data on the service, as well as controls that allow them to exercise their data subject rights to view, access, export and delete

---

personal data held by Microsoft. Bing maintains user data in accordance with Microsoft security standards. [See Best Practice PG3, and PD9.](#)

- Bing allows users and other stakeholders to easily report private content appearing in search results that violates local laws, including in the EU the Right to Be Forgotten, and Bing policies prohibiting the indexing of private content published without consent such as NCII, credit cards, or other private information, or images of minor users. Bing’s content review teams are trained on legal obligations and Bing policies in order to efficiently action such requests. Bing engages in regular tests and audits of its system to ensure its ability to respond. [See Best Practice PE5.](#)
- Bing’s terms of use governing activity on its generative AI features prohibit the use of the service to violate others’ privacy. Safety systems in generative AI features are designed to prevent abuse, such as blurring faces in images before they are used as prompts in visual search and preventing the generation of images with individual faces. Bing labels all images as generated by Bing Image Creator to limit the possibility of misuse. [See Best Practice PE3.](#)
- Bing takes steps to prevent the creation of suggestions that could undermine the right to private and family life and provides users with easy mechanisms to report problematic suggestions for removal. [See Best Practice PD1, PD8, PD9.](#)
- Microsoft, including Bing, also responds to periodic evaluations by the third-party independent organization [Ranking Digital Rights](#) before it publishes its annual ratings on Bing our practices, governance and leadership on the protection of freedom of expression and privacy.
- Bing is transparent about its privacy practices and content removal policies, as well as reports on content removed due to privacy concerns such as RTBF or NCII. [See Best Practice PT1, PT2, PT3, PT4, and PT5.](#)

**Residual risk score: Low.** Bing’s mitigations are designed to minimize the risk that materials negatively impacting private and family life will be disseminated via Bing while balancing other fundamental rights and interests such as free expression/access to information. Bing users cannot create, publish or share content on Bing which limits risk exposure related to user content or user interactions. Bing also takes steps to effectively protect the security of personal data provided to the service by users to avoid inadvertent use or disclosure. Bing’s policies and practices supporting these mitigations are sufficiently mature to address these risks.

**Additional mitigations needed:** Bing must continue to navigate the inherent nuance associated with providing broad access to information balanced with the need to protect users from harmful content. To do this, Bing must invest in proactive approaches to reducing the harms created by results that may negatively impact private and family life, and continue to invest in its existing defensive infrastructure. Bing should do this by taking a leadership role in the industry’s approach to integrating safety and risk reduction into the future of search, as well as expanding mechanisms for gathering and seeking out feedback from users, experts, civil society, industry partners, and researchers, including expanding Bing’s ability to share data with researchers. Bing should continue to enhance its ability to identify issues specific to local markets via partnerships with local experts in the markets where it operates. The deliverable for this exercise is to provide additional transparency reporting to the public,

---

further define internal accountabilities, and ensure highly detailed documentation of policies and procedures.

### C. To the protection of personal data

---

Summary of risks: Bing users could be harmed by inappropriate processing of personal data by the platform, such as by lack of transparency, lack of controls, lack of legal basis for processing, lack of access to remedies such as the Right to Be Forgotten, or by improper storage/inadvertent disclosure. Bing users or other data subjects may be harmed by private content appearing in search results that was published without consent, or autosuggestions disclosing personal details. Generative AI features in Bing could also be used to create materials that violate a data subject's privacy, or be used to find information about private individuals.

Inherent risk score: The right to protection of personal data is an individual issue of significant severity, and encountering materials that negatively impact the right to privacy in online search engines can lead to harm to individual users and to third party victims. However, per Bing's content removal reports, and in light of the size of the Bing index, far less than 1% of the index includes materials subject to removal under Bing's policies and practices regarding inappropriately published personal information (including under the Right to Be Forgotten). Bing has also not been the subject of a data breach regarding user data. The inherent risk related to data protection in Bing, absent mitigations, is **moderate**.

#### Summary of key risk mitigations:

- Bing has a robust privacy and security infrastructure that includes full-time professional compliance managers, requires pre-launch feature reviews and mitigations, adherence to standards for data handling and security by internal employees and vendors, training for all employees, and completion of Data Protection Impact Assessments to ensure appropriate risk mitigations and compliance with GDPR and other privacy laws. **See Best Practice PD2, PD3, PD5, PD6, and PE1.**
- Users of Bing are given controls over collection and use of personal data on the service, as well as controls that allow them to exercise their data subject rights to view, access, export and delete personal data held by Microsoft. Bing maintains user data in accordance with applicable Microsoft security standards. **See Best Practice PG3, and PD9.**
- Bing allows users and other stakeholders to easily report private content appearing in search results that violates local laws, including the Right to Be Forgotten, or Bing policies prohibiting the indexing of private content published without consent, such as NCII, sensitive personal data such as credit cards or other private information, or images of minors. Bing's content review teams are trained on legal obligations and Bing policies in order to efficiently take action on such requests and has access to local legal experts as needed. Bing engages in regular tests and audits of its system to ensure its ability to respond. **See Best Practice PE5.**

- 
- Bing’s terms of use governing activity on its generative AI features prohibit the use of the service to infringe on others’ privacy. Safety systems in generative AI features are designed to prevent privacy abuse, such as blurring faces in images used in prompts in visual search.  
[See Best Practice PE3.](#)
  - Bing takes steps to prevent the creation of, and reactively removes as necessary, autosuggestions that could undermine the right to private and family life. Bing provides users with easily accessible mechanisms for reporting problematic suggestions. [See Best Practice PD1, PD8, and PD9.](#)
  - Bing is transparent about its privacy practices and content removal policies, which are available in all member state languages, as well as reports on content removed due to privacy concerns such as RTBF or NCII. [See Best Practice PT1, PT2, PT3, PT4, and PT5.](#)

**Residual risk score:** **Low.** Bing’s mitigations are designed to minimize the risk that materials negatively impacting privacy are disseminated via Bing while balancing other fundamental rights and interests such as free expression/access to information. Bing also takes steps to effectively protect the security of personal data provided to the service by users to avoid inadvertent or harmful use or disclosure. Bing’s policies and practices supporting these mitigations are sufficiently mature to address these risks.

**Additional mitigations needed:** While Bing has robust privacy infrastructure to address privacy harms that may occur via content published in search results, as well as robust infrastructure to safely manage Bing user data, Bing should continue to invest in its defensive and privacy infrastructure so that it can continue to meet a high bar for privacy protection. Bing should continue to work with cross-Microsoft subject matter experts in privacy to stay abreast of new developments in privacy law and new technologies to better prevent privacy harms in search. In addition, Bing should further improve the protection of private data by publishing additional transparency documentation on how Bing delivers search results and any applicable personal data involved in the search engine operations as features evolve.

## **D. To freedom of expression and information, including the freedom and pluralism of the media**

---

**Summary of risks:** Bing’s ranking and relevance systems could be ineffective at returning useful search results, making information difficult to find, or ranking algorithms could be biased in ways that negatively affect media pluralism. Abuse by webmasters of Bing’s ranking algorithms could make it harder for users to find relevant content. Over-removal of content on Bing products and features, including overly limiting safety systems in suggestions or other recommendations, in ads, or in generative AI features, could limit free expression or enable large tech companies to control information available to the public. Bing could inadvertently apply its guidelines for inclusion in the News vertical that are biased against particular news outlets. Advertising policies could inadvertently restrict the ability of certain advertisers to express their views. Abuse of Bing’s reporting features could limit the free expression of third-party content owners.



---

**Inherent risk score:** The right to free expression and access to information online, and media pluralism, is an individual and societal issue of significant severity, an issue further exacerbated by the crucial role search engines play in enabling access to information online, coupled with the limited number of services offering global search indexes. The inherent risk related to the right to free expression and media pluralism via Bing, absent mitigations, is **high**.

**Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring that its ranking and relevance systems help users find the most relevant, highest authority content available in response to their search queries. Bing regularly tests the efficacy of its ranking and relevance systems using objective, repeatable metrics and employs hundreds of people dedicated to ranking and relevance improvements. Bing also employs a defensive search team designed to remedy algorithmic issues in high priority subject areas and deploys targeted algorithmic interventions as needed to ensure users are able to access the information they are seeking. Bing works to ensure its interventions are effective in all languages and regions in which Bing offers the service. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs, and through ingestion of user and stakeholder feedback. **See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Bing is continuing to refine its safety systems to avoid unnecessarily restricting access to conversation topics in generative AI features. Bing regularly reviews the efficacy of its interventions using established metrics. **See Best Practice PD1, PD5, PD7, and PI1.**
- Bing limits removal of content to narrow scenarios to avoid unduly impacting access to information. Bing provides training and oversight to its content review teams to ensure consistent application of policies, including reviews of decisions, and provides escalation paths to local legal experts as needed. For government demands, Bing employs additional safeguards to ensure any actions taken are narrow, specific, submitted in writing, and based on valid legal orders. Bing has automated safeguards in place to trigger additional reviews as needed. **See Best Practices PE1, PE2.**
- Users and webmasters who believe their content was affected in error or believe that Bing incorrectly actioned a complaint are provided mechanisms for redress; if decisions are reversed the site will be reindexed. **See Best Practice PE6.**
- Bing avoids proactive automated removal of content except in the narrow case of CSEAI and spam. **See Best Practice PE3.**
- Bing is transparent about its ranking parameters and its content moderation policies and provides in-product notices to inform users when content has been removed. Bing Webmasters have access to a dashboard that provides information about how their content has been indexed. Bing publishes information about content removals on a biannual basis. **See Best Practice PG1, PG3, PT1, PT2.**

- 
- Bing is biannually audited for its commitments to human rights, including free expression, as part of its membership in the Global Network Initiative. [See Best Practice PG5.](#)
  - Microsoft, including Bing, also responds to periodic evaluations by the third-party independent organization [Ranking Digital Rights](#) before it publishes its annual ratings on Bing our practices, governance and leadership on the protection of freedom of expression and privacy. [See Best Practice PG5.](#)
  - Bing and its partners across Microsoft meet regularly with external experts from civil society on issues of free expression to advance our policies and practices and better mitigate risks. In addition to cofounding the Global Network Initiative and upholding GNI principles on freedom of expression and privacy, Microsoft also participates in the [Advisory Network](#) (a multistakeholder advisory group) of the [Freedom Online Coalition](#), a coalition of 37 governments working to advance internet freedom. [See Best Practice PG5.](#)

**Residual risk score: Low.** Bing takes significant steps to ensure that the right to free expression is upheld in Bing, and that content removal is limited to narrow, specified scenarios. Bing’s policies and practices support free expression and Bing engages with civil society, including biannual audits with the Global Network Initiative, to ensure it upholds its commitments to free expression. Bing’s policies and practices supporting these mitigations are sufficiently mature to address these risks.

**Additional mitigations needed:** Search engines must navigate and balance the inherent nuance associated with providing broad access to information balanced with the need to protect users from harmful content. Bing should continue to enhance its transparency regarding ranking and moderation processes, including content removals, and continue to work with the Global Network Initiative and other expert partners to ensure its systems are appropriately balancing free expression and other interests, with a particular focus on markets and languages across the EU. Bing should continue to expand its network of research partnerships and enhance its ability to provide researchers with access to data to identify possible harms and solutions. Microsoft and Bing should continue to take a leadership position in working across industry to develop standards and best practices for generative AI features.

## **E. To non-discrimination**

---

**Summary of risks:** While users cannot post or share content through Bing, thus limiting user behavior as a vector of possible discrimination risk in Bing, web content shown in search results can include content that negatively affects the right of nondiscrimination for users or third parties. There is a risk that websites ranked by Bing contain harmful or illegal discriminatory materials, calls for violence against protected groups, hateful speech, harmful misinformation, or otherwise promote harmful attitudes or stereotypes. Bing’s ranking algorithms or moderation processes for Bing products and features could inadvertently bias search results towards certain viewpoints. Lack of available material on the web or prioritization of certain viewpoints could lead to a lack of diversity in search results in ways that perpetuate harmful stereotypes. Bing’s internal content moderation decisions and safety

---

systems, both in traditional search and in generative AI features, could inadvertently bias the system towards certain viewpoints, or overblocking could prevent display of materials important to allow individuals to engage in public life or exercise other important interests. Malicious actors may attempt to work around safety protections to promote or create harmful materials. Ads on Bing could be targeted in discriminatory ways that affect protected groups' ability to access crucial services like housing or employment.

**Inherent risk score:** The right to nondiscrimination is an individual and societal issue of significant severity, and encountering materials that negatively impact nondiscrimination in online search engines can lead to harm to individual users and potentially to third party victims. Bing's content removal reports, which include materials subject to removal for hate speech or other illegal discriminatory content, represent far less than 1% of the content in the index in light of the size of the Bing index. Bing also tracks the number of queries subjected to defensive interventions out of the total number of queries submitted to the service, which are an additional indicator of the frequency with which users engage with potentially problematic content across the systemic risk areas: these queries constitute less than 1% of the total number of queries in the search services. The inherent risk related to data protection in Bing, absent mitigations, is **moderate**.

#### **Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring its page crawlers and algorithms prioritize high quality content to avoid inadvertently returning discriminatory materials to users. Bing's ranking principles consider low quality content to include "pages that call for violence, name-calling, offensive statements, or use derogatory language to make a point are generally considered low quality" and such pages are, accordingly, ranked lower. Bing works to ensure its interventions are effective in all languages and regions in which Bing offers the service. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs, and through ingestion of user and stakeholder feedback. **See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Bing has a defensive search team that is dedicated to identifying and remedying, via targeted algorithmic interventions, high impact issues in search results, such as misinformation, hateful speech, and other problematic content that could negatively impact the right to freedom from discrimination. Bing regularly reviews the efficacy of its interventions through internal metrics as well as social listening channels, which are regularly reviewed with leadership, to ensure that Bing's interventions are performing as expected and not inadvertently introducing additional bias or other harm. **See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Where appropriate, Bing adds information to help ensure users are well informed and not misled or harmed by materials appearing in search results, such as answers pointing users to authoritative sources when standard search results are likely to lack high authority content, adds public service announcements or warnings, and prevents autosuggestions likely to lead to

---

harmful discriminatory materials. Bing works to ensure such materials are available across markets and languages in which Bing is offered. [See Best Practice PT5.](#)

- Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal discriminatory materials from governments, users, or other stakeholders, meeting required legal timeframes for content removal where applicable. Support teams are trained on applicable requirements and policies and provided escalation paths where needed for complex issues, including to local legal experts. The Bing team also relies on Microsoft's extensive team of subject matter experts and external experts to identify and quickly address emerging concerns. Bing engages in regular tests and audits of its system to ensure its ability to respond. [See Best Practice PD1, PD3, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, PI2, PI3, PI4, PI5, PT1, PT2, and PT3.](#)
- Bing and cross-Microsoft SMEs engage with external stakeholders in areas of key policy priorities, such as terrorist and violent extremist content, misinformation, CSEAI, and copyright infringement, to ensure that our internal policies, practices, and standards are addressing their key concerns. [See Best Practice PD4, PD5, PD6, PD7, PD8, PG1, PG5, PG6, PE1.1., PE1.2, PE6.1, PE6.3, PE7, PE8, PE9, PI2, and PI4.](#)
- Bing relies on Microsoft's extensive cross-company compliance infrastructure to proactively review new products and features to ensure they meet the bar for Microsoft's policy commitments in key areas such as privacy, accessibility, digital safety, and responsible AI. [See Best Practice PD2, PD3, PD4, PD5, and PG1.](#)
- Bing is transparent about its policies and actions with respect to user and webmaster content, and ensures this content is available in all relevant languages and markets. [See Best Practice PG1, PG3, PT1, PT2, PT3, PT4, and PT5.](#)
- These same mitigations apply to Bing's generative AI features, with additional enhanced safety features such as classifiers, filters, and a bespoke metaprompt that further limit the likelihood of harmful content appearing in Bing Chat/Image Creator features. Bing has engaged in extensive responsible AI reviews regarding generative AI features in order to ensure outputs are not biased or discriminatory. [See Best Practice PD1, PD2, PD4, PD5, PD6, PD7, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, PI2, PI3, PI4, PI5, PT1, PT2, PT3, PT4, and PT5.](#)
- Bing is continually working to ensure that its generative features do not over-block outputs so that users are able to access the information they seek. Bing works to be transparent about how its safety systems work for generative AI features, and ensures that users can always find third-party links to content responsive to their queries in standard search results. [See Best Practice PD1, PG1, PG3, PE3, and PE6.](#)
- Users of Bing's generative AI services are subject to a Code of Conduct that prohibits use of the service to create harmful or discriminatory content, and users who attempt to bypass Bing's safety systems to violate this restriction may be banned from the service. [See Best Practice PG1 and PG3.](#)
- The [Microsoft Advertising Policies](#) set out the requirements for ad content including criteria upon which ad content will be removed. Microsoft Advertising works to ensure that it enforces these

---

requirements in unbiased and non-discriminatory ways. Microsoft advertising is constantly improving its internal systems and process to ensure the advertising content served is legal, clear, truthful, and accurate. Microsoft Advertising's support teams work closely with its advertisers to review complaints around suspicious online activity, and they work across internal teams to verify data accuracy and integrity. [See Best Practice PD1, PD2, PD3, PD4, PD6, PD7, and PE4.](#)

**Residual risk score: Low.** Bing's policies and practices in ranking and moderation are designed to minimize harmful display of discriminatory content while avoiding undue impact on free expression. Bing's policies and practices supporting these mitigations are sufficiently mature to address these risks.

**Additional mitigations needed:** Bing should continue to invest in its ranking systems to prioritize high-quality material and continue to explore other mechanisms for reducing the likelihood that material in search results causes harm to nondiscrimination without unduly limiting access to information, such as through answers, indicators of content provenance or fact checks, and ensure that these interventions are available in all relevant markets and languages. Bing can also continue to improve its policies and procedures addressing its approach to reduction of algorithmic bias. These policies and procedures should include information such as the actions that development teams must take to reduce skewed outcomes, the procedures and assessments for ensuring training data is adequately representative, and the testing and approval chains required to be included in each feature. Metrics must be assigned to these efforts, and Bing should set goals and timeframes for further improvements. Bing should also continue to expand its connections with third party expert stakeholders, researchers, and civil society organizations to stay abreast of possible new threat vectors as well as potential new effective interventions. Bing should continue to identify additional investments that can specifically improve its systems across all EU countries and languages. Bing can continue to expand its partnerships with researchers, including enabling access to data, to help identify new solutions.

## F. To respect for the rights of the child

---

**Summary of risks:** Bing users cannot post or share content or interact with each other, which limits the risk that Bing users could harm children through online conduct, such as grooming or bullying. However, child users could be exposed to problematic content appearing in organic or paid search results, as well as interactions with Bing's generative AI features. Bing's safety systems could also negatively impact children if they unduly limit children's access to information or services; children have a right to engage in civic life and access information. Bing's data practices could harm children if they are not designed with the best interests of the child in mind. Bing users could also seek out child sexual exploitation and abuse imagery through the platform, which causes ongoing harm to child survivors of sexual abuse.

**Inherent risk score:** The rights of the child, and protection of children online, are an individual and societal issue of significant severity. However, the Bing service is not known to be used by a large

---

number of children (well less than 1% of EU Bing MAU are known to be users under 18), and Bing prevents known child users under the age of digital consent (including all users under 13) from accessing the full slate of Bing generative AI features. The inherent risk related to the rights of the child in Bing, absent mitigations, is **moderate**.

**Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring its crawlers and algorithms prioritize high quality content to avoid inadvertently returning harmful materials to users. Bing prevents autosuggestions likely to lead to harmful materials. Bing regularly tests the efficacy of its ranking and relevance systems using objective, repeatable metrics and employs hundreds of people dedicated to ranking and relevance improvements. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs. **See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Bing has a defensive search team that is dedicated to identifying and remedying, via targeted algorithmic interventions, high impact issues in search results, such as misinformation, hateful speech, and other problematic content that could negatively impact child users. Bing regularly reviews the efficacy of its interventions to ensure that they are performing as expected and not inadvertently introducing additional bias or other harms. **See Best Practice PD1, PD2, and PD5.**
- Bing also adds information to help ensure users are well informed and not misled or harmed by materials appearing in search results, such as answers pointing users to authoritative sources when search results lack high authority content, such as pointing users to suicide intervention resources when queries indicate a potential suicide intent. Bing also includes a number of digital literacy features to provide users with robust information regarding the content they are viewing, including content provenance indicators, fact checks, and answers. Bing’s generative AI Image Creator tools include a watermark on all images to indicate provenance as an AI-generated image. Bing’s SafeSearch tool by default blocks viewing of adult content for all search users. **See Best Practice PD1, PD2, PD3, PD4, PD5, PD6, PD7, PD8, PG1, PG3, PG5, PE 1.1, PE1.3, PE6.1, PE3, PT5, PE8, PE9, PI1, and P12.**
- Bing has features to provide families with greater control over their experience, including via Microsoft Family Safety tools that allow monitoring of search and to set certain filters for their children, including in SafeSearch. Child users whose accounts are subject to Family Safety monitoring are notified. Bing also enables family network administrators to set SafeSearch or chat restrictions for a home network or device as they deem appropriate. **See Best Practice PG2 and PD9.**
- Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal or policy-violating materials from governments, users, or other stakeholders, meeting required legal timeframes for content removal where applicable. Support teams are trained on applicable requirements and policies and provided escalation paths where needed for complex issues. The Bing team also relies on Microsoft’s extensive team of subject matter experts and external experts to identify and quickly address emerging concerns. Bing engages in regular

---

tests and audits of its system to ensure its ability to respond. [See Best Practice PD1, PD3, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, PI2, PI3, PI4, PI5, PT1, PT2, and PT3.](#)

- Bing and cross-Microsoft SMEs engage with external stakeholders on issues of child development and child protection, to ensure that our product designs, internal policies, practices, and standards are addressing their key concerns. [See Best Practice PD4, PD5, PD6, PD7, PD8, PG1, PG5, PG6, PE1.1, PE1.2, PE6.1, PE6.3, PE7, PE8, PE9, PI2, and PI4.](#)
- Bing relies on Microsoft’s extensive cross-company compliance infrastructure to proactively review new products and features to ensure they meet the bar for Microsoft’s policy commitments in key areas such as privacy, accessibility, digital safety, and responsible AI, which include specific standards on offering services to children. For example, privacy reviews ensure that child users are defaulted to the highest level of privacy protection and that disclosures and consent experiences are written in child-friendly language. [See Best Practice PD2, PD3, PD4, PD5, PD6.](#)
- Bing is transparent about its policies and actions with respect to user and webmaster content, and with its data practices. Microsoft offers a child-friendly version of its privacy statement to better inform younger users of data practices. [See Best Practice PG1, PG3, PT1, PT2, PT3, PT4, and PT5.](#)
- These same mitigations apply to Bing’s generative AI features, with additional enhanced safety features such as classifiers, filters, and a bespoke metaprompt that further limit the likelihood of harmful content appearing in Bing Chat/Image Creator features. Bing’s in-product disclosures and disclaimers around engaging with an AI system are written to be consumable by all ages. Bing has engaged in extensive responsible AI reviews regarding generative AI features in order to minimize the likelihood of harm. Bing prevents known users under the age of consent from accessing the full slate of conversational AI features. [See Best Practice PD1, PD2, PD4, PD5, PD6, PD7, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, PI2, PI3, PI4, PI5, PT1, PT2, PT3, PT4, and PT5.](#)
- Bing is continually working to ensure that its generative features appropriately balance safety considerations and the risk of over-blocking outputs so that users are able to access the information they seek. Where Bing’s safety filters prevent responses in conversational settings, Bing works to be transparent about how these safety systems operate. Relevant links to third-party content always remain available in standard search results. [See Best Practice PD1, PG1, PG3, PE3, and PE6.](#)
- Users of Bing’s generative AI services are held to a Code of Conduct that prohibits use of the service to create harmful or discriminatory content; users who violate this restriction may be banned from use of the service. [See Best Practice PG1 and PG3.](#)
- Bing proactively uses hash-matching technologies (including PhotoDNA and MD5) to detect matches to known CSEAI, to avoid it from appearing in the index using PhotoDNA, reactive reporting and removal, and via threat intelligence provided by third party expert partners. [See Best Practice PD1, PD2, PD5, PG5, PG6, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PI2, PI3, PI4, PI5, PT1, PT2, and PT3.](#)

- 
- Microsoft Advertising, which powers ads on Bing, has clear and regularly enforced content policies and practices that prevent known child users from being shown behaviorally targeted ads or ads for adult materials such as alcohol or gambling. [See Best Practice PG1.](#)

**Residual risk score: Low.** Bing’s algorithmic and moderation policies and practices are designed to ensure all users, including children, are not misled or harmed by content in search results. Bing’s data collection practices for known child users are designed with the best interests of the child in mind. Bing takes aggressive actions to eradicate CSEAI from appearing in the Bing index. Bing’s policies and practices supporting these mitigations are sufficiently mature to address risks to the rights of the child.

**Additional mitigations needed:** Bing must continue to navigate the inherent nuance associated with providing broad access to information balanced with the need to protect users from harmful content. For children, this means providing research capabilities and technologies that allow children to learn about the world, but ensuring they are not exposed to harmful content. To further mitigate this risk, Bing must invest in proactive approaches to understanding the possible harms to child users and the interventions most effective at balancing competing interests in this age group. Bing should do this by taking a leadership role in the industry’s approach to integrating safety and risk reduction into the future of search, as well as expanding mechanisms for gathering and seeking out feedback from users, families, expert stakeholders, industry partners, and researchers, including expanding Bing’s ability to share data with researchers. Bing should continue to identify additional investments that can specifically improve its systems across all EU countries and languages. The deliverable for this exercise is to provide additional transparency reporting to the public, identify additional interventions to help uphold the best interests of children, further define internal accountabilities, expand its relationships with third party stakeholders, and ensure highly detailed documentation of policies and procedures.

## **G. To a high level of consumer protection**

---

**Summary of risks:** Given that Bing works to index as much of the world wide web as possible, it has no control over the content appearing on third-party sites in the index, and that users may have valid reasons to seek out content that could be harmful or illegal in other contexts, there is a risk that content that could damage consumer protection will appear in the index, such as pages seeking to sell dangerous, counterfeit, or fraudulent goods or services, technology used to create scams or fraud (e.g., credit card pin readers), pages with fake reviews of products or services (or similar pages promoting mis or disinformation), other types of scams, or sites containing malware. Such content could also appear in ads on Bing or in generative AI features. Autosuggest features could enhance such harm by serving to lead users to harmful content. Sites in the index may contain pages that collect private information about users for use in ID theft or other fraud, such as credit card numbers or account information. On the other hand, overremoval of search results related to harmful products (in the interest of protecting users) could inadvertently limit users’ ability to access useful information about the safety of these products, how to report problems to authorities, or remedy harms. Bing could also fail to meet consumer protection expectations within its own service if it fails to adhere to



---

its own promises regarding ranking, moderation, or data handling. If Bing failed to adequately protect information held about users (e.g., search queries), a data breach could occur, making this content available to fraudsters. As Bing users cannot post or share content on the platform, nor sell goods or services directly, there is little risk stemming from user behaviors (e.g., fake accounts). Bing advertisements could be offered through fake accounts and/or contain harmful content.

**Inherent risk score:** The right to a high level of consumer protection is an individual and societal issue of significant severity, and encountering materials that negatively impact consumer protection in online search engines can lead to harm to individual users and potentially to third party victims. Bing's content removal reports, which include materials subject to removal for illegality such as consumer protection law violations and intellectual property violations, represent far less than 1% of the content in the Bing index (which contains hundreds of billions of pages). Bing also tracks the number of queries subjected to defensive interventions out of the total number of queries submitted to the service, which are an additional indicator of the frequency with which users engage with potentially problematic content across the systemic risk areas: these queries constitute less than 1% of the total number of queries in the search services. The inherent risk related to consumer protection in Bing, absent mitigations, is **moderate**.

#### **Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring its page crawlers and algorithms prioritize high quality content to avoid inadvertently returning websites likely to include content negatively impacting consumer protection, such as dangerous or harmful products or services, counterfeit merchandise, or pages containing private information that could be used for ID theft or fraud. Bing's Webmaster Guidelines detail the types of improper search engine optimization (SEO) tactics – often used to promote scams or fraud - that can lead to pages being ranked lower. Bing works to ensure its interventions are effective in all languages and regions in which Bing offers the service. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs, and through ingestion of user and stakeholder feedback. **See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Bing has a defensive search team that is dedicated to identifying and remedying, via targeted algorithmic interventions, areas where its ranking algorithms are not performing as expected. These interventions target high impact issues in search results where users have a heightened likelihood of being harmed, such as scams/fraud, misinformation, and other problematic content that could negatively impact consumer protection. Bing regularly reviews the efficacy of its interventions to ensure that they are performing as expected and not inadvertently introducing additional bias or other harm. Bing's defensive interventions are measured across markets and languages. **See Best Practice PD1, PD2, PD4, PG1, PG2, PGPG5, PE3, PE6.1, PI1, PI2., PE9, PT1, and PT4.**
- Bing also adds information to help ensure users are well informed and not misled or harmed by materials appearing in search results, such as answers pointing users to authoritative sources when search results lack high authority content, and prevents autosuggestions likely to lead to

---

harmful materials. In novel interfaces such as Bing chat, Bing works to remind users that they are engaging with an AI system and to remind users of the importance of verifying facts in the source material. These materials are designed to be available in all languages and regions Bing offers. [See Best Practice PD1, PD2, PD3, PD4, PD5, PD6, PD7, PD8, PG1, PG3, PG5, PE 1.1, PE1.3, PE6.1, PE3, PT5, PE8, PE9, P11, and P12.](#)

- Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal materials from governments, users, or other stakeholders, meeting required legal timeframes for content removal where applicable. Bing also has a policy to remove privacy-impacting materials that could be used for scams or fraud, such as pages containing account information or credit card numbers, offers users reporting mechanisms to inform Bing of such pages in the index. Support teams are trained on applicable requirements and policies and provided escalation paths where needed for complex issues, including to local legal and policy teams as needed. Bing engages in regular tests and audits of its system to ensure its ability to respond. [See Best Practice PD1, PD3, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, P11, P12, P13, P14, P15, PT1, PT2, and PT3.](#)
- Bing maintains user data in its systems in accordance with cross-Microsoft security and data handling standards to minimize the possibility of a data breach. [See Best Practice PD2, PD5, and PE1.](#)
- These same mitigations apply to Bing’s generative AI features, with additional enhanced safety features such as classifiers, filters, and a bespoke metaprompt that further limit the likelihood of harmful content appearing in Bing Chat/Image Creator features. Bing has engaged in extensive responsible AI reviews regarding generative AI features in order to ensure outputs are not biased or discriminatory. [See Best Practice PD1, PD2, PD4, PD5, PD6, PD7, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, P11, P12, P13, P14, P15, PT1, PT2, PT3, PT4, and PT5.](#)
- Users of Bing’s generative AI services are subject to a Code of Conduct that prohibits use of the service to create content enabling fraud, and users who attempt to bypass Bing’s safety systems in order to violate this restriction may be banned from the service. [See Best Practice PG3.](#)
- Microsoft Advertising, which powers ads on Bing, has clear and regularly enforced content policies and practices that prevent advertisements negatively impacting consumer protection. The [Microsoft Advertising Policies](#) set out the requirements for ad content including criteria upon which ad content will be removed. Microsoft requires our advertisers and partners to comply with our policies throughout their use of our services. Microsoft Advertising Policies prohibit advertising content that is misleading, deceptive, and fraudulent. Microsoft Advertising also has a set of Relevance and Quality Policies to manage the relevancy and quality of the advertisements that it serves through its advertising network. These policies deter advertisers from luring users onto sites using questionable or misleading tactics. [See Best Practice PG1.](#)
- Bing users may report advertising for takedown through the “feedback” form found on the Bing’s pages. Additionally, ads that may be promoting illegal content or that may appear to violate the Microsoft Advertising terms or policies may be submitted through one of these forms: Low

---

quality ad submission & escalation or Intellectual Property Concern Form. [See Best Practice PD8, PD9, PG1, PG2, PG3, PG4, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE5, PE6, PE6.1, PE6.2, PE6.3, PE7, PT1, PT2, and PT3.](#)

- Microsoft Advertising is constantly improving its internal systems and process to ensure the advertising content served is legal, clear, truthful, and accurate. Microsoft Advertising employs a robust filtration system to detect robotic traffic and other harmful cyber-attacks. This system uses various algorithmic systems to detect and neutralize illegal, invalid or malicious online traffic which may arise from or result in click fraud, phishing, malware, or account compromise. Microsoft Advertising has several teams of security engineers, support agents, and traffic quality professionals dedicated to continually developing and improving this traffic filtration and network monitoring system. Microsoft Advertising's support teams work closely with its advertisers to review complaints around suspicious online activity, and they work across internal teams to verify data accuracy and integrity. [See Best Practice PD1, PD2, PD3, PD4, PD6, PD7, and PE4.](#)

**Residual risk score:** **Low.** Bing's algorithmic and moderation policies and practices are designed to ensure its users are not misled or harmed by content in search results, including content that could negatively affect consumer protection, while avoiding undue impact on other important rights such as free expression/access to information. Bing protects user data with Microsoft-standard security practices to avoid data breaches. Bing's policies and practices supporting these mitigations are sufficiently mature to address risks to consumer protection.

**Additional mitigations needed:** Bing must continue to navigate the inherent nuance associated with providing broad access to information, including accessing information about potentially harmful content, balanced with the need to protect users from harmful content. To ensure a high level of consumer protection, Bing should continue to invest in high quality defensive systems, as well as continued support for public service announcements and warnings where appropriate, to ensure users are not harmed by content in search results. Bing can expand its partnerships with third party expert stakeholders to better understand emerging concerns in this area and identify new solutions, including via expansion of research partnerships (and researcher data access) and across additional geographies and languages. In addition, Bing should continue to publish up to date transparency documentation on how Bing delivers search results and how it protects personal data involved in the search engine operations as features evolve.

### **Foreseeable negative effects on democratic processes, civic discourse and electoral processes, and public security.**

---

**Summary of risks:** While Bing users cannot post or share information on the Bing platform, thus minimizing the risk that user content or interactions "go viral" or lead to large scale harm in civic discourse, democratic/electoral processes or public security, Bing does index third-party content on the broader internet, which may often include materials harmful to these important interests. If not appropriately mitigated, Bing users could be harmed by low quality or actively deceptive content appearing in search results, including misleading information about elections, electoral processes,

---

---

election outcomes and general public security. Ads appearing on Bing may also include such content, though its policies generally prohibit such content. Lack of authoritative information or data voids in less trafficked languages or geographies could lead to lower quality search results in certain markets. Bing's generative AI features could also be used to consume or create the types of content identified above, including synthetic media ("deepfakes") designed to mislead. Intentional actions by malicious actors could bypass Bing's safety systems and lead to increased visibility of this content. Bing's search suggestions could similarly serve to lead Bing users to this type of content. Bing's safety systems in search or chat could inadvertently introduce bias into search results if enforced inconsistently; in generative features Bing's safety systems could lead to overblocking that undermines users' ability to find related content. If not vetted appropriately, Bing's partnerships with internal and third-party experts on this topic could introduce bias into Bing systems.

**Inherent risk score:** Ensuring the integrity of democratic processes, civic discourse, and public security is a societal issue of significant severity. While users cannot communicate with each other in Bing and content cannot "go viral" through Bing, which minimizes the risk of uncivil discourse on the platform or misinformation shared by users, Bing does help users find relevant high quality, authoritative content on the web. Given that a search engine like Bing plays a key role in ensuring people have access to high quality answers to their research questions online, search engines play an important role in protecting democratic institutions. In measuring the frequency with which this issue occurs on the platform, Bing tracks the number of queries subjected to defensive interventions out of the total number of queries submitted to the service: these queries constitute less than 1% of the total number of queries in the search services. The inherent risk related to protection of democratic processes, civic discourse, and public security in Bing, absent mitigations, is **high**.

**Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring that its ranking and relevance systems help users find the most relevant, highest authority content available in response to their search queries. Bing regularly tests the efficacy of its ranking and relevance systems using objective, repeatable metrics and employs hundreds of people dedicated to ranking and relevance improvements. Bing also employs a defensive search team designed to remedy algorithmic issues in high priority subject areas and deploys targeted algorithmic interventions as needed to ensure users are able to access the information they are seeking. Bing works to ensure its interventions are effective in all languages and regions in which Bing offers the service. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs, and through ingestion of user and stakeholder feedback. [See Best Practice PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.](#)
- Bing works with internal teams including the Microsoft Threat Assessment Center (MTAC) and external experts in elections, public safety and misinformation to quickly identify and mitigate threats in these areas and prepare for key elections globally, including in the EU. Through Microsoft's Democracy Forward and MTAC teams, Microsoft offers mediums for election authorities, including in the EU and EU member states, to have lines of communication with

---

Microsoft to identify possible foreign information operations targeting elections. In addition, Microsoft works to identify and track nation-state information operations targeting democracies across the world and works with trusted third-party partners, including NewsGuard, Global Democracy Index (GDI), and EFE, to provide early indicators of narratives, hashtags, or information operations that can be leveraged to inform early detection and defensive search strategies. [See Best Practice PD1 and PI4.](#)

- Where search results may not return high authority content, such as in a data void, Bing works to ensure the digital literacy of its users through additional tools like answers, fact checks, and trustworthiness signals, such as NewsGuard ratings that allow users to better evaluate their sources of information. In novel interfaces such as Bing chat, Bing works to remind users that they are engaging with an AI system and to remind users of the importance of verifying facts in the source material. Microsoft is also developing tools and working with multistakeholder partnerships to combat disinformation that can pose risks to these interests, such as the Coalition for Content Provenance and Authenticity (C2PA), which is a content provenance standard that allows for watermarking of generated images. As another example, Bing also works with the Institute for Strategic Dialogue to surface “counter narratives” related to terrorist and violent extremist content to help deter users who may express an interest in recruitment. These interventions are available in multiple languages and markets. [See Best Practice PD1 and PE8.](#)
  - Bing works to prevent the display of search suggestions that can serve to lead users to harmful misleading content, and provides users with reporting mechanisms to quickly action problematic suggestions that do appear. [See Best Practice PD1, PD9, PE1, and PE2.](#)
  - Bing limits removal of content to narrow scenarios to avoid unduly impacting access to information. Bing provides training and oversight to its content review teams to ensure consistent application of policies, including reviews of decisions, and provides escalation paths for difficult questions, including to local legal experts as needed. For government demands, Bing employs additional safeguards to ensure any actions taken are narrow, specific, submitted in writing, and based on valid legal orders. [See Best Practice PD1, PD3, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, P11, P12, P13, P14, P15, PT1, PT2, and PT3.](#)
  - Bing is continuing to refine its safety systems to avoid unnecessarily restricting access to conversation topics in generative AI features, and to be transparent about how these safety systems operate. Where generative AI features do not provide a response, Bing users can still find access to relevant third-party links in standard search results. [See Best Practice PG3.](#)
  - Users and webmasters who believe their content was affected in error or believe that Bing incorrectly actioned a complaint are provided mechanisms for redress; if decisions are reversed the site will be reindexed. [See Best Practice PE6.](#)
  - Bing is transparent about its ranking parameters and its content moderation policies and provides in-product notices to inform users when content has been removed. These materials are available in all member state languages. Bing Webmasters have access to a dashboard that provides
-

---

information about how their content has been indexed. Bing publishes information about content removals on a biannual basis. [See Best Practice PT1, PT2, PT3, PT4, and PT5.](#)

- Bing engages with internal and external experts in combating misinformation in democratic processes and public safety, such as the Microsoft Democracy Forward team, Microsoft's Threat Analysis Center, NewsGuard, EFE, Reporters Without Borders and GDI. Bing is a signatory to the EU's Code of Practice on Disinformation and reports biannually on its ability to uphold its commitments under the Code. Bing also is committed to upholding Microsoft's information integrity principles and has worked to reflect these commitments in Bing's Trustworthy Search Principles. Microsoft and Bing also participate in crisis situation working groups, such as those developed in the EU for COVID-19 and the war in Ukraine, to share information across industry and ensure practices meet regulatory expectations. [See Best Practice PD4, PD5, PD6, PD7, PD8, PG1, PG5, PG6, PE1, PE6, PE7, PE8, PE9, PI2, and PI4.](#)
- Bing and its partners across Microsoft meet regularly with external experts from civil society in order to ensure Microsoft's and Bing's policies and practices meet expectations. Bing is biannually audited for its global commitments to human rights, including free expression, by the Global Network Initiative. [See Best Practice PD4, PD5, PD6, PD7, PD8, PG1, PG5, PG6, PE1, PE6, PE7, PE8, PE9, PI2, and PI4.](#)
- Microsoft Advertising, who powers ads on Bing, has clear and regularly enforced content policies and practices that prevent advertisements that could negatively impact democratic processes, including a prohibition on political advertising. [See Best Practice PG1.](#)
- Bing's Ad Repository in the [Microsoft Ad Library](#) also provides greater transparency to consumers about the ads they see on the Bing platform, including who's behind the ads and why they are shown. [See Best Practice PG1.](#)

**Residual risk score: Low.** Bing's algorithmic and moderation policies and practices are designed to ensure its users are not misled or harmed by content in search results, including content that could negatively affect civic discourse, public security, or electoral processes, while avoiding undue impact on other important rights such as free expression/access to information and nondiscrimination. Bing partners closely with Microsoft's related policy teams, such as Democracy Forward, Microsoft Threat Analysis Center, and Digital Safety, as well as external experts on misinformation and disinformation, elections, and public safety, in order to obtain up to date information about new vectors for risk and take action as needed. Bing's participation in the Code of Practice on Disinformation and related working groups also helps ensure that Bing's processes for countering these harms are robust. Bing's policies and practices supporting these mitigations are sufficiently mature to address these risks.

**Additional mitigations needed:** Search engines must navigate and balance the inherent nuance associated with providing broad access to information balanced with the need to protect users from harmful content, and to avoid introducing bias into Bing's search results via its ranking or moderation activities. To further enhance its ability to mitigate harms to democratic processes, civic discourse and electoral processes, and public security, Bing should continue to expand its ability to ensure that users have access to high authority content when searching for matters of significant public concern

---

---

through ranking improvements and enhanced search features, and continue to iterate on its generative AI safety systems to ensure users are able to access this information through chat features. Bing should also continue to grow its partnerships with internal and external experts on misinformation, elections, and public security, as well as its research partnerships (including expanding access to data) to further address these risks. Bing should continue to identify additional investments that can specifically improve its systems across all EU countries and languages.

### **Negative effects to public health, safety of minors, serious negative consequences to a person's physical or mental wellbeing, or gender-based violence**

---

**Summary of risks:** While Bing users cannot post or share information on the Bing platform, thus minimizing the risk that user content or interactions lead to harms to public health, safety of minors, physical or mental wellbeing, or gender-based violence, Bing does index third-party content from the broader internet, which may often include materials that could have negative effects on these important interests. If not appropriately mitigated, Bing users could be harmed by low quality or actively deceptive content appearing in search results, including for example by misinformation or disinformation about public health issues (including vaccines), sites purporting to sell illegal or gray market pharmaceuticals or other potentially dangerous products, hateful and harmful speech that could harm users' mental health, content promoting suicide or self-harm, content with particular harm to minors (e.g., adult content), or content such as CSEAI or NCII that leads to harms to child victims or based on gender. Ads appearing on Bing may also include such content, although it has policies in place to limit such ads. If not appropriately mitigated, Bing's generative AI features could also be used to consume or create the types of content identified above. Intentional actions by malicious actors could bypass Bing's safety systems or ranking algorithms and lead to increased visibility of this content. Bing's search suggestions could similarly serve to lead Bing users to this type of content. Bing's safety systems in search or chat could inadvertently introduce bias into search results if enforced inconsistently; in generative features Bing's safety systems could also lead to overblocking that undermines users' ability to find related content. Users may place an inordinate amount of trust in generative AI outputs that are incomplete or inaccurate. Users could violate Bing's terms of use and code of conduct to use generative AI tools to create convincing content that (if shared on other platforms) could lead to harm to public health, or mental or physical harm to individuals, including minors.

**Inherent risk score:** Ensuring a high degree of public health, the safety of minors, upholding a person's physical/mental wellbeing, and eliminating gender-based violence are societal issues of significant severity. While users cannot communicate with each other in Bing or cause content to "go viral" through Bing, minimizing the risk that individual users will use the platform to amplify harmful messages or misinformation, Bing does help users find relevant third-party content on the web. Given that a search engine like Bing plays a key role in ensuring people have access to high quality answers to their research questions online, search engines play an important role in upholding these rights. In measuring the frequency with which this issue occurs on the platform, Bing tracks the number of queries subjected to defensive interventions out of the total number of queries submitted to the

---

service: these queries constitute less than 1% of the total number of queries in the search services. The inherent risk related to protection of these important interests, absent mitigations, is **high**.

### **Summary of key risk mitigations:**

- Bing invests significant time and resources into ensuring its crawlers and algorithms prioritize high quality content to avoid inadvertently returning low quality harmful materials to users who have not expressed a clear interest in finding it. Bing works to ensure its interventions are effective in all languages and regions in which Bing offers the service. Bing regularly measures the efficacy of its ranking algorithms using metrics and makes changes as needed, including regularly reviewed and updated OKRs, and through ingestion of user and stakeholder feedback. **See Best Practices PD1, PD2, PD4, PD5, PD8, PD9, PG1, PG3, P4, PG5, PG6, PE 1.1, PE1.2, PE2, PE 3, PE5, PE7, PE8, PE9, PI1, PI2, PI3, PI4, and PT5.**
- Bing has a defensive search team that is dedicated to identifying and remedying, via targeted algorithmic interventions, high impact issues in search results, such as misinformation, public health concerns, hateful speech, self-harm materials, and other problematic content that could negatively impact these interests. Bing maintains a set of metrics to track, monitor and review the efficacy of its interventions on an ongoing basis to ensure that they are performing as expected and not inadvertently introducing additional bias or other harms. **See Best Practices PD1, PD5, PD7, PI1, and PI3.**
- Bing also adds information to help ensure users are well informed and not misled or harmed by materials appearing in search results, such as answers pointing users to authoritative sources when search results lack high authority content, which are available across the markets and languages where Bing offers services. Bing also works to prevent the display of suggestions likely to lead to harmful or offensive materials and gives users in-product mechanisms to report problematic suggestions. **See Best Practice PD1, PD2, PD3, PD4, PD5, PD6, PD7, PD8, PG1, PG3, PG5, PE 1.1, PE1.3, PE6.1, PE3, PT5, PE8, PE9, PI1, and P12.**
- Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal materials from governments, users, or other stakeholders, meeting required legal timeframes for content removal where applicable. Support teams are trained on applicable requirements and policies and provided escalation paths where needed for complex issues. Bing engages in regular tests and audits of its system to ensure its ability to respond. **See Best Practice PD1, PD3, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, P12, P13, P14, P15, PT1, PT2, and PT3.**
- Bing proactively uses hash-matching technologies (including PhotoDNA and MD5) to detect matches to known CSEAI, to avoid it from appearing in the index using PhotoDNA and via threat intelligence provided by third party expert partners. **See Best Practice PD1, PD2, PD5, PG5, PG6, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, P6.1, PI2, P13, P14, P15, PT1, PT2, and PT3.**
- Bing and cross-Microsoft subject matter experts (SMEs) engage with external stakeholders in areas of key policy priorities around the world, such as terrorist and violent extremist content, misinformation, responsible AI, CSEAI, and copyright infringement, to ensure that our internal policies, practices, and standards are addressing their key concerns. The Bing team also relies on



---

Microsoft's extensive team of subject matter experts and external experts to identify and quickly address emerging concerns in these areas, such as Digital Safety or Democracy Forward. [See Best Practice PD4, PD5, PD6, PD7, PD8, PG1, PG5, PG6, PE1.1., PE1.2, PE6.1, PE6.3, PE7, PE8, PE9, PI2, and PI4.](#)

- With respect to misinformation and disinformation risks, Bing supplements its internal threat identification and safety systems with further mitigations as signatory of the EU's Code of Practice on Disinformation, which requires biannual reporting on specific interventions (such as indicators of content provenance, elevation of high authority answers on problematic topics, and fact checks) to improve Bing's ability to fight misinformation across the EU, including in key areas such as public health/COVID 19. [See Best Practices PG5, PE9, PT1, and PT4.](#)
- Bing relies on Microsoft's extensive cross-company compliance infrastructure to proactively review new products and features to ensure they meet the bar for Microsoft's policy commitments in key areas such as privacy, accessibility, digital safety, and responsible AI. [See Best Practice PD2, PD3, PD4, PD5, and PD6.](#)
- Bing is transparent about its policies and actions with respect to user and webmaster content, and makes these disclosures available in all EU languages. [See Best Practice PG1, PG3, PT1, PT2, PT3, PT4, and PT5.](#)
- These same mitigations apply to Bing's generative AI features, with additional enhanced safety features such as classifiers, filters, and a bespoke metaprompt that further limit the likelihood of harmful content appearing in Bing Chat/Image Creator features. Bing has engaged in extensive responsible AI reviews regarding generative AI features in order to ensure outputs are not biased or discriminatory. Bing is continually working to ensure that its generative features do not over-block outputs so that users are able to access the information they seek. Bing has created in-product disclaimers, FAQs, and other explainers to prevent users from putting inordinate trust in responses in chat. Bing is also working to effectively label generated output so that it cannot be used to misinform on other platforms, such as including watermarks on generated images and through participation cross-industry efforts to improve identification techniques like the Coalition for Content Provenance and Authenticity. Bing's work as a signatory on the COP on Disinformation also includes addressing misinformation risks in generative AI features. [See Best Practice PD1, PD2, PD4, PD5, PD6, PD7, PD8, PD9, PG1, PG2, PG3, PG4, PG5, PG6, PG7, PE1.1, PE1.2, PE1.3, PE2, PE3, PE4, PE5, PE6.1, PE6.2, PE6.3, PE7, PE8, PE9, PI1, PI2, PI3, PI4, PI5, PT1, PT2, PT3, PT4, and PT5.](#)
- Microsoft Advertising, which powers advertisements that appear on Bing, has clear and regularly enforced content policies that prevent advertising of harmful materials. [See Best Practice PG1.](#)

**Residual risk score: Low.** Bing's algorithmic and moderation policies and practices are designed to ensure its users are not misled or harmed by content in search results, including content that could negatively public health or individual wellbeing, while avoiding undue impact on other important rights such as free expression/access to information. Bing's policies and practices supporting these mitigations are sufficiently mature to address these risks.

---

**Additional mitigations needed:** Bing leverages industry best practices to ensure prioritization of high-quality material and reduction of the prevalence of dangerous or misleading information that could lead to these harms. Bing should continue to improve these systems and continue to partner with third party experts and other stakeholders to quickly identify new areas of concern and implement new, effective mitigations. Bing can continue to expand its work with external researchers, including expanding researcher access to data. Bing should continue to work to identify local resources that can help support Bing’s in-product interventions (especially for topics impacting user health and wellbeing) across all EU markets. Metrics must be assigned to these efforts, and Bing should set goals and timeframes for further improvements.

## Key Takeaways

### Key strengths

Bing has established system-wide processes that help mitigate the areas of highest risk across the platform:

- Bing has designed, and continually updates, its ranking algorithms to ensure that its search results consistently provide high quality, authoritative information to users while not unduly limiting access to information, given Bing's important role in upholding free expression and access to information. Bing has hundreds of employees dedicated to improving its core search algorithms and upholding Bing's trustworthy search principles. Bing supplements these processes with a specific "defensive search" team designed to identify and remedy algorithmic failures in high profile areas, such as those falling under each of the identified systemic risks, using targeted algorithmic improvements. These processes are designed to be effective across all markets and languages where Bing is offered. Bing regularly reviews the efficacy of its ranking features using established metrics and OKRs, and also reviews and acts on user feedback as necessary.
- Where users are seeking low authority content that could be misleading or harmful, and Bing determines that such interventions are likely to be helpful (and not exacerbate the problem), Bing works to include additional contextual information to search results to ensure users are not harmed by content in search results. These interventions may include answers, public service announcements, site-level warnings, indicators of content provenance such as NewsGuard ratings, fact checks, counternarratives for users seeking terrorist and violent extremist content to dissuade recruitment, and other high-quality information to supplement what is returned in the main search results. Bing works to ensure this information is available across the markets and languages where its services are offered.
- To uphold free expression and access to information, Bing works to avoid removing content from search results except in limited, narrow scenarios where legal demands or other important interests warrant removal. Bing has long established processes to ingest reports of concern from users and other stakeholders, including governments. Complaint review is conducted by a global team of language and policy enforcement experts that adhere to strict limits with respect to the time in which the review must be completed. For government demands, Bing has designed its systems to avoid enabling censorship and works with civil society, including biannual audits by the Global Network Initiative, to ensure compliance with Microsoft's human rights principles. Bing's processes also enable removal of other content in line with local laws (such as defamation, copyright, or privacy) and Bing policies, such as those requiring removal of privacy-impacting materials (including materials used for doxing, identity theft, or NCII) and is aggressive in working to prevent the indexing of CSEAI.
- Bing provides transparency about its algorithms, ranking principles, content policies through publication in How Bing Delivers Search Results and the Bing Webmaster Guidelines, and cross-company documentation, such as the Microsoft Privacy Statement and the Microsoft Services Agreement. These materials are made available in all relevant EU member state languages.

- Bing works with cross-Microsoft teams to ensure compliance with internal policies and standards in a variety of high priority policy areas, such as privacy, digital safety, responsible AI, information integrity, diversity and inclusion, and accessibility. New product features in Bing are reviewed for compliance with each of these requirements before features are launched. Bing works with these cross-company teams of subject matter experts in order to identify new areas of concern for mitigation and to understand emerging legal requirements in the markets where it operates.
- Bing works to be data-driven in its processes and has implemented systematic mechanisms for verifying the efficacy of its systems and interventions, including biannually updated OKRs, and through regular review of metrics designed to identify areas for continued improvement, including with key executives in regularly recurring meetings.
- Data provided by users in Bing is handled in accordance with Microsoft privacy and security standards to ensure compliance with GDPR and other laws, and to avoid data breach.
- Bing and its internal Microsoft partners work with external researchers and subject matter experts (including data sharing relationships) to supplement its internal knowledge of key subject areas, to quickly identify new threats, and to obtain feedback on the efficacy of its processes.
- Bing's new generative AI features build on the processes and safety systems outlined above and were developed in regular consultation with Microsoft's Responsible AI team, which supplemented these processes with additional protections relevant to net-new potential harms raised by AI features. Bing has worked to provide extensive transparency on its development process and published regular updates about its learnings in blog posts and through [The New Bing: Our Approach to Responsible AI](#), public FAQs, and in-product disclosures. Bing is continually updating these features and disclosures in response to user and regulatory feedback.
- Since Bing generally does not allow users the ability to post content or communicate directly with other users, this limits Bing's risk exposure as to many other systemic risk areas common to other online platforms, such as content virality, impersonation, or user rights regarding content removals and appeals. The risk of such harms is therefore low, and as such Bing has (appropriately) not dedicated significant resources to terms of use enforcement or internal complaint resolution processes. Bing has robust processes in place that will ensure that any new features that may implicate user content will be proactively reviewed and additional risk mitigations implemented prior to launch.

## Looking forward

Bing consistently looks for opportunities to be more proactive in its approach to understanding, analyzing, and mitigating systemic risks. By finding ways to even more deeply integrate trust and safety into every facet of the product development cycle and invest in innovative approaches, Bing seeks to continually improve its risk posture. Key focus areas over the next two years include:

- Continue to navigate the inherent nuance associated with providing broad access to information balanced with the need to protect users from harmful content;
- Continue to expand transparency of Bing's principles, processes, and content moderation activities;
- Expand mechanisms for gathering and seeking out feedback from users, civil society, regulators, industry partners, and other stakeholders, and create more formal processes for reflecting this feedback, as appropriate, in Bing's product improvement processes;
- Expand Bing's relationships with third party experts to quickly identify new areas of concern and identify and implement more effective interventions for issues of significant concern, including coverage across EU markets and languages;
- Expand Bing's network of relationships with internal and external researchers, including creating more formal processes by which Bing can share data with researchers;
- Further document policies, processes, and responsibilities for Bing's compliance and safety operations to ensure internal accountabilities and adherence to procedures, including improving processes for handoffs between internal teams, and further advance Microsoft's and Bing's compliance functions; and
- Take a leadership role in the industry's approach to integrating safety and risk reduction into generative artificial intelligence and emerging technologies.

## Conclusion

After evaluating the purpose and design of the Bing search services, the key systemic risks associated with the functionality and features available on Bing, and the safety systems and other risk mitigation processes implemented by Bing, the Risk Assessment team has determined that while there is room for Bing's safety systems to continue to mature to address additional risks more effectively, Bing's existing measures are sufficient to mitigate key systemic risks on the platform. Going forward, Bing should build on its strong foundation to further improve its ability to prevent harm to its users, such as though creating more formal processes for identifying, documenting and mitigating risks in new products and features, iterating on its transparency documentation, expanding its relationships with third party expert partners to expand its ability to quickly identify and mitigate new areas of harm, expanding its programs to enable researcher access to data, continuing to be an active participant in cross-industry efforts, and to continue to be a leader in the industry in integrating safety and risk reduction into new technologies such as artificial intelligence.

