Code of Practice on Disinformation – Report of Microsoft for the period 1 July – 31 December 2023

## Table of Contents

Executive Summary	
Subscription per Service overview	
II. Scrutiny of Ad Placements	
Commitment 1	
Commitment 2	
Commitment 3	
III. Political Advertising	
Commitment 4	
Commitment 5	
Commitment 6	
Commitment 7	
Commitment 8	
Commitment 9	
Commitment 10	
Commitment 11	
Commitment 12	
Commitment 13	
IV. Integrity of Services	
Commitment 14	
Commitment 15	

Commitment 16	
V. Empowering Users	93
Commitment 17	94
Commitment 18	
Commitment 19	
Commitment 20	
Commitment 21	131
Commitment 22	
Commitment 23	
Commitment 24	
Commitment 25	157
VI. Empowering the research community	158
Commitment 26	159
Commitment 27	
Commitment 28	
Commitment 29	
VII. Empowering the fact-checking community	174
Commitment 30	175
Commitment 31	
Commitment 32	
Commitment 33	
VIII. Transparency Centre	
Commitment 34	
Commitment 35	

Commitment 36	
IX. Permanent Task-Force	193
Commitment 37	
X. Monitoring of Code	196
Commitment 38	197
Commitment 39	198
Commitment 40	199
Commitment 41	
Commitment 42	201
Commitment 43	
Commitment 44	
Reporting on the services' response during a period of crisis	204
War of aggression by Russia on Ukraine	
Israel-Hamas Conflict	214
Reporting on the services' response during an election	222
European Elections 2024	223

#### **Executive Summary**

Microsoft is pleased to file this report on our compliance with the commitments of the strengthened 2022 EU Code of Practice on Disinformation. At Microsoft, we are committed to instilling trust and security across our products and services, and across the broader web. We recognise that fighting disinformation is a key element to creating a trustworthy and safe online environment and continue to increase our efforts to counter these threats. We also recognize that there is not a one-size-fits-all approach to this work, and instead there needs to be a whole of society strategy that recognizes that not all services or platforms are the same and that different measures may be more effective than others in improving the information environment for our users.

Microsoft appreciates that since the last report was filed, the focus on artificial intelligence (AI) and interest in understanding how AI could affect the spread of disinformation have continued to grow. While AI certainly poses challenges in the information integrity space, we also see many opportunities for AI to assist and streamline defenders' work in detecting and assessing influence operations. To be clear, challenges include improved efficiency of malicious content creation and dissemination, as threat actors will be able to create high-quality inauthentic content quickly and cost-efficiently, which requires Microsoft to evolve and respond to changing tactics. However, Microsoft is fully committed to utilizing best in class tools and technology to help mitigate the risks of its services being used to further disinformation.

Serving as a leader in AI research, we are committed to proactively publicize our threat detection efforts for the benefit of the AI community, regulators, and broader society. As such, we have adopted six focus areas to combat the harmful use of deceptive AI:

- 1. A strong safety architecture
- 2. Durable media provenance and watermarking
- 3. Safeguarding our services from abusive content and conduct
- 4. Robust collaboration across industry and with governments and civil society
- 5. Modernized legislation to protect people from the abuse of technology
- 6. Public awareness and education

While this report focuses on the second half of calendar year 2023, ahead of the 2024 European elections, it is pertinent to highlight in this report that Microsoft came together with the tech sector at the Munich Security Conference in February 2024 to take an important step forward against AI deepfakes during pivotal times, specifically the elections that will take place around the globe this year. Microsoft and LinkedIn are two of 20 companies that announced a new <u>Tech Accord to Combat Deceptive Use of AI in 2024</u> <u>Elections</u>. The goal is straightforward but critical – to combat video, audio, and images that fake or alter the appearance, voice, or actions of political candidates, election officials, and other key stakeholders or that provide false information to voters about when, where, and how they can lawfully vote.

The Tech Accord's commitments will make it more difficult for bad actors to use legitimate tools to create deepfakes, while simultaneously simplifying the process for users to identify authentic content. This focuses on the work of companies that generate AI content as well as those that distribute it, and calls on them to strengthen the safety architecture in AI services by assessing risks and strengthening controls to help prevent abuse. For its part, Microsoft has already taken steps to meet the commitments in the Tech Accord with our <u>new election protection commitments</u>, by further implementing content provenance, establishment of reporting channels and improved detection capability. For example:

- Microsoft is harnessing the data science and technical capabilities of our AI for Good Lab and Microsoft Threat Analysis Center teams to better detect deepfakes on the internet. We will call on the expertise of our Digital Crimes Unit to invest in new threat intelligence work to pursue the early detection of AI-powered criminal activity.
- We have also launched a new web page <u>Microsoft-2024 Elections</u> where a political candidate can report to us a concern about a deepfake of themselves. In essence, this empowers political candidates around the world to aid with the global detection of deepfakes.
- In addition, Microsoft will launch <u>Content Credentials as a Service</u> to enable political candidates around the world to digitally sign and authenticate media using the Coalition for Content Provenance and Authenticity's (C2PA) digital watermarking credentials.
- Microsoft will deliver a series of trainings on election security and AI dedicated to political groups in the European Parliament.

Beyond our commitment to combat deceptive use of AI in this year's elections, we are in the process of implementing a set of additional actions safeguarding candidates, election campaigns, election authorities, and voters:

- Microsoft's newly formed "Campaign Success Team" will advise and support political campaigns as they navigate the world of AI, combat the spread of cyber influence campaigns, and protect the authenticity of their own content and images.
- Microsoft will provide access to a new "Election Communications Hub" to support democratic governments around the world as they build secure and resilient election processes.
- Ahead of the European elections, Microsoft will establish a Situation Room, bringing together resources across the company to monitor, support, and protect the European Parliament elections.
- Bing Search is undertaking a multifaceted approach to election integrity and will also integrate specialized answers and information panels for the European elections, with a link to official voting information.

Microsoft is taking a cross-product, whole-of-company approach to ensure the responsible implementation of AI. This starts with our <u>Responsible AI Principles</u>. Building on those principles in June of 2022 Microsoft released our <u>Responsible AI Standard v.2</u> and <u>Information Integrity Principles</u> to help set baseline standards and guidance across product teams. Recognizing that there is an important role for government, academia and civil society to play in the responsible deployment of AI, we also created a <u>roadmap for the governance of AI</u> across the world as well as creating a vision for the responsible advancement of AI, both inside Microsoft and throughout the world, <u>including specifically in Europe</u>. For more information on Microsoft's commitment to Responsible AI and ongoing internal and external efforts, we encourage you to review our <u>Responsible AI hub</u>, which offers a range of information, tools, and resources related to the ethical and responsible use of AI technologies. It includes detailed information about Microsoft's internal Responsible AI processes and tools which can be used to responsibly develop and deploy AI products.

Additionally, we will continue to build upon these approaches to Responsible AI. For example, recognizing both the enormous potential for generative and other forms of AI to transform the world of work in positive ways and the potential risks AI presents in that context, LinkedIn recently published its framework of <u>Responsible AI Principles</u>, which is inspired by and aligned with Microsoft's Responsible AI Principles. LinkedIn provides more details on these principles in our response to Commitment 15, including how we've put these principles into practice with respect to three recently launched Generative AI products.

Microsoft has developed tools, including Project Providence with Truepic, and multistakeholder partnerships, such as the Coalition for Content Provenance and Authenticity (C2PA), to combat the rise of manipulated or AI created media. <u>Project Providence</u> is a proof-of-concept mobile app that collects images and video with authentication information embedded. This could assist any organization or person to document events, places, or objects anywhere in the world and provides proof that images were unaltered and authentic. Microsoft is a founding member of C2PA, alongside Adobe, Intel, Truepic, Twitter, the BBC, and other tech and media companies.

Additionally, Microsoft has partnerships to strengthen the company's capacity and understanding of global threats to disinformation: first, with EFE Verifica (<u>EFE</u>), to track false narratives spreading in Spain, Latin America, and Spanish speaking populations; and second, with Reporters Sans Frontières (RSF), to use their <u>Journalism Trust Initiative (JTI)</u> dataset in Microsoft products. These partnerships are part of a larger effort to empower Microsoft users to better understand the information they consume across our platforms and products.

Microsoft is going beyond efforts to simply combat disinformation by helping users become more educated on information sources and potential disinformation through media literacy and digital education. Media literacy campaigns are not designed to tell individuals what to believe or how to think; rather, they are about equipping people to think critically and make informed decisions about what information they consume. Microsoft continues to invest in our Information Literacy Program, <u>partnering</u> with leading news and media literacy nonprofits, including the News Literacy Project, a collaboration led by The Trust Project on the Trust Indicators and Verified, to develop campaigns built on industry research and best practices. Microsoft provides funding for the research and development of the campaigns as well as threat intelligence insights, technical expertise, and in-kind ad space on Microsoft platforms to promote the programs. Microsoft supported the work of Minecraft game developers in their efforts to launch a game, exploring core media literacy concepts within the Minecraft Education world and the work of the Microsoft Education team to launch a free app in Microsoft Teams that helps educators and students to form effective queries and identify reliable resources. Finally, Microsoft has been in ongoing conversations with several fact-checking organizations to expand our fact-checking capability to further meet our commitments under the Code of Practice and explore innovative ways to integrate trustworthiness signals and fact-checks into our evolving technology platforms, beyond those that we already have.

Microsoft has subscribed to the Code of Practice with the following services:

- <u>Microsoft Advertising</u> is our proprietary advertising platform, which serves the vast majority of ads displayed on Bing Search and provides advertising to most other Microsoft services that display ads, as well as many third-party services. Microsoft Advertising works both with advertisers, who provide it with advertising content, and publishers, such as Bing Search, who display these advertisements on their services. Microsoft Advertising employs a distinct set of policies and enforcement measures with respect to each of these two categories of business partners to prevent the spread of disinformation through advertising.
- <u>Bing Search</u> is an online search engine with the primary objective of connecting users to the most relevant search results from the web. Users come to Bing with a specific research topic in mind and expect Bing to provide links to the most relevant and authoritative third-party websites on the Internet that are responsive to their search terms. Therefore, addressing misinformation or disinformation in organic search results often requires a different approach than may be appropriate for other types of online services, as over-moderation of content in search could have a significant negative impact on the right to access information, freedom of expression, and media plurality. Therefore, 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 remediation efforts may on occasion involve removal of content from search results (where legal or policy considerations warrant removal), in many cases, Bing has found that actions 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 measures 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 prevent users from being unexpectedly exposed to harmful content in search results that they did not expressly seek to find. Copilot in Bing provides 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. It is built on longstanding safety systems in search, supplemented by additional protections for new risks related to AI like conversational drift, hallucinations, and jailbreaking. Microsoft has partnered closely with Microsoft's Responsible AI team to proactively address these harms and has been transparent about its approach in <u>Copilot in Bing: Our approach to Responsible AI</u>. Bing continues to evolve these features based on user and external stakeholder feedback.

• <u>LinkedIn</u> is a real identity online social networking service for professionals to connect and interact with other professionals, grow their professional network and brand, and seek career development opportunities. LinkedIn is part of its members' professional identity and has a specific purpose. Activity on the platform and content members share can be seen by current and future employers, colleagues, potential business partners and recruitment firms, among others. Given this audience, members by and large tend to limit their activity to professional areas of interest and expect the content they see to be professional in nature. LinkedIn is committed to keeping its platform safe, trusted, and professional and respects the laws that apply to its services. On joining LinkedIn, members agree to abide by LinkedIn's <u>User Agreement</u> and its <u>Professional Community Policies</u>, which expressly forbid members from posting information that is false or misleading.

As a company, we have made significant progress in the past six months to empower users to better understand the information they consume across our platforms and products. Over the course of the next six months, we'll continue to make this information transparent and public. Specifically, we will focus on the following areas:

- Further de-funding the mechanisms nation-states are using to push their narratives and propaganda and regularly evaluating and improving user and advertiser policies as needed.
- Meet the <u>voluntary commitments</u> we have made for the responsible advancement of AI technology. These include deploying new state-of-the-art provenance tools to help the public identify AI-generated audio-visual content and understand its provenance, as well as enhanced detection and response capability.
- Ensuring Microsoft and LinkedIn products are developed consistent with Microsoft's Responsible AI Standards and LinkedIn's Responsible AI Principles, as relevant, and that risks associated with AI systems are mitigated to provide safe, trustworthy, and ethical experiences for users and, further, ensuring that our information integrity principles are integrated into AI systems included in Microsoft products.
- Continuing to proactively expose and share information about cross-platform foreign influence operations and actioning such intelligence appropriately through defensive search and other techniques. This includes expansion of the trusted third parties Microsoft uses to inform its work detecting and disrupting these influence operations. This also includes adding trusted third parties in additional languages, ensuring global coverage for our information integrity work. We will be particularly focused on cross platform information sharing related to elections.
- Strengthening our efforts and expanding our funding in the areas of media literacy and critical thinking, aiming to include vulnerable groups and having greater language access. As part of our focus areas and commitments under the Tech Accord we will increase our partnerships to increase AI literacy efforts and build greater understanding of provenance and other trustworthiness indicators.
- Providing users with additional tools and functionality to help understand the trustworthiness of the sites or domains they are visiting and empowering them to make informed decisions about those sources.

- Supporting good faith research into disinformation and broader disinformation trends and tactics.
- Continue to share learnings pertaining to generative AI as products and services evolve and new threats or TTPs emerge. In addition, Microsoft will continue to regularly evaluate, implement, and share best practices for addressing disinformation trends and TTPs as we navigate the technological changes posed by the malicious use of AI.
- Develop new partnerships for expanded fact-checking coverage in the EU and continue to explore further ways to help users evaluate content on our services.
- Enhance existing research tooling to provide enhanced data reporting and continue to deliver relevant data and research to support research into the spread of disinformation.
- Educating users on generative AI features, including their risks and limitations, and providing the broader public and research community with information on our approach to Responsible AI

Unless stated otherwise, data provided under this report covers a reporting period of 1 July 2023 to 31 December 2023 ("Reporting Period").

### Subscription per Service overview

Commitments	Measures	LinkedIn	Microsoft Advertising	Bing Search	Microsoft Corporation				
II. Scrutiny of Ad Placements									
	Measure 1.1	$\boxtimes$	$\boxtimes$						
	Measure 1.2	$\boxtimes$	$\boxtimes$						
1	Measure 1.3	$\boxtimes$	$\boxtimes$						
I	Measure 1.4	$\boxtimes$							
	Measure 1.5	$\boxtimes$	$\boxtimes$						
	Measure 1.6	$\boxtimes$	$\boxtimes$						
	Measure 2.1	$\boxtimes$	$\boxtimes$						
2	Measure 2.2	$\boxtimes$	$\boxtimes$						
2	Measure 2.3	$\boxtimes$	$\boxtimes$						
	Measure 2.4	$\boxtimes$	$\boxtimes$						
	Measure 3.1	$\boxtimes$	$\boxtimes$						
3	Measure 3.2	$\boxtimes$	$\boxtimes$						
	Measure 3.3	$\boxtimes$	$\boxtimes$						
		III. Political	Advertising						
	Measure 4.1	$\boxtimes$	$\boxtimes$						
4	Measure 4.2	$\boxtimes$	$\boxtimes$						
5	Measure 5.1	$\boxtimes$	$\boxtimes$						
	Measure 6.1								
	Measure 6.2								
6	Measure 6.3								
	Measure 6.4								
	Measure 6.5								
7	Measure 7.1								
1	Measure 7.2								

	Measure 7.3	$\boxtimes$	$\boxtimes$		
	Measure 7.4				
8	Measure 8.1				
0	Measure 8.2				
9	Measure 9.1				
9	Measure 9.2				
10	Measure 10.1				
10	Measure 10.2				
	Measure 11.1				
11	Measure 11.2				
	Measure 11.3				
	Measure 11.4				
	Measure 12.1				
12	Measure 12.2				
	Measure 12.3				
	Measure 13.1				
13	Measure 13.2				
	Measure 13.3				
		IV. Integrity	of Services		
	Measure 14.1	$\boxtimes$		$\boxtimes$	
14	Measure 14.2	$\boxtimes$		$\boxtimes$	
	Measure 14.3	$\boxtimes$		$\boxtimes$	
15	Measure 15.1	$\boxtimes$		$\boxtimes$	
	Measure 15.2	$\boxtimes$		$\boxtimes$	
16	Measure 16.1	$\boxtimes$		$\boxtimes$	
10	Measure 16.2	$\boxtimes$			
V. Empowering Users					
	Measure 17.1	$\boxtimes$		$\boxtimes$	
17	Measure 17.2	$\boxtimes$		$\boxtimes$	
	Measure 17.3	$\boxtimes$		$\boxtimes$	

	Measure 18.1	$\boxtimes$			
18	Measure 18.2	$\boxtimes$		$\boxtimes$	
	Measure 18.3	$\boxtimes$		$\boxtimes$	
19	Measure 19.1	$\boxtimes$		$\boxtimes$	
19	Measure 19.2	$\boxtimes$		$\boxtimes$	
20	Measure 20.1	$\boxtimes$		$\boxtimes$	
20	Measure 20.2	$\boxtimes$		$\boxtimes$	
	Measure 21.1	$\boxtimes$		$\boxtimes$	
21	Measure 21.2	$\boxtimes$			
	Measure 21.3	$\boxtimes$		$\boxtimes$	
	Measure 22.1	$\boxtimes$			
	Measure 22.2	$\boxtimes$		$\boxtimes$	
	Measure 22.3	$\boxtimes$		$\boxtimes$	
22	Measure 22.4				
	Measure 22.5				
	Measure 22.6				
	Measure 22.7	$\boxtimes$		$\boxtimes$	
23	Measure 23.1	$\boxtimes$		$\boxtimes$	
25	Measure 23.2	$\boxtimes$		$\boxtimes$	
24	Measure 24.1	$\boxtimes$			
25	Measure 25.1				
25	Measure 25.2				
	VI. I	Empowering the	Research Commu	nity	
	Measure 26.1	$\boxtimes$		$\boxtimes$	
26	Measure 26.2	$\boxtimes$		$\boxtimes$	
	Measure 26.3	$\boxtimes$		$\boxtimes$	
	Measure 27.1	$\boxtimes$		$\boxtimes$	
27	Measure 27.2	$\boxtimes$		$\boxtimes$	
21	Measure 27.3	$\boxtimes$		$\boxtimes$	
	Measure 27.4	$\boxtimes$		$\boxtimes$	

	Measure 28.1	$\boxtimes$		$\boxtimes$	
28	Measure 28.2	$\boxtimes$		$\boxtimes$	
20	Measure 28.3	$\boxtimes$		$\boxtimes$	
	Measure 28.4	$\boxtimes$		$\boxtimes$	
	Measure 29.1				
29	Measure 29.2				
	Measure 29.3				
	VII. Em	powering the Fa	ct-checking Com	nunity	
	Measure 30.1	$\boxtimes$		$\boxtimes$	
30	Measure 30.2	$\boxtimes$		$\boxtimes$	
30	Measure 30.3	$\boxtimes$		$\boxtimes$	
	Measure 30.4	$\boxtimes$		$\boxtimes$	
	Measure 31.1	$\boxtimes$			
31	Measure 31.2	$\boxtimes$		$\boxtimes$	
51	Measure 31.3	$\boxtimes$		$\boxtimes$	
	Measure 31.4	$\boxtimes$		$\boxtimes$	
	Measure 32.1	$\boxtimes$		$\boxtimes$	
32	Measure 32.2	$\boxtimes$			
	Measure 32.3	$\boxtimes$		$\boxtimes$	
33	Measure 33.1				
		VIII. Transpa	rency Centre		
	Measure 34.1				$\boxtimes$
	Measure 34.2				$\boxtimes$
34	Measure 34.3				$\boxtimes$
	Measure 34.4				$\boxtimes$
	Measure 34.5				$\boxtimes$
	Measure 35.1				$\boxtimes$
35	Measure 35.2				$\boxtimes$
55	Measure 35.3				$\boxtimes$
	Measure 35.4				$\boxtimes$

	Measure 35.5			$\boxtimes$
	Measure 35.6			$\boxtimes$
	Measure 36.1			$\boxtimes$
36	Measure 36.2			$\boxtimes$
	Measure 36.3			$\boxtimes$
		IX. Permaner	nt Task-Force	
	Measure 37.1			$\boxtimes$
	Measure 37.2			$\boxtimes$
37	Measure 37.3			$\boxtimes$
57	Measure 37.4			$\boxtimes$
	Measure 37.5			$\boxtimes$
	Measure 37.6			$\boxtimes$
		X. Monitoring	g of the Code	
38	-			$\boxtimes$
39	-			$\boxtimes$
	Measure 40.1			$\boxtimes$
	Measure 40.2			$\boxtimes$
40	Measure 40.3			$\boxtimes$
40	Measure 40.4			$\boxtimes$
	Measure 40.5			$\boxtimes$
	Measure 40.6			$\boxtimes$
	Measure 41.1		$\boxtimes$	$\boxtimes$
41	Measure 41.2			$\boxtimes$
	Measure 41.3			$\boxtimes$
42	-			$\boxtimes$
43	-			$\boxtimes$
44	-			$\boxtimes$

# II. Scrutiny of Ad Placements Commitments 1 - 3

#### **II. Scrutiny of Ad Placements**

Commitment 1

Relevant signatories participating in ad placements commit to defund the dissemination of disinformation, and improve the policies and systems which determine the eligibility of content to be monetised, the controls for monetisation and ad placement, and the data to report on the accuracy and effectiveness of controls and services around ad placements

	C.1	M 1.1	M 1.2	M 1.3	M 1.4	M 1.5	M 1.6
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Microsoft	Microsoft	Microsoft	Microsoft		Microsoft	Microsoft
	Advertising	Advertising	Advertising	Advertising		Advertising	Advertising

	LinkedIn	Microsoft Advertising
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	<ul> <li>As outlined below, LinkedIn: <ul> <li>has implemented and enforces policies concerning misinformation and disinformation.</li> <li>provides information and tools to give advertisers transparency and control regarding the placement of their advertising.</li> <li>has integrated brand safety tools and services.</li> </ul> </li> </ul>	<ul> <li>Microsoft Advertising implemented the following measures: <ul> <li>Iterated on its automated detection mechanisms to enforce its policies on Information Integrity. Since the last reporting period, Microsoft Advertising improved its controls to further prevent advertising from serving on web domains that violates our policies.</li> <li>Further developed automated data reporting functionalities</li> </ul> </li> </ul>

		to meet the Code's reporting compliance efforts.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	LinkedIn plans to continue to assess	Microsoft Advertising plans to
	its policies and services and to	continue to assess its policies,
	update them as warranted.	services, reporting, and
		accreditations to update them as
		warranted.

Measure 1.1	LinkedIn	Microsoft Advertising
QRE 1.1.1	LinkedIn prohibits misinformation and disinformation on its platform,	Microsoft Advertising prohibits misinformation and disinformation on its
<b>_</b>	whether in the form of organic content or in the form of advertising	network.
	content. LinkedIn's Professional Community Policies, which apply to all	
	content on LinkedIn's platform, expressly prohibit false and misleading	As reported in QRE 1.2.1, Microsoft Advertising works with selected, trustworthy
	content, including misinformation and disinformation:	publishing partners and requires these partners to abide by strict brand safety-
		oriented policies to avoid providing revenue streams to websites engaging in
	• Do not share false or misleading content. Do not share	misleading, deceptive, harmful, or insensitive behaviours.
	content that is false, misleading, or intended to deceive. Do	
	not share content to interfere with or improperly influence an	Additionally, in December 2022, Microsoft Advertising rolled out revised network-
	election or civic process. Do not share content that directly	wide policies to avoid the publishing and carriage of harmful Disinformation and
	contradicts guidance from leading global health organisations	the placement of advertising next to Disinformation content. Such policies
	and public health authorities; including false information	prohibit ads or sites that contain or lead to Disinformation. To enforce this policy,
	about the safety or efficacy of vaccines or medical treatments.	"We may use a combination of internal signals and trusted third-party data or
	Do not share content or endorse someone or something in	information sources to reject, block, or take down ads or sites that contain
	exchange for personal benefit (including personal or family	disinformation or send traffic to pages containing disinformation. We may block
	relationships, monetary payment, free products or services, or	

other value), unless you have included a clear and conspicuous	at the domain level landing pages or sites that violate this policy." Please see <u>here</u>
notice of the personal benefit you receive and have otherwise	for our main policy page.
complied with our Advertising Policies.	
LinkedIn provides specific examples of false and misleading content	
that violates its policy via a Help Center article on False or Misleading	
Content.	
<u>content.</u>	
Linkedle's Advertising Deligion in some sets the shore provision and	
LinkedIn's <u>Advertising Policies</u> incorporate the above provision, and	
similarly prohibit misinformation and disinformation. In addition,	
LinkedIn's Advertising Policies also prohibit fraudulent and deceptive	
ads, and require that claims in an ad have factual support:	
o Fraud and Deception: Ads must not be fraudulent or	
deceptive. Your product or service must accurately match the	
content of your ad. Any claims in your ad must have factual	
support. Do not make deceptive or inaccurate claims about	
competitive products or services. Do not imply you or your	
product are affiliated with or endorsed by others without their	
permission. Additionally, make sure to disclose any pertinent	
partnerships when sharing advertising content on LinkedIn.	
Do not advertise prices or offers that are inaccurate - any	
advertised discount, offer or price must be easily discoverable	
· · ·	
from the link in your ad.	
Of note, unlike some other platforms, LinkedIn does not allow	
members to monetise or run ads against their content, nor does it offer	
an ad revenue share program. Thus, members publishing	
disinformation on LinkedIn are not able to monetise that	
disinformation or collect advertising revenue via LinkedIn. LinkedIn has	
instead reported the number of ads it restricted on its platform during	
the period.	

SLI 1.1.1 –	Methodology of data measurement:		Methodology of data measureme	ent:
SLI 1.1.1 – Numbers by actions enforcing policies above (specify if at page and/or domain level)	The table below reports metrics concerning ads LinkedIn restricted under the misinformation policies in QRE 1.1.1.		Microsoft Advertising assesses a individual ads that we preve participating in the Microsoft Ad the Microsoft Advertising service number web domains that we blo the last reporting period, we hav prevent ad calls on web domains these domains, thus enforcing of improvements, however, Microsof "Impression" field as describe measurement options for the upp - "Domain" represents the num ad impression in a Memb domains across all regions (r individual country level be	the impact of its actions by reporting on the ented from monetizing on web properties vertising network (i.e., "publisher sites" that use es to display ads on their properties), and the ocked from participating in our ad network. Since we made additional system upgrades to further that we blocked and eliminated impressions on ur policies more effectively. As a result of these off Advertising is no longer able to report on the ed below. We are assessing further data
	The number of ads LinkedIn The number of impressions the restricted under its misinformation ads received before they were policies in QRE 2.1.1 between 1 July restricted. – 31 December 2023		Impression	Domain
Level	Ads	Impressions	Impression	Domain
Member States		· · ·	· · · ·	
Austria	0	0		
Belgium	0	0		

Bulgaria	0	0		
Croatia	0	0		
Cyprus	0	0		
Czech Republic	0	0		
Denmark	0	0		
Estonia	0	0		
Finland	0	0		
France	1	0		
Germany	3	3,111		
Greece	0	0		
Hungary	0	0		
Ireland	0	0		
Italy	2	0		
Latvia	0	0		
Lithuania	0	0		
Luxembourg	0	0		
Malta	0	0		
Netherlands	1	0		
Poland	0	0		
Portugal	0	0		
Romania	0	0		
Slovakia	0	0		
Slovenia	0	0		
Spain	0	0		
Sweden	0	0		
Iceland	0	0		
Liechtenstein	0	0		
Norway	0	0		
Total EU	7	3,111	N/A	570
Total EEA	7	3,111	N/A	570

	Methodology of data measurement:	Methodology of data measurement:
	Following the methodology developed by the Task-force	Following the methodology developed by the Task-force Subgroup on Ac
	Subgroup on Ad Scrutiny, this SLI considers the impressions	Scrutiny, this SLI considers the impressions of ads or sources that were blocked
	to ads or sources that were blocked and applies an agreed- upon conversion factor to those impressions.	and applies an agreed-upon conversion factor to those impressions.
		It should be noted, however, that Microsoft Advertising is a "pay per click" (PPC
	As reported above, LinkedIn restricted 7 ads between 1 July	ad network. This means that advertisers are charged following a click on the ad
	- 31 December 2023 under its misinformation policies in	and not based on the number of impressions served. We therefore calculated the
	QRE 1.1.1.	approximate financial value in the table by using the "blended CPM" value
SLI 1.1.2 - Preventing the		provided by Ebiquity plc. and the following equation:
flow of legitimate	We calculated the approximate financial value in the table by	
advertising investment to	using a "blended CPM" value and the following equation:	(Impressions/1000) x BlendedCPM; where CPM means "Cost Per Mille."
sites or content that are		
designated as	(Impressions/1000) x BlendedCPM; where CPM means "Cost	As described above, Microsoft Advertising prevents serving any ads on well
disinformation	Per Mille."	domains that spread disinformation. Such proactive measures block all ad traffic
		upfront (i.e., we act at the earliest point of identification, which could mean prio
	In most cases, the ads did not receive any impressions	to an impression being served), which brings the monetization on those web
	before they were restricted, resulting in a low estimated	properties down to zero. The figures shown in the table below are the estimated
	Euro value under the methodology.	impact based on the numbers we reported in our previous report. We expect a consistent impact across reporting periods.
		As described above, Microsoft Advertising continues to iterate on its proactive mechanisms to prevent potentially infringing content from serving to consumers
		which will be reflected in future reports.

<sup>&</sup>lt;sup>1</sup> Request placed between a seller and buyer of advertising that can detail amongst other things website, specific content, targeting data inclusive of audience or content. <sup>2</sup> Comprehensive calculation of the number of people who have been reached by a piece of media content by passive exposure (viewing a piece of content) or active engagement (visiting a destination).

	Euro value of ads demonetised	Euro value of ads demonetised
Member States		
Austria	0	€ 159.9605
Belgium	0	€ 166.0040
Bulgaria	0	0
Croatia	0	0
Cyprus	0	€ 2.8625
Czech Republic	0	€ 9.7775
Denmark	0	€ 187.1820
Estonia	0	€ 0.7800
Finland	0	€ 26.9685
France	0	€ 133.7750
Germany	€12.44	€ 2,599.5360
Greece	0	0
Hungary	0	0
Ireland	0	€ 133.8925
Italy	0	€ 732.5895
Latvia	0	0
Lithuania	0	€ 0.0390
Luxembourg	0	€ 3.6750
Malta	0	0
Netherlands	0	€ 366.4960
Poland	0	€ 1.7595
Portugal	0	€ 21.0375
Romania	0	€ 3.5120
Slovakia	0	0
Slovenia	0	0
Spain	0	€ 75.3600
Sweden	0	0
Iceland	0	0
Liechtenstein	0	€ 0.3390
Norway	0	€ 138.2710

Total EU	€12.44	€ 4,625.2070
Total EEA	€12.44	€ 4,763.8170

Measure 1.2	LinkedIn	Microsoft Advertising
	LinkedIn does not offer an ad revenue share program and does not allow	Microsoft Advertising works with selected, trustworthy publishing partners
	members to monetise content they post to LinkedIn by running ads	and requires these partners to abide by strict brand safety-oriented policies
	against it.	to avoid providing revenue streams to websites engaging in misleading,
		deceptive, harmful, or insensitive behaviors. These publishers also benefit
	LinkedIn displays ads in two environments: (1) on the LinkedIn platform,	from the set of measures identified above that Microsoft Advertising takes
	which accounts for the vast majority of ads; and (2) on the LinkedIn	with regard to advertisers, which ensures that these partners receive high-
	Audience Network, which allows LinkedIn advertisers to extend their	integrity, non-deceptive ads from the Microsoft Advertising platform.
	reach to professionals on a curated network of thirty-seven-hundred	
	third-party publishers selected by LinkedIn (for example, Nasdaq.com,	Microsoft Advertising's policies with respect to these publishers include a
	CNN.com, Vogue.com, Realtor.com).	comprehensive list of prohibited content that ads cannot serve against.
		Prohibited content includes, but is not limited to, Disinformation, sensitive
	With respect to the first category – ads displayed on the LinkedIn	political content (e.g., extreme, aggressive, or misleading interpretations of
	platform – as noted in response to QRE 1.1.1, unlike other platforms,	news, events, or individuals), unmoderated user-generated content, and
QRE 1.2.1	LinkedIn does not offer a content monetisation or an ad revenue share	unsavory content (such as content disparaging individuals or organisations).
-	program. Thus, no member content is monetised or demonetised, and	Publishers are required to maintain a list of prohibited terms and provide us
	there is no ability for a member publishing disinformation to collect any	with information on their content management practices where applicable.
	advertising revenue share from LinkedIn.	In addition to content requirements, publishers are required to abide by
		restrictions against engaging in business practices that are harmful to users
	With respect to the second category – ads displayed on the LinkedIn	(e.g., distributing malware).
	Audience Network – LinkedIn takes a number of steps to help ensure	
	LinkedIn advertisers' ads appear in a trusted environment and that	Microsoft Advertising reviews publisher properties and domains for policy
	publishers that systematically provide harmful disinformation are not	compliance, including compliance with restrictions on prohibited content.
	included in the LinkedIn Audience Network.	In this review, Microsoft Advertising also considers feedback from its
		advertisers to help ensure a safe environment for the delivery of their
	• First, the LinkedIn Audience Network is a curated network of	advertisements, and maintains a review process to investigate related
	third-party sites and apps selected by LinkedIn. LinkedIn does	advertiser complaints. Publishers are promptly notified of properties or domains that violate Microsoft Advertising's policies; such properties and
	not allow any blog, application, or website to join the LinkedIn	domains that violate microsoft Advertising's policies; such properties and domains are not approved by Microsoft for live ad traffic. If a property or
		uomains are not approved by microsoft for live ad traffic. If a property or

	<ul> <li>Audience Network and display ads; rather, LinkedIn selects the publishers that are included in the network.</li> <li>Second, LinkedIn has integrated with partners, such as Integral Ad Science and DoubleVerify, to help monitor the quality and brand safety of the publishers in the LinkedIn Audience Network and filter out publisher inventory that falls short of standards, such as GARM's brand safety floor.</li> <li>Third, LinkedIn regularly reviews the publishers included in the LinkedIn Audience Network to ensure they meet LinkedIn standards and are serving LinkedIn advertisers.</li> </ul> To date, LinkedIn has periodically removed publishers from the LinkedIn Audience Network, but has not had to remove any publisher as a result of publishing disinformation.	domain is already live, and later found in violation of Microsoft Advertising's policies, it is removed from the network until the publisher remedies the issue. As stated in our policies, "We may use a combination of internal signals and trusted third-party data or information sources to reject, block, or take down ads or sites that contain disinformation or send traffic to pages containing disinformation." Specific to Disinformation, Microsoft Advertising partner with Information Integrity experts, such as NewsGuard and GDI, as source and references of Disinformation domains. Microsoft Advertising is actively blocking domains that these sources deem as Disinformation.
SLI 1.2.1	Methodology of data measurement:         As stated in response to QRE 1.2.1, LinkedIn does not allow third parties to monetise content they post to LinkedIn by running ads against it and has not had to remove any publisher from the LinkedIn Audience Network for publishing disinformation.         Accordingly, the metrics for this SLI for the period 1 July – 31 December 2023 are zero.	Methodology of data measurement As reported in SLI 1.1.1, since Microsoft Advertising blocks domains globally, not at the Member State level, we are providing the global aggregate number of 570 barred during the current reporting period. For clarity, this is an incremental figure which applies in addition to all previously blocked domains. Microsoft Advertising did not bar any accounts during the relevant reporting period. This is because we take actions by blocking web domains across the entire network, irrespective of which ad account may be promoting them. Microsoft Advertising did not roll out any new policies in the reporting period. Therefore, the relevant "nr. of policy reviews" is zero.

	Nr of policy reviews	Nr of update to policies	Nr of accounts barred	Nr of	Nr of policy	Nr of	Nr of	Nr of
				domains	reviews	update	accounts	domains
				barred		to	barred	barred
						policies		
Member States								
Global					Zero (0)	Zero (0)	Zero (0)	570
Total EU	-	-	-	-				
Total EEA	-	-	-	-				

Measure 1.3	LinkedIn	Microsoft Advertising
QRE 1.3.1	LinkedIn provides a range of information and tools to give advertisers transparency and control regarding the placement of their advertising. For example, for ads on the LinkedIn platform, LinkedIn publishes a Feed Brand Safety score for advertisers and the public. The Feed Brand Safety score measures the number of ad impressions on the LinkedIn platform that appeared adjacent to – that is, immediately above or below within the LinkedIn feed – content removed for violating LinkedIn's Professional Community Policies, including disinformation. From July through December 2023, the Feed Brand Safety score was 99%+ safe. More information about LinkedIn's Feed Brand Safety Score is available <u>here</u> . In addition, LinkedIn publishes for advertisers and the public a semiannual transparency report, which discloses the amount of violating member content, including misinformation, that LinkedIn removed from the platform during the period. For the period from 1 January to 30 June 2023, for example, LinkedIn removed 85,216 pieces of misinformation from the platform. LinkedIn's most recent transparency report is available <u>here</u> . For ads on the LinkedIn Audience Network, as discussed in QRE 1.2.1, LinkedIn provides tools to assist advertisers in controlling where their	<ul> <li>Microsoft Advertising provides its customers with campaign reporting and functionalities to monitor and control ad placement across the Microsoft Advertising network. Such transparency controls are generally available via the campaign User Interface (UI) and through customer support. Transparency controls include: <ul> <li>Ad delivery reports at the domain level: data reports show the website/domain where the ads are served.</li> <li>Site exclusions: ability to exclude certain websites/domains from the ad campaigns to prevent ads from serving on such websites.</li> <li>Negative keywords exclusions: ability to exclude certain keywords from the ad campaigns to prevent ads from serving against search queries containing such keywords.</li> <li>Syndication Publisher Network Opt-Out: ability to prevent any ad delivery on the extended publisher network. In which case, all ads will serve on owned and operated properties</li> </ul> </li> </ul>

	ads appear within the network. For example, advertisers can set up category-level blocking based on the Interactive Advertising Bureau's (IAB) publisher category taxonomy to prevent their ads from running on certain types of publishers within the network. Similarly, advertisers can review the list of publishers within the network and create custom allow lists and block lists to ensure their ads are placed on apps and sites that meet an advertiser's specific standards.	
Measure 1.4	LinkedIn	Microsoft Advertising
QRE 1.4.1	This Measure is not relevant or pertinent for LinkedIn as it does not buy advertising on behalf of others, inclusive of advertisers, and agencies.	This measure is not relevant or pertinent to Microsoft Advertising as it does not buy advertising, inclusive of advertisers, and agencies.
Measure 1.5	LinkedIn	Microsoft Advertising
QRE 1.5.1	As indicated in response to QRE 1.2.1, LinkedIn does not offer a content monetisation or an ad revenue share program. Thus, no member content is monetised or demonetised, and there is no ability for a member publishing disinformation on LinkedIn to collect advertising revenue share. As a result, LinkedIn has not undertaken independent third-party audits relative to monetisation and disinformation.	This measure is not relevant or pertinent to Microsoft Advertising, as it does not produce first party reporting. Please see QRE 1.5.2 for the relevant actions.
QRE 1.5.2	Not applicable.	Microsoft Advertising undergoes yearly Media Rating Council (MRC) accreditations via third-party audit. The MRC accreditation certifies Microsoft Advertising's click measurement systems adheres to the industry standards for counting ad clicks and the processes supporting this technology are accurate. Here is the Microsoft Advertising's <u>MRC accreditation letter</u> . This <u>article</u> provides a summary of the click measurement processes and methods Microsoft Advertising employs to measure and count clicks. For additional information, please visit the <u>IAB/MRC click measurement guidelines</u> . Microsoft Adverting undergoes yearly audit by the Network Advertising Initiative (NAI) as part of the annual member's compliance review process.

Measure 1.6	LinkedIn	Microsoft Advertising
	LinkedIn has integrated a number of brand safety tools and services	As described in QRE 1.2.1, Microsoft Advertising partner with Information Integrity
	to help advertisers understand and control the placement of their	experts, such as NewsGuard and GDI, as source and references of Disinformation
	advertising, and help avoid the placement of advertising next to	domains. Microsoft Advertising is actively blocking domains that these sources deem
	disinformation content and/or in places or sources that repeatedly publish disinformation.	as Disinformation.
		Please see QRE 1.3.1 for transparency and control functionalities.
	First, it's worth noting that LinkedIn endeavors to limit the	
	disinformation that may appear on its platform in the first place. As	
	set out in response to QREs 17.1.1 / 18.1.3 / 18.2.1 / 23.2.1, LinkedIn	
	has implemented automated and manual systems and processes to	
	detect and remove content that violates our policies, including disinformation, and to take action on violative content when it's	
	reported to us. Further, LinkedIn limits and controls the publishers that are included in the LinkedIn Audience Network, discussed in	
QRE 1.6.1	response to QRE 1.2.1.	
QKE 1.0.1	Second, LinkedIn has partnered with third parties, such as Integral Ad	
	Science and DoubleVerify, to evaluate and filter advertising inventory	
	on LinkedIn Audience Network publisher sites that falls short of	
	standards, such as GARM's brand safety floor. These partners help	
	evaluate and filter third-party publisher advertising inventory before	
	a bid is placed, and decrease instances when an ad may run on an	
	unsafe or low-quality page.	
	In addition, LinkedIn has implemented a Brand Safety Hub within	
	LinkedIn Campaign Manager. As part of the hub, advertisers can	
	control what publisher apps and sites their ads appear on within the	
	LinkedIn Audience Network. For example, advertisers can create	
	custom block lists and allow lists of publisher sites within the LinkedIn	
	Audience Network that meet an advertiser's specific standards.	
	Similarly, advertisers can apply third-party brand safety tools to their	

	campaigns, including IAB publisher categories and DoubleVerify brand suitability profiles.	
	This QRE is not relevant or pertinent for LinkedIn as it does not buy	This QRE is not relevant or pertinent as Microsoft Advertising does not buy
QRE 1.6.2	advertising on behalf of others, inclusive of advertisers, and agencies.	advertising.
QRE 1.6.3	This QRE is not relevant or pertinent as LinkedIn is not a brand safety tool provider.	This QRE is not relevant or pertinent as Microsoft Advertising is not a brand safety tool provider.
QRE 1.6.4	This QRE is not relevant or pertinent as LinkedIn is not a ratings service.	This QRE is not relevant or pertinent as Microsoft Advertising is not a ratings service.
SLI 1.6.1	Methodology of data measurement: Not applicable.	Methodology of data measurement: N/A

II. Scrutiny of Ad Placements							
	Commitment 2						
Relevant Signatories participating in advertising commit to prevent the misuse of advertising systems to disseminate Disinformation in the form of advertising messages.							
	C.2 M 2.1 M 2.2 M 2.3 M 2.4						
We signed up to the following LinkedIn LinkedIn LinkedIn LinkedIn LinkedIn LinkedIn							
measures of this commitment:	Microsoft Advertising	Microsoft Advertising	Microsoft Advertising	Microsoft Advertising	Microsoft Advertising		

	LinkedIn	Microsoft Advertising
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	As outlined below, LinkedIn:	Microsoft Advertising implemented
		the following measures:

	<ul> <li>has implemented and enforces policies concerning misinformation and disinformation.</li> <li>partners with others to facilitate the flow of information to tackle purveyors of disinformation.</li> <li>has implemented systems and procedures to help ensure ads on LinkedIn comply with its Advertising Policies.</li> <li>provides information to advertisers when ads are restricted about the policies at issue.</li> </ul>	<ul> <li>Iterated on its automated detection mechanisms, including new automated classifiers to detect misleading claims relating to false information and consumer scams, such as financial scams, unsupported pricing claims and sensationalized ads, and misleading celebrity endorsements.</li> <li>Further developed automated data reporting functionalities to meet the Code's reporting compliance efforts.</li> </ul>
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures LinkedIn plans to continue to assess its policies and services and to update them as warranted.	Planned Implementation Measures Microsoft Advertising plans to continue to assess its policies and services and to update them as

Measure 2.1	LinkedIn	Microsoft Advertising	
QRE 2.1.1	As noted in response to QRE 1.1.1, LinkedIn prohibits misinformation and	d As described in QRE 2.2.1, Microsoft Advertising employs a set of policies that	
	disinformation on its platform, whether in the form of organic content or	prohibit advertising content that is misleading, deceptive, fraudulent, or	
	in the form of advertising content. LinkedIn's Professional Community	harmful, including misinformation and disinformation.	

<u>Policies</u>, which apply to all content on the platform, expressly prohibit false and misleading content, including <u>misinformation and</u> <u>disinformation</u>:

**Do not share false or misleading content.** Do not share content that is false, misleading, or intended to deceive. Do not share content to interfere with or improperly influence an election or other civic process. Do not share content that directly contradicts guidance from leading global health organisations and public health authorities; including false information about the safety or efficacy of vaccines or medical treatments. Do not share content or endorse someone or something in exchange for personal benefit (including personal or family relationships, monetary payment, free products or services, or other value), unless you have included a clear and conspicuous notice of the personal benefit you receive and have otherwise complied with our <u>Advertising Policies</u>.

LinkedIn provides specific examples of false and misleading content that violates its policy via a Help Center article on <u>False or Misleading</u> <u>Content.</u>

LinkedIn's <u>Advertising Policies</u> incorporate the Professional Community Policy provision, and similarly prohibit misinformation and disinformation. In addition, LinkedIn's Advertising Policies separately prohibit fraudulent and deceptive ads, and require that claims in an ad have factual support:

 Fraud and Deception: Ads must not be fraudulent or deceptive. Your product or service must accurately match the content of your ad. Any claims in your ad must have factual support. Do not make deceptive or inaccurate claims about competitive products or services. Do not imply you or your product are affiliated with or endorsed by others without their permission. Additionally, make sure to disclose any pertinent partnerships when sharing advertising

In December 2022, Microsoft Advertising rolled out revised network-wide policies to avoid the publishing and carriage of harmful Disinformation and the placement of advertising next to Disinformation content. Such policies prohibit ads or sites that contain or lead to Disinformation. To enforce this policy, "We may use a combination of internal signals and trusted third-party data or information sources to reject, block, or take down ads or sites that contain disinformation or send traffic to pages containing disinformation. We may block at the domain level landing pages or sites that violate this policy." Please see <u>here</u> for our main policy page.

	content on LinkedIn. Do not advertise prices or offers that are inaccurate – any advertised discount, offer or price must be easily discoverable from the link in your ad.	
	Methodology of data measurement:	Methodology of data measurement:
SLI 2.1.1 – Numbers by actions enforcing policies above	<ul> <li>The table below reports the number of ads LinkedIn restricted under the misinformation policies in QRE 2.1.1 above between 1 July – 31 December 2023, broken out by EEA Member State.</li> <li>The metrics are assigned to EEA Member State based on the primary country targeting of the ad.</li> <li>In total, 7 ads were restricted between 1 July – 31 December 2023.</li> <li>The following factors may contribute to the number of ads reported by LinkedIn being low: <ul> <li>LinkedIn is primarily a business-to-business advertising platform - that is, businesses marketing their products and services to other businesses and members in a professional capacity.</li> <li>Because of the business-to-business nature of LinkedIn's advertising platform, ads on LinkedIn may cost more than ads placed in other settings, impacting the ads run on LinkedIn.</li> </ul> </li> </ul>	Microsoft Advertising assesses the impact of its actions by reporting on the ads that were loaded in the Microsoft Advertising campaign system but that we blocked from being displayed across our network, and estimating the number of individual ad impressions that such ads would have generated (i.e., ads that would have been served to consumers). "Unique Ads" are the individual advertisements (or ad creatives) loaded in the Microsoft Advertising campaign system that we blocked. "Blocked Impressions" means the individual ad impressions that such Unique Ads would have generated if not blocked. "Unique Domains" means the web domain or URLs that the Unique Ads would have directed customers to. For example, if Advertiser A loads 10 ads in the campaign system promoting two websites associated with disinformation that would have been displayed 100 times, Microsoft Advertising will report 10 Unique Ads, 100 Blocked Impressions, and two Unique Domains. In practice, however, no consumer ever see any of these ads. "Account suspension" means suspending access to the Microsoft Advertising services to advertisers who willingly or repeatedly violate our terms or policies prohibiting phishing, malware, or payment Instrument fraud. Suspended customers cannot serve ads until they redress the violation. The reporting period is July 1, 2023 to December 31, 2023.

	The number of ads LinkedIn restricted under the misinformation policies	Unique Ads	Blocked	Unique	Account
	in QRE 2.1.1 between 1 July - 31 December 2023		Impressions	Domains	suspension
Member States					
Austria	0	149,314	1,088,400	3	931
Belgium	0	729,946	4,591,329	4	1,427
Bulgaria	0	73,334	256,081	2	373
Croatia	0	73,435	145,436	2	113
Cyprus	0	91,302	326,225	2	175
Czech Republic	0	1,065,362	366,016	2	621
Denmark	0	123,627	879,064	3	1,277
Estonia	0	46,196	80,738	2	889
Finland	0	353,655	408,633	3	960
France	1	96,126	52,023,672	7	6,386
Germany	3	2,197,667	27,254,107	6	23,203
Greece	0	102,423	516,546	2	372
Hungary	0	55,550	381,798	2	149
Ireland	0	85,079	3,569,570	3	408
Italy	2	252,985	3,551,970	5	3,628
Latvia	0	42,089	129,170	2	274
Lithuania	0	263,048	140,949	2	128
Luxembourg	0	14,225	248,857	2	118
Malta	0	59,866	439,723	2	85
Netherlands	1	93,543	5,557,107	4	4,776
Poland	0	57,057	1,438,951	2	2,470
Portugal	0	81,365	884,977	2	279
Romania	0	305,164	595,882	2	522
Slovakia	0	116,189	179,558	2	186
Slovenia	0	160,722	93,622	2	350
Spain	0	124,649	19,851,516	3	3,748
Sweden	0	108,808	2,165,388	4	1,008
Iceland	0	171,484	73,178	2	86

Liechtenstein	0	43,402	10,712	2	15
Norway	0	61,043	1,007,499	2	846
Total EU	7	7,026,152	127,165,285	77	54,856
Total EEA	7	7,198,655	128,256,674	83	55,803

Measure 2.2	LinkedIn	Microsoft Advertising
	LinkedIn works with numerous partners to facilitate the flow of	Microsoft Advertising employs dedicated operational support and engineering
	information to tackle purveyors of disinformation, including	resources to enforce its advertising policies detailed below, combining
	disinformation spread by state-sponsored and institutional actors.	automated and manual enforcement methods to prevent or take down
		advertisements that violate its policies. Every ad loaded into the Microsoft
	LinkedIn maintains an internal Threat Prevention and Defense team	Advertising system is subject to these enforcement methods, which leverage
	composed of threat investigators and intelligence personnel to address	machine-learning techniques, automated screening, the expertise of its
	disinformation. This team works with various other internal teams,	operations team, and dedicated user safety experts. In addition, Microsoft
	including our Artificial Intelligence modeling team, to develop leads into	Advertising conducts a manual review of all advertisements flagged to its
	threat actor campaigns. The leads are then manually verified, and	customer support team and removes advertisements that violate its policies.
	confirmed TTPs (Tactics, Techniques, and Procedures) and IOCs	
	(Indicators of Compromise) relating to threat actors are shared with	As stated in our policies on Disinformation, "We may use a combination of
	other external stakeholders, including, for example, industry peers. Any	internal signals and trusted third-party data or information sources to reject,
QRE 2.2.1	associated disinformation content is verified by our internal or external	block, or take down ads or sites that contain disinformation or send traffic to
<b>_</b>	fact-checkers, and coordinated inauthentic behaviours (CIBs) are also	pages containing disinformation." Microsoft Advertising partners with
	removed by our Threat Prevention and Defense team.	Information Integrity experts, such as NewsGuard and GDI, as source and
		references of Disinformation domains. Microsoft Advertising is actively blocking
	LinkedIn, along with its parent company, Microsoft, is heavily involved in	domains that these sources deem as spreading Disinformation.
	threat exchanges. These threat exchanges take various forms, such as: 1)	
	regular discussion amongst industry peers to discuss high-level trends	In addition to the policy on Disinformation, Microsoft Advertising's Misleading
	and campaigns; and, 2) one-on-one engagement with individual peer	Content Policies prohibit advertising content that is misleading, deceptive,
	companies to discuss TTPs and IOCs. This exchange of information leads	fraudulent, or that can be harmful to users, including advertisements that
	to a better understanding of the incentives of sophisticated and well-	contain unsubstantiated claims, or that falsely claim or imply endorsements or
	funded threat actors and how they evolve their TTPs to achieve those	affiliations with third party products, services, governmental entities, or
	goals, which assists us in their identification and removal.	organisations. 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

	LinkedIn always stands ready to receive and investigate any leads we receive from peers and other external stakeholders. In addition to one- on-one engagement with peers, we also consume intelligence from vendors and investigate any TTPs and IOCs made available in peer disclosures. In turn, we also regularly release information about disinformation on our platform in publicly-available transparency reports and blog posts.	users onto sites using questionable or misleading tactics (e.g., by prohibiting advertisements that lead users to sites that misrepresent the origin or intent of their content).
Measure 2.3	LinkedIn	Microsoft Advertising
	All advertising that runs on LinkedIn's platform is subject to LinkedIn's Advertising Policies. LinkedIn has implemented both automated and manual systems to help ensure that advertising on the platform complies with its Advertising Policies, and that ads that do not comply with its policies are removed.	<ul><li>Please see QRE 2.2.1. Microsoft Advertising blocks sites or domains that our Information Integrity expert partners deem as spreading Disinformation.</li><li>Microsoft Advertising also rejects all ads associated with such domains and instructs its publishing partners to block ads from showing on such domains.</li></ul>
	When an advertiser submits an advertising campaign, the campaign is evaluated by LinkedIn automated systems. If those systems determine a campaign may violate LinkedIn's policies, the campaign is forwarded to LinkedIn's advertising review team for manual review.	
QRE 2.3.1	The advertising review team is trained in LinkedIn's Advertising Policies and dedicated to advertising review. LinkedIn also employs a dedicated team of trainers, who not only support the onboarding of new ad reviewers, but also provide ongoing educational opportunities for reviewers. LinkedIn similarly employs quality assurance analysts, who provide one- on-one coaching, as well as regular monthly forums to discuss reviewers' most frequent challenges. For complex issues, reviewers have direct access to global advertising policy managers through regular office hours and dedicated escalation pathways.	
	LinkedIn members may also report ads that they believe violate LinkedIn's advertising policies, and when members report ads LinkedIn's advertising	

	review team reviews them. To report an ad, members can click on the	
	three-dot icon in the upper right-hand corner of every ad and select the	
	"Hide or report this ad" option. Members are then directed to select a	
	reporting reason, with "Misinformation" provided as a reporting option.	
	Methodology of data measurement:	Methodology of data measurement:
	The table below reports metrics concerning ads LinkedIn restricted under	Microsoft Advertising removed ads after the ads were shown to consumers
	the misinformation policies in QRE 2.1.1. The metrics include: (1) the	during the relevant reporting period. Since we remove ads running in all
	number of ads LinkedIn restricted under the misinformation policies in	countries, the "Nr. of ads removed" count is the same across all Member States.
	QRE 2.1.1 between 1 July - 31 December 2023, broken out by EEA	The low figures are due to the proactive nature of our actions, since we block
	Member State; (2) the number of impressions those ads received before	most disinformation content before it shows to consumers.
	they were restricted. The metrics are assigned to EEA Member States	
	based on the primary country targeting of the ad.	Please see SLI 2.1.1 (Unique Ads) for the number of ads prohibited.
	In total, 7 ads were restricted between 1 July - 31 December 2023.	The reporting period is July 1, 2023 to December 31, 2023.
<b>CU 2 2 4</b>	The following factors may contribute to the number of ads reported by	
SLI 2.3.1	LinkedIn being lower than other platforms:	
	- LinkedIn is primarily a business-to-business advertising platform	
	that is, businesses marketing their products and services to other	
	businesses and members in a professional capacity.	
	- Because of the business-to-business nature of LinkedIn's	
	advertising platform, ads on LinkedIn may cost more than ads on	
	other platforms, impacting the ads run on LinkedIn.	
	- The size of LinkedIn's advertising business, and the number of ads	
	run on LinkedIn, may be smaller than other platforms.	

	The number of ads LinkedIn restricted under the	The number of impressions the ads received before they were	Nr of ads removed (as well as reach of ads before they were successfully	Nr of ads prohibited (same as Unique Ads in SLI 2.1.1)
	misinformation policies in QRE 2.1.1 between 1 July – 31 December 2023	restricted	removed)	
Member States				
Austria	0	0	190	149,314
Belgium	0	0	190	729,946
Bulgaria	0	0	190	73,334
Croatia	0	0	190	73,435
Cyprus	0	0	190	91,302
Czech Republic	0	0	190	1,065,362
Denmark	0	0	190	123,627
Estonia	0	0	190	46,196
Finland	0	0	190	353,655
France	1	0	190	96,126
Germany	3	3,111	190	2,197,667
Greece	0	0	190	102,423
Hungary	0	0	190	55,550
Ireland	0	0	190	85,079
Italy	2	0	190	252,985
Latvia	0	0	190	42,089
Lithuania	0	0	190	263,048
Luxembourg	0	0	190	14,225
Malta	0	0	190	59,866
Netherlands	1	0	190	93,543
Poland	0	0	190	57,057
Portugal	0	0	190	81,365
Romania	0	0	190	305,164
Slovakia	0	0	190	116,189
Slovenia	0	0	190	160,722
Spain	0	0	190	124,649

Sweden	0	0	190	108,808
Iceland	0	0	190	171,484
Liechtenstein	0	0	190	43,402
Norway	0	0	190	61,043
Total EU	7	3,111	190	7,026,152
Total EEA	7	3,111	190	7,198,655

Measure 2.4	LinkedIn	Microsoft Advertising
QRE 2.4.1	<ul> <li>When LinkedIn rejects or restricts an ad for violation of its policies, as described in QRE 2.3.1, LinkedIn sends the advertiser an email notification. The email notification outlines the rejection reason and advertising policy that the ad has violated.</li> <li>The notification also provides advertisers instructions regarding how they can address the violation, including by revising the ad in LinkedIn Campaign Manager to address the violations, or by contacting their sales representative or LinkedIn customer support if they require clarification or believe there has been a mistake.</li> <li>Because advertisers can address rejections a number of ways – by revising and resubmitting the advertisement, by creating a new advertisement that complies with LinkedIn's policies, or by contacting their LinkedIn sales representative or customer support – LinkedIn does not report "appeal" and "appeal grant" metrics for ad rejections. LinkedIn has provided metrics on the number of ad restrictions as part of SLI 2.3.1 above.</li> </ul>	Microsoft Advertising notifies its advertisers customers of policy violations through the following: - Prompts in the campaign User Interface (UI) - Email notifications (for example, for account suspension) - Notifications from the assigned Account representatives, as applicable Advertisers may appeal an editorial decision through the conflict-resolution process described here: <u>How do I challenge a disapproval? (microsoft.com)</u>
SLI 2.4.1	Methodology of data measurement: LinkedIn does not report "appeal" and "appeal grant" metrics for ad rejections as outlined in our response to QRE 2.4.1. LinkedIn has	Methodology of data measurement: Microsoft Advertising tracks appeals (at advertising creative and keyword level) in the aggregate and not at the Member State level due to the nature

	provided metrics on the number c	f ad restrictions as part of SLI 2.3.1	of its services. We are providing the aggregate global value of 835 total	
	above. a		appeals during the relevant reporting period.	
	Nr of appeals	Proportion of appeals that led to	Nr of appeals	Proportion of appeals that led to a
		a change of the initial decision		change of the initial decision
Global	N/A		835	275

		II. Scrutiny of Ad Plac	ements	
		Commitment 3	3	
players, expanding to	organisations active in	ng and placing digital advertising commit n the online monetisation value chain, suc im to increase the effectiveness of scrutiny	h as online e-payment services, e-co	ommerce platforms and relevant
	C.3	M 3.1	М 3.2	М 3.3
We signed up to the	LinkedIn	LinkedIn	LinkedIn	LinkedIn
following measures of	Microsoft	Microsoft Advertising	Microsoft Advertising	Microsoft Advertising

	LinkedIn	Microsoft Advertising
In line with this commitment, did you deploy new implementation measures (e.g., changes to	No	No
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	N/A	N/A
Do you plan to put further implementation measures in place in the next 6 months to	No	No
substantially improve the maturity of the implementation of this commitment? [Yes/No]		

this commitment:

Advertising

If yes, which further implementation measures do you plan to put in place in the next 6	Planned Implementation Measures	Planned Implementation Measures
months?		
	N/A	N/A

Measure 3.1	LinkedIn	Microsoft Advertising
QRE 3.1.1	As discussed as part of QRE 2.2.1, LinkedIn works with numerous	Microsoft Advertising partners with Information Integrity experts as source
	partners to facilitate the flow of information to tackle purveyors of	and references of Disinformation domains.
	disinformation, including disinformation spread by state-sponsored	
	and institutional actors.	Microsoft Advertising is a member and an active participant in several trade groups, including the IAB Europe and European Interactive Digital
	LinkedIn maintains an internal Threat Prevention and Defense team	Advertising Alliance (EDAA).
	composed of threat investigators and intelligence personnel to	
	address disinformation. This team works with various other internal	As part of the broader Microsoft engagements, Microsoft Advertising
	teams, including our Artificial Intelligence modeling team, to develop	participates in threat exchange discussions with industry peers and trusted
	leads into threat actor campaigns. The leads are then manually	third parties which increase the effectiveness of our enforcement actions.
	verified, and confirmed TTPs (Tactics, Techniques, and Procedures)	
	and IOCs (Indicators of Compromise) relating to threat actors are	
	shared with other external stakeholders, including, for example,	
	industry peers. Any associated disinformation content is verified by	
	our internal or external fact-checkers, and coordinated inauthentic	
	behaviours (CIBs) are also removed by our Threat Prevention and	
	Defense team.	
	LinkedIn, along with its parent company, Microsoft, is heavily involved	
	in threat exchanges. These threat exchanges take various forms, such	
	as: 1) regular discussion amongst industry peers to discuss high-level	
	trends and campaigns; and, 2) one-on-one engagement with	
	individual peer companies to discuss TTPs and IOCs. This exchange	
	of information leads to a better understanding of the incentives of	
	sophisticated and well-funded threat actors and how they evolve	
	their TTPs to achieve those goals, which assists us in their	
	identification and removal.	

	LinkedIn always stands ready to receive and investigate any leads we receive from peers and other external stakeholders. In addition to one-on-one engagement with peers, we also consume intelligence from vendors and investigate any TTPs and IOCs made available in peer disclosures. In turn, we also regularly release information about disinformation on our platform in publicly-available transparency reports and blog posts.	
Measure 3.2	LinkedIn	Microsoft Advertising
QRE 3.2.1	Please see the response to QRE 3.1.1. In addition, LinkedIn is a member of and participates in GARM.	Please see the response to QRE 3.1.1.
Measure 3.3	LinkedIn	Microsoft Advertising
QRE 3.3.1	Please see the response to QRE 3.1.1. In addition, as discussed in response to QRE 1.6.1, LinkedIn partners with companies including Integral Ad Science and DoubleVerify to help evaluate and filter advertising inventory on LinkedIn Audience Network publisher sites that falls short of standards, such as GARM's brand safety floor.	Please see the response to QRE 3.1.1.

# III. Political Advertising Commitments 4 - 13

	III. Political Advertising		
	Commitment 4		
	Relevant Signatories commit to adopt a common definition of "politica	l and issue advertising".	
	C.4	M 4.1	M 4.2
We signed up to the following measures of this commitment:	LinkedIn Microsoft Advertising	LinkedIn Microsoft Advertising	LinkedIn Microsoft Advertising

	LinkedIn	Microsoft Advertising
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	No
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	N/A	N/A
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	LinkedIn will review its policies and	Microsoft Advertising will review its
	enforcement in light of the	policies and enforcement in light of
	upcoming Regulation on	the upcoming Regulation on
	Transparency and Targeting of	Transparency and Targeting of
	Political Advertising and the	Political Advertising and the
	associated common guidance to be	associated common guidance to be
	issued in accordance with Art. 8.2 of	issued in accordance with Art. 8.2 of
	the Regulation.	the Regulation.

Measure 4.1	LinkedIn	Microsoft Advertising
Measure 4.2	LinkedIn	Microsoft Advertising
QRE 4.1.1 (for	LinkedIn's Advertising Policies do not allow political advertising, and	Microsoft Advertising policies prohibit ads for election-related content,
QRE 4.1.1 (for measures 4.1 and 4.2)	LinkedIn's <u>Advertising Policies</u> do not allow political advertising, and LinkedIn has not allowed political advertising since 2018. Among other things, LinkedIn Advertising policies prohibit "ads advocating for or against a particular candidate, party, or ballot proposition or otherwise intended to influence an election outcome" and "ads fundraising for or by political candidates, parties, political action committees or similar organisations, or ballot propositions." In addition, LinkedIn's Advertising Policies prohibit certain types of advertisements that might be considered issue based. For example, "ads exploiting a sensitive political issue even if the advertiser has no explicit political agenda" are also prohibited.	Microsoft Advertising policies prohibit ads for election-related content, political candidates, parties, ballot measures and political fundraising globally; similarly, ads aimed at fundraising for political candidates, parties, political action committees ("PACs"), and ballot measures also are barred. All Microsoft and third-party services that rely on Microsoft Advertising to serve advertisements on their platforms benefit from these robust, and robustly enforced set of policies. Furthermore, Microsoft prohibits political advertising across Microsoft media properties and platforms. Microsoft Advertising's policies also prohibit certain types of advertisements that might be considered issue-based. More specifically, "advertising that exploits political agendas, sensitive political issues or uses 'hot button' political issues or names of prominent politicians is not allowed regardless of whether the advertiser has a political agenda," and "advertising that
		exploits sensitive political or religious issues for commercial gain, or promote extreme political or extreme religious agendas or any known associations with hate, criminal or terrorist activities" is also prohibited. See here: <u>Political content - Microsoft Advertising</u>
QRE 4.1.2 (for	Microsoft looks forward to the upcoming Regulation on Transparency	Microsoft looks forward to the upcoming Regulation on Transparency and
measures 4.1 and 4.2)	and Targeting of Political Advertising and the associated common guidance to be issued in accordance with Art. 8.2 of the Regulation, as well as participating in related upcoming work to be organized in the context of the Taskforce.	Targeting of Political Advertising and the associated common guidance to be issued in accordance with Art. 8.2 of the Regulation, as well as participating in related upcoming work to be organized in the context of the Task-force.

### III. Political Advertising

<u> </u>			
Com	mitm	ient 5	

Relevant Signatories commit to apply a consistent approach across political and issue advertising on their services and to clearly indicate in their advertising policies the extent to which such advertising is permitted or prohibited on their services.

	C.5	M 5.1
We signed up to the following	LinkedIn	LinkedIn
measures of this commitment:	Microsoft Advertising	Microsoft Advertising

	LinkedIn	Microsoft Advertising
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	No
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	The current implementation	The current implementation
	measures employed by LinkedIn	measures employed by Microsoft
	meet this commitment.	Advertising suffice to meet this
		commitment.
Do you plan to put further implementation measures in place in the next 6 months to	No	No
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	Not applicable as LinkedIn currently	Not applicable as Microsoft
	prohibits all political advertising, as	Advertising currently prohibits all
	outlined under QRE 5.1.1.	political advertising, as outlined
		under QRE 5.1.1.

Measure 5.1	LinkedIn	Microsoft Advertising
QRE 5.1.1	LinkedIn's Advertising Policies do not allow political advertising, and	Microsoft Advertising policies prohibit ads for election-related content,
	LinkedIn has not allowed political advertising since 2018.	political candidates, parties, ballot measures and political fundraising
		globally; similarly, ads aimed at fundraising for political candidates, parties,
	Among other things, LinkedIn Advertising policies prohibit "ads	political action committees ("PACs"), and ballot measures also are barred. All
	advocating for or against a particular candidate, party, or ballot	Microsoft and third-party services that rely on Microsoft Advertising to serve
	proposition or otherwise intended to influence an election outcome"	advertisements on their platforms benefit from these robust, and robustly
	and "ads fundraising for or by political candidates, parties, political	enforced, set of policies.
	action committees or similar organisations, or ballot propositions." In	
	addition, LinkedIn's Advertising Policies prohibit certain types of	Furthermore, Microsoft prohibits political advertising across Microsoft
	advertisements that might be considered issue based. For example,	media properties and platforms. Microsoft Advertising's policies also
	"ads exploiting a sensitive political issue even if the advertiser has no	prohibit certain types of advertisements that might be considered issue-
	explicit political agenda" are also prohibited.	based. More specifically, "advertising that exploits political agendas,
		sensitive political issues or uses 'hot button' political issues or names of
		prominent politicians is not allowed regardless of whether the advertiser has
		a political agenda," and "advertising that exploits sensitive political or
		religious issues for commercial gain, or promote extreme political or
		extreme religious agendas or any known associations with hate, criminal or
		terrorist activities" is also prohibited.

	III. Political Advertising					
	Commitment 6					
-	Relevant Signatories commit to make political or issue ads clearly labelled and distinguishable as paid-for content in a way that allows users to understand that the content displayed contains political or issue advertising					nderstand
	C.6	M 6.1	M 6.2	M 6.3	M 6.4	M 6.5
We signed up to the following measures of this commitment:	Commitment 6 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they do not allow political or issue-based advertising as set out in more detail under measure 5.1.	N/A	N/A	N/A	N/A	N/A

### **III. Political Advertising**

Commitment 7

Relevant Signatories commit to put proportionate and appropriate identity verification systems in place for sponsors and providers of advertising services acting on behalf of sponsors placing political or issue ads. Relevant signatories will make sure that labelling and user-facing transparency requirements are met before allowing placement of such ads.

	C.7	M 7.1	M 7.2	M 7.3	M 7.4
We signed up to the following	LinkedIn	N/A	N/A	LinkedIn	N/A
measures of this commitment:	Microsoft Advertising			Microsoft Advertising	

	LinkedIn	Microsoft Advertising
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	No
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	Not applicable
Do you plan to put further implementation measures in place in the next 6 months to	No	No
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6	Planned Implementation Measures	Planned Implementation Measures
months?	Not applicable	Not applicable

Measure 7.3	LinkedIn	Microsoft Advertising
QRE 7.3.1	As set out in QRE 5.1.1, LinkedIn's Advertising Policies prohibit	As set out in QRE 5.1.1, Microsoft Advertising prohibits ads for election-
	political advertising. Before submitting a campaign, customers must	related content, political candidates, parties, ballot measures and political
	agree that their ad complies with LinkedIn's policies.	fundraising globally; similarly, ads aimed at fundraising for political
		candidates, parties, political action committees ("PACs"), and ballot
	As detailed in QRE 2.3.1, LinkedIn has implemented both automated	measures also are barred. All Microsoft and third-party services that rely on
	and manual systems to help ensure that advertising on the platform	

	<ul> <li>complies with its Advertising Policies, and that ads that do not comply with its policies are removed. These enforcement systems apply equally to prohibited political advertising, as well as other violations of LinkedIn's Advertising Policies.</li> <li>In addition to LinkedIn's pre-emptive enforcement, LinkedIn members may also report ads that they believe violate LinkedIn's advertising policies, and when members report ads LinkedIn's advertising review team reviews them. To report an ad, members can click on the three-dot icon in the upper right of every ad and select the "Hide or report this ad" option.</li> </ul>	Microsoft Advertising to serve advertisements on their platforms benefit from these robust, and robustly enforced, set of policies. Microsoft Advertising's policies also prohibit certain types of advertisements that might be considered issue-based. More specifically, "advertising that exploits political agendas, sensitive political issues or uses 'hot button' political issues or names of prominent politicians is not allowed regardless of whether the advertiser has a political agenda," and "advertising that exploits sensitive political or religious issues for commercial gain, or promote extreme political or extreme religious agendas or any known associations with hate, criminal or terrorist activities" is also prohibited. In addition, to comply with local laws in Canada and France, Microsoft Advertising has prohibited advertising content related to debates of general interest linked to an electoral campaign in those jurisdictions.
		See here: Political content - Microsoft Advertising
QRE 7.3.2	As set out in QRE 5.1.1, LinkedIn's Advertising Policies prohibit political advertising. Ads that do not comply with LinkedIn's Advertising Policies are removed.	Microsoft Advertising does not offer its advertising services to customers or partners that may promote political content. For example, a political party would be made aware by our customer support team that it cannot run political ads campaigns through our ad network.
		Microsoft Advertising employs dedicated operational support and engineering resources to enforce restrictions on political advertising using a combination of proactive and reactive mechanisms. On the proactive side, Microsoft Advertising has implemented several processes designed to identify and block political ads from showing across its advertising network, including restrictions on certain terms and from certain domains. For example, we compile a list of terms comprised of known political parties, candidates, and ballot measures and block any ads that may be otherwise shown on a search engine search for such terms. On the reactive side, if Microsoft Advertising becomes aware that an ad suspected of violating its policies is being served to our publishers—for instance, because a user flagged that ad to our customer support team—the offending ad is

	promptly reviewed and, if it violates our policies, taken down. Users can report advertising which may be in violation of the Microsoft Advertising policies through the publisher-specific reporting form or via this form: <u>Low</u> <u>quality ad submission &amp; escalation - Microsoft Advertising</u>
	The above actions apply across all the websites that use Microsoft Advertising to serve ads on their properties, whether owned and operated by Microsoft (like Bing) or third-party websites.

III. Political Advertising				
Commitment 8				
Relevant Signato	Relevant Signatories commit to provide transparency information to users about the political or issue ads they see on their service.			
		M 8.1	M 8.2	
We signed up to the following	Commitment 8 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they do not allow political or	<b>M 8.1</b> N/A	M 8.2	

III. Political Advertising					
	Commitment 9				
Relevant Signatories comn	Relevant Signatories commit to provide users with clear, comprehensible, comprehensive information about why they are seeing a political or issue ad.				
	C.9	M 9.1	M 9.2		
We signed up to the following	Commitment 9 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they do not allow political or	N/A	N/A		
measures of this commitment:	issue-based advertising as set out in more detail under measure 5.1.				

III. Political Advertising				
Commitment 10				
<b>v</b>	to maintain repositories of political or issue advertising and ensure their currentness, completeness, usabi and issue advertising served, along with the necessary information to comply with their legal obligations ar e.	,	-	
	C.10	M 10.1	M 10.2	
We signed up to the following	Commitment 10 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they do not allow political or	N/A	N/A	
measures of this commitment:	issue-based advertising as set out in more detail under measure 5.1.			

III. Political Advertising				
Commitment 11				
Relevant Signatories commit to provide application programming interfaces (APIs) or other interfaces enabling users and researchers to perform customised searches within their ad repositories of political or issue advertising and to include a set of minimum functionalities as well as a set of minimum search criteria for the application of APIs or other interfaces."				
C.11	M 11.1	M 11.2	M 11.3	M 11.4
Commitment 11 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they	N/A	N/A	N/A	N/A
do not allow political or issue-based advertising as set out in more detail under measure 5.1.				

III. Political Advertising			
Commitment 12			
Relevant Signatories commit to increase oversight of political and issue advertising and constructively assist, as appropriate, in the creation, implementation and improvement of political or issue advertising policies and practices.			
C.12	M 12.1	M 12.2	M 12.3

We signed up to the following	Commitment 12 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they do not allow	N/A	N/A	N/A
measures of this commitment:	political or issue-based advertising as set out in more detail under measure 5.1.			

Commitment 13

Relevant Signatories agree to engage in ongoing monitoring and research to understand and respond to risks related to Disinformation in political or issue advertising.

	C.13	M 13.1	M 13.2	M 13.3
We signed up to the following	Commitment 13 is not relevant or pertinent for LinkedIn and Microsoft Advertising as they do not allow	N/A	N/A	N/A
measures of this commitment:	political or issue-based advertising as set out in more detail under measure 5.1.			

## IV. Integrity of Services Commitments 14 - 16

### **IV. Integrity of Services**

#### Commitment 14

In order to limit impermissible manipulative behaviours and practices across their services, Relevant Signatories commit to put in place or further bolster policies to address both misinformation and disinformation across their services, and to agree on a cross-service understanding of manipulative behaviours, actors and practices not permitted on their services. Such behaviours and practices, which should periodically be reviewed in light with the latest evidence on the conducts and TTPs employed by malicious actors, such as the AMITT Disinformation Tactics, Techniques and Procedures Framework, include:

The following TTPs pertain to the creation of assets for the purpose of a disinformation campaign, and to ways to make these assets seem credible:

- 1. Creation of inauthentic accounts or botnets (which may include automated, partially automated, or non-automated accounts)
- 2. Use of fake / inauthentic reactions (e.g. likes, up votes, comments)
- 3. Use of fake followers or subscribers
- 4. Creation of inauthentic pages, groups, chat groups, fora, or domains
- 5. Account hijacking or impersonation

The following TTPs pertain to the dissemination of content created in the context of a disinformation campaign, which may or may not include some forms of targeting or attempting to silence opposing views. Relevant TTPs include:

- 6. Deliberately targeting vulnerable recipients (e.g. via personalized advertising, location spoofing or obfuscation)
- 7. Deploy deceptive manipulated media (e.g. "deep fakes", "cheap fakes"...)
- 8. Use "hack and leak" operation (which may or may not include doctored content)
- 9. Inauthentic coordination of content creation or amplification, including attempts to deceive/manipulate platforms algorithms (e.g. keyword stuffing or inauthentic posting/reposting designed to mislead people about popularity of content, including by influencers)
- 10. Use of deceptive practices to deceive/manipulate platform algorithms, such as to create, amplify or hijack hashtags, data voids, filter bubbles, or echo chambers
- 11. Non-transparent compensated messages or promotions by influencers
- 12. Coordinated mass reporting of non-violative opposing content or accounts

	C.14	M 14.1	M 14.2	M 14.3
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search	Bing Search	Bing Search

	LinkedIn	Bing Search
n line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
our terms of service, new tools, new policies, etc)? [Yes/No]		
f yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	Bing has refined and improved
		existing defensive search and AI
		mitigations. Bing regularly
		evaluates and implements
		additional measures and builds
		upon existing policies, practices,
		and mitigations to help address
		TTPs that could potentially be used
		to help spread disinformation, as
		discussed throughout this report.
		Bing has also improved content
		provenance measures for Copilot i
		Bing. The Copilot in Bing feature to
		create images includes C2PA
		Content Credentials, developed by
		the Coalition for Content
		Provenance and Authenticity
		(C2PA), which sets technical
		standards for certifying the source
		and history (or provenance) of
		media content.

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No	Microsoft also recently launched a new web page – <u>Microsoft-2024</u> <u>Elections</u> – where political candidates can report alleged deepfakes of themselves to Microsoft. Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	Not applicable	Microsoft's Content Credentials as a Service will be launched in private preview for political campaigns and elections purposes. This tool enables campaigns to add Content Credentials to their digital media. It will be available to select private preview customers at no cost for the 2024 election cycle and will be the first-of-a-kind SaaS application for adding content credentials to images and video.
		Bing also regularly reviews and evaluates policies and practices related to Copilot in Bing and may adjust and update policies as needed. Bing's existing programs are already designed to address these issues, but Bing regularly evaluates the efficacy of its measures and endeavors to
		improve and work to respond

	quickly to new threats or issues as
	they arise.

Measure 14.1	LinkedIn	Bing Search
QRE 14.1.1	LinkedIn's <u>User Agreement</u> (in particular section 8 <i>LinkedIn "Dos and Don'ts"</i> ) and our <u>Professional Community Policies</u> - which are accepted by every member when joining LinkedIn - detail the impermissible manipulative behaviours and practices that are prohibited on our platform. Fake accounts, misinformation, and inauthentic content are not allowed, and we take active steps to remove it from our platform.	Bing Search is an online search engine, the primary purpose of which is to provide a searchable index of webpages available on the internet to help users find the content they are looking for. Bing Search does not host the content or control the operation, content, or design of indexed websites. Users come to Bing Search with a specific research topic in mind and expect Bing to provide links to the most relevant and authoritative third-party websites on the Internet that are responsive to their search terms. Bing Search does not allow users to post and share content or otherwise enable content to go "viral" through user-to-user exchanges of information on Bing.

employing social media schemes (e.g., fake followers – TTP 3), using inauthentic domains (TTP 4), or keyword stuffing (TTP 9) are considered abusive practices that are addressed in Bing's ranking system and Webmaster Guidelines. In addition, in connection with generative AI features, Microsoft has implemented measures intended to address TTP No. 7 (related to deceptive deepfakes), which are discussed in more detail below.

As noted above, Bing's primary mechanism for combatting manipulative behaviors in search results is via its ranking algorithms and systems designed to identify and combat attempts to abuse search engine optimization techniques (i.e., spam). Bing Search describes the main parameters of its ranking systems in depth in <u>How Bing Delivers Search Results</u>. Abusive techniques and examples of prohibited SEO activities are described in more detail in the <u>Bing Webmaster Guidelines</u>. These pages provide a deeper, more detailed understanding of how Bing search algorithms work that cannot be fully detailed here given the space limitations of this report and complexity of how search engines rank websites.

As described in these documents, Bing's ranking algorithms are designed to identify and prioritize the highest quality, most authoritative content available online that is relevant to the user's query, and by preventing abusive search engine optimization techniques (spam).

One of the key ranking techniques Bing uses to prevent low quality or misinformative websites from returning high in search results is through the "quality and credibility" score. Determining the quality and credibility (QC) of a website includes evaluating the clarity of purpose of the site, its usability, and presentation. QC also consists of an evaluation of the page's "authority", which includes factors such as:

 Reputation: What types of other websites link to the site? A wellknown news site is considered to have a higher reputation than a brand-new blog.dLevel of discourse: Is the purpose of the content solely to cause harm to individuals or groups of people? For

<ul> <li>example, a site that promotes violence or resorts to name-calling or bullying will be considered to have a low level of discourse, and therefore lower authority, than a balanced news article.</li> <li>Level of distortion: How well does the site differentiate fact from opinion? A site that is clearly labeled as satire or parody will have more authority than one that tries to obscure its intent.</li> <li>Origination and transparency of the ownership: Is the site reporting first-hand information, or does it summarize or republish content from others? If the site doesn't publish original content, do they attribute the source? A first-hand account published on a personal blog could have more authority than unsourced content.cln addition to its ranking algorithms, Bing Search's general abuse/spam policies prohibit certain practices intended to manipulate or deceive the Bing Search algorithms. These policies include prohibitions intended to address the TTPs employed by malicious actors in the spread of disinformation. Pursuant to the Webmaster Guidelines, Bing may take action on websites employing spam tactics (such as social media schemes, keyword stuffing, malicious behavior, cloaking, link schemes, or misleading structured data markups) or that otherwise violate the Webmaster Guidelines, including by applying ranking penalties (such as demoting a website) or delisting a website from the index. d</li> <li>Note that it is not feasible to distinguish between general spam tactics and spam tactics employed by malicious actors specifically for the purpose of spreading disinformation. Therefore, Bing Search has not presented data on the amount of spam detected and actioned under its policies since these figures are indicative of actions taken toward spam verall and presently cannot be used to provide an accurate assessment of whether it pertains to spam used in connection with disinformation campaigns or spam used for another purpose (e.g., phishing).</li> </ul>	
spam tactics employed by malicious actors specifically for the purpose of spreading disinformation. Therefore, Bing Search has not presented data on the amount of spam detected and actioned under its policies since these figures are indicative of actions taken toward spam overall and presently cannot be used to provide an accurate assessment of whether it pertains to spam used in connection with disinformation campaigns or spam used for	<ul> <li>bullying will be considered to have a low level of discourse, and therefore lower authority, than a balanced news article.</li> <li>Level of distortion: How well does the site differentiate fact from opinion? A site that is clearly labeled as satire or parody will have more authority than one that tries to obscure its intent.</li> <li>Origination and transparency of the ownership: Is the site reporting first-hand information, or does it summarize or republish content from others? If the site doesn't publish original content, do they attribute the source? A first-hand account published on a personal blog could have more authority than unsourced content.cln addition to its ranking algorithms, Bing Search's general abuse/spam policies prohibit certain practices intended to manipulate or deceive the Bing Search algorithms. These policies include prohibitions intended to address the TTPs employed by malicious actors in the spread of disinformation. Pursuant to the Webmaster Guidelines, Bing may take action on websites employing spam tactics (such as social media schemes, keyword stuffing, malicious behavior, cloaking, link schemes, or misleading structured data markups) or that otherwise violate the Webmaster Guidelines, including by applying ranking penalties (such as demoting a website) or delisting a website from</li> </ul>
	spam tactics employed by malicious actors specifically for the purpose of spreading disinformation. Therefore, Bing Search has not presented data on the amount of spam detected and actioned under its policies since these figures are indicative of actions taken toward spam overall and presently cannot be used to provide an accurate assessment of whether it pertains to spam used in connection with disinformation campaigns or spam used for

Generative AI Features (Copilot in Bing)
In February 2023, Microsoft launched new generative AI features, including
an AI-enhanced web search experience that allows users to quickly and easily
obtain answers to their queries through a chat experience that is "grounded"
in high authority web search results (meaning chat answers include links to
supporting webpages). These AI features were originally known as Bing Chat
and Image Creator, but since have undergone continued improvements and
have meaningfully evolved into a new distinct family of AI services under the
brand Microsoft Copilot.
Copilot is an AI-powered research assistant, planner, and creative partner for
users across the Microsoft ecosystem, and is offered as a standalone service
and with associated endpoints in other Microsoft services, designed to help
users complete tasks in a variety of productivity scenarios. One such
endpoint is in Bing Search, which integrates Copilot functionality in order to
provide users with a modern, natural language-based search interface
("Copilot in Bing").
Copilot in Bing continues to rely on the same infrastructure and mitigations
previously discussed in Microsoft's last report. As we described in the last
report, Copilot in Bing combines traditional search functionality with the
capabilities of large language models so that users can ask more complex
and nuanced questions and engage in more natural conversations on search
topics. Much like a research librarian, Copilot in Bing uses AI to concisely
summarize relevant information in search results, with links to supporting
webpages where users can continue their research. Users can also use
Copilot to generate creative content, such as code, poems, jokes, stories, and
images.
Copilot in Bing's primary functionality is, like traditional Bing search, to
provide users with links to third party content responsive to their search
queries. As such, the ranking algorithms and spam/abuse policies described
above continue to be Bing's primary defense against manipulation and
abuse, supplemented by interventions designed specifically to address

manipulation in generative AI features. As to answers triggering creative inspiration, Microsoft has worked continuously to improve and adjust safety mitigations, policies, and user experiences within Copilot to minimize the risk they may be used for manipulative purposes. Additional information on how Microsoft approached responsible AI in the development of Copilot in Bing is available in Copilot in Bing: our Approach to Responsible AI. TTP 10 is still the most relevant TTP to Copilot in Bing and generative AI search experiences, as users cannot post or share content directly on the Bing service, even in Copilot in Bing. In addition, Microsoft undertakes specific mitigations to address TTP 7 given the risks that users may attempt to use generative AI to create deep fakes or manipulated media to spread disinformation. Although Bing does not have the ability to monitor third party platforms for publication of content created through Bing's services, Bing has implemented safeguards to help to minimize the risk that bad actors can use Bing generative AI experiences to create mis/disinformation. Microsoft published the Copilot AI Experiences Terms (previously the Bing Conversational Experiences and Image Creator Terms, referred to here as "Supplemental Terms") to advise users on prohibited conduct and content. These Supplemental Terms primarily address TTPs No. 10 and 7 by restricting attempts to create or spread mis/disinformation or deceptive images using Copilot. Users that violate the Supplemental Terms and Code of Conduct may be suspended from the service. In addition, Copilot works to prevent generation of problematic text or images by blocking user prompts that (i) violate the Code of Conduct or (ii) are likely to lead to creation of material that violates the Code of Conduct. Repeated attempts to produce prohibited content or other violations of the Code of Conduct may also result in service or account suspension. For further information as to how Bing Search and Copilot implement these policies see QRE 14.1.2.

QRE 14.1.2	LinkedIn works with numerous partners to facilitate the flow of	As discussed under QRE 14.1.1, TTP No. 10 tends to be the primary
	information to tackle purveyors of disinformation, including	mechanism for manipulation and abuse in the context of search engines and
	disinformation spread by state-sponsored and institutional actors.	is addressed through Bing's ranking systems and abuse policies (for both
		traditional search and Copilot in Bing).
	LinkedIn maintains an internal Trust and Safety team composed of	
	threat investigators and intelligence personnel to address	Blocking content in organic search results based solely on the truth or falsity
	disinformation. This team works with various other internal teams,	of the content can raise significant concerns relating to fundamental rights
	including our Artificial Intelligence modeling team, to develop leads	of freedom of expression and the freedom to receive and impart
	into threat actor campaigns. The leads are then manually verified, and	information. Instead of blocking access to content to address these TTPs,
	confirmed TTPs (Tactics, Techniques, and Procedures) and IOCs	Bing Search focuses on ranking its organic search results so that trusted,
	(Indicators of Compromise) relating to threat actors are shared with	authoritative news and information appears first, and provides tools to help
	other external stakeholders, including, for example, industry peers.	its users evaluate the trustworthiness of certain sites and ensure they are not
		misled or harmed by the content that appears in search results. Bing Search
		takes actions to promote high authority, high quality content and thereby
	behaviours (CIBs) are also removed by our Trust and Safety team.	reduce the impact of misinformation appearing in Bing Search results. This
		includes Bing Search's continued improvement of its ranking algorithms to
		ensure that the most authoritative, relevant content is returned at the top of
		search results, regular review and actioning of disinformation threat
		intelligence, partnership with fact-checking and media literacy
		organizations, contributing to and supporting the research community, and
		implementation and enforcement of clear policies concerning the use of
		manipulative tactics on Bing Search, among other initiatives described
	of the incentives of sophisticated and well-funded threat actors and	
	how they evolve their TTPs to achieve those goals, which assists us in	
	their identification and removal.	Although the Bing Search algorithm endeavors to prioritize relevance,
		quality, and credibility in all scenarios, in some cases Bing Search identifies a
		threat that undermines the efficacy of its algorithms. When this happens,
		Bing Search employs "defensive search" strategies and interventions to
		counteract threats and TTPs in accordance with its trustworthy search
		principles (which are discussed in further detail in the "Our Approach to
	peer disclosures. In turn, we also regularly release information about	<b>-</b>
	policy-violating content on our platform in publicly available	
	transparency reports and blog posts, including for example How	

report published in October 2023.

We're Protecting Members From Fake Profiles, Automated Fake "Defensive search interventions" may include algorithmic interventions (such Account Detection, and An Update on How We Keep Members Safe. | as quality and credibility boosts or demotions of a website), restricting The LinkedIn<u>Community Report</u> also describes actions we take on autosuggest or related search terms to avoid directing users to problematic content that violates our Professional Community Policies and User queries, prioritizing additional features promoting high authority Agreement. It is published twice per year and covers the global information (e.g., Answers, Public Service Announcements), and in limited detection of fake accounts, spam and scams, content violations and cases manual interventions for individual reported issues or broader areas copyright infringements. The most recent reporting period covered 1 more prone to misinformation or disinformation (e.g., elections, January to 30 June 2023. LinkedIn also publishes Transparency pharmaceutical drugs, or COVID-19). Bing actively monitors manipulation Reports covers the EU under the Digital Services Act, with the first trends in identified high-risk areas and deploys mitigation methods as needed to ensure users are provided with high quality, high authority search results.

> In addition to defensive search, Bing Search regularly monitors for other violations of its Webmaster Guidelines, including attempts to manipulate the Bing Search algorithm through prohibited practices such as cloaking, link spamming, keyword stuffing, and phishing. Bing Search dedicates meaningful resources to maintaining the integrity of the platform, promoting high authority, relevant results, and reducing spam (including spam aimed at distributing low authority information and manipulative content). Bing Search utilizes a combination of human intervention and Aldriven analysis to regularly review, detect, and address spam tactics occurring on Bing Search. When Bing Search detects websites deploying manipulative techniques or engaging in spam tactics, those websites may incur ranking penalties or be removed from the Bing Search index altogether.

> Microsoft also works to identify and track nation-state information operations targeting democracies across the world and works with trusted third-party partners, including NewsGuard, the Global Democracy Index (GDI), and Spanish-language news agency EFE, to provide early indicators of narratives, hashtags, or information operations that can be leveraged to inform early detection and defensive search strategies. Through Microsoft's Democracy Forward team and the Microsoft Threat Assessment Center

(MTAC), Microsoft also offers mediums for election authorities, including in the EU and EEA Member States, to have lines of communication with Microsoft to identify possible foreign information operations targeting elections.

The above measures also apply to Copilot in Bing. Responses provided by the Copilot feature are "grounded" in search results, which are based on the same ranking algorithms and moderation infrastructure that are used by Bing's traditional web search, and, as such, benefit from Bing's longstanding safety infrastructure described above. Nonetheless, Microsoft recognizes that generative AI technology may also raise new risks and possibilities of harm that are not present in traditional web search and has supplemented its existing threat identification and mitigation processes with additional risk assessments and mitigation processes based on <u>Microsoft's Responsible AI program</u>.

Microsoft's Responsible AI program is designed to identify potential harms, measure their propensity to occur, and build mitigations to address them. Guided by our <u>Responsible AI Standard</u>, Microsoft sought to identify, measure, and mitigate potential harms and misuse of new generative AI experiences while securing the transformative and beneficial uses that these tools provide. Since its launch initially as Bing Chat and through its evolution to Copilot, Microsoft has implemented a range of safety mitigations to help address, among other things, impermissible content, behaviours, and other TTPs that could potentially be used to create or spread misinformation.

Below are several examples of our iterative approach to identify, measure, and mitigate potential harms, including the spread of misinformation. Additional information, which is updated regularly, can be found in <u>Copilot</u> in <u>Bing: Our approach to responsible Al</u>.

- **Grounded Results.** Copilot in Bing is grounded in web search results. This means that we center the response provided to a user's query on high-

ranking content from the web, and we provide links to websites so that users can learn more and evaluate the credibility and information presented by reviewing the source material. We describe Bing's approach to search ranking, including the main parameters of ranking, in more detail in <u>How</u> <u>Bing Delivers Search Results</u>.

- **Pre-launch and ongoing testing.** Before launching Copilot, Microsoft conducted extensive "red team" testing. A multidisciplinary team of experts conducted numerous rounds of testing to evaluate how well the system responded when pressed to produce harmful responses, surface potential avenues for misuse, and identify capabilities and limitations. Post-release, Copilot experiences are integrated into Microsoft engineering organizations' existing production measurement and testing infrastructure. For example, red team testers from different regions and backgrounds continuously and systematically attempt to compromise the system, and their findings are used to expand the datasets that Microsoft uses for improving the system. More information on Microsoft's approach to red-team testing is available at <u>Microsoft AI Red Team building future of safer AI | Microsoft Security Blog</u>.

- Classifiers, Metaprompting, and Filtering Interventions: Microsoft has created special mitigations in the form of "classifiers" and "metaprompting" to help reduce the risk of certain harms and misuse of generative AI features. *Classifiers* classify text to flag different types of potentially harmful content in search queries, chat prompts, or generated responses. Microsoft uses AI-based classifiers and content filters, which apply to all search results and relevant features; it also designed additional prompt classifiers and content filters specifically to address possible harms raised by new generative AI features such as Copilot in Bing. Flags lead to potential mitigations, such as not returning generated content to the user, diverting the user to a different topic, or redirecting the user to traditional search. Metaprompting involves giving instructions to the model to guide its behavior, including so that the system behaves in accordance with Microsoft's AI Principles and user expectations. Microsoft has also implemented additional filtering and

classifiers to prevent chat responses from returning what Bing considers "low authority" content as part of an answer and to help address impermissible content, behaviours, and other TTPs (e.g., TTP No. 7) that could potentially be used to create or spread misinformation.

- **Content Provenance Tools.** Microsoft also makes it clear that images created in Copilot in Bing are Al-generated by including content provenance information in each image. These content provenance features use cryptographic methods to mark and sign Al-generated content with metadata about its source and history. The invisible digital watermark feature shows the source, time, and date of original creation, and this information cannot be altered. Providing clear indications of image provenance helps reduce the risk of deepfakes (e.g., TTP No. 7) and helps users identify when an image was generated with the assistance of Microsoft generative Al tools. Microsoft has partnered with other industry leaders to create the Coalition for Content Provenance and Authenticity (C2PA) standards body to help develop and apply content provenance standards across the industry.

- **Expanded and Prominent Reporting Functionality.** Copilot in Bing allows users to submit feedback and report their concerns, which are then reviewed by Microsoft's operations teams. Microsoft has made it easy for users to report problematic content they encounter while using generative AI features in Bing by including a "Feedback" portal on the footer of every Bing or Copilot page, with direct links to its "Report a Concern" tool.

- **Regular Improvements Based on Real World Usage.** Microsoft continues to make changes to Copilot in Bing regularly to improve product performance, update existing mitigations, and implement new mitigations in response to our learnings based on real-world usage of the product.

- **Operations and incident response.** We also use Microsoft's ongoing monitoring and operational processes to address when Copilot in Bing

		<ul> <li>receives signals or receives a report indicating possible misuse or violations of the terms of use.</li> <li>Cooperation with Industry Partners. The third-party content that grounds Copilot chat answers relies on the same ranking algorithms and</li> </ul>
		defensive interventions that power traditional Bing search, including reliance on signals of page authority that Bing receives from its third-party partners, such as GDI, NewsGuard (a signatory to the Code), EFE, RSA Journalism Trust Index, and fact-checks using the ClaimReview protocol. For example, Bing uses NewsGuard trustworthiness ratings as one of the indicators of trustworthiness of websites to help users access relevant, reliable, and high- authority content in Copilot. More information about NewsGuard integrations with Bing's new generative experiences is available at " <u>How</u> <u>Microsoft Uses NewsGuard to Support its Trusted, Transparent Technology</u> <u>- NewsGuard.</u> "
		Our approach to identifying, measuring, and mitigating harms will continue to evolve as we learn more, and we continue to make improvements based on feedback from users, civil society groups, and other third party stakeholders.
		Microsoft also recently launched a new web page – <u>Microsoft-2024</u> <u>Elections</u> – where political candidates can report alleged deepfakes of themselves to Microsoft. See also response to QRE 14.1.1.
Measure 14.2	LinkedIn	Bing Search
QRE 14.2.1	LinkedIn's Professional Community Policies prohibit misinformation and misinformation is removed from the LinkedIn platform. Where LinkedIn removes such content pursuant to our false and misleading content policies, LinkedIn notifies members of the action taken. Members that repeatedly post misinformation are permanently restricted.	The <u>Webmaster Guidelines</u> – and related defensive search and spam interventions – are global policies that are enforced globally by Bing Search, including in EU Member States. Websites that appear in Bing search results (in traditional search or in generative AI chat) are not hosted by Bing Search and, as such, Bing Search has limited information about the hosting location of these third-party websites. When addressing spam activity, Bing Search takes action at the global level (which necessarily carry through to Copilot AI

State-sponsored attempts to post misinformation, if any, are removed. Further, LinkedIn's professional focus shapes the type of content we see on platform. People tend to say things differently when their colleagues and employer are watching. Accordingly, our members don't tend to use LinkedIn to engage in the mass dissemination of misinformation, and bad actors generally need to create fake accounts to peddle misinformation.	features reliant on the Bing search index) to benefit Bing Search users in all countries (including EU Member States). Bing Search's defensive search interventions are also applied at a global level (thereby encompassing all EU member states) and automatically applied to queries searched in all EU languages. Metrics on defensive search interventions are provided in SLI 14.2.1. See also responses to QRE 14.1.1-2.
To ensure their content reaches a large audience, bad actors need to either connect with real members or post content that real members will like—both of which are hard to achieve on LinkedIn given our professional focus. The mass dissemination of false information, as well as artificial traffic and engagement, therefore, requires the mass creation of fake accounts, which we have various defences to prevent and limit.	
To evolve to the ever-changing threat landscape, our team continually invests in new technologies for combating inauthentic behaviour on the platform. We are investing in artificial intelligence technologies such as advanced network algorithms that detect communities of fake accounts through similarities in their content and behaviour, computer vision and natural language processing algorithms for detecting AI-generated elements in fake profiles, anomaly detection of risky behaviours, and deep learning models for detecting sequences of activity that are associated with abusive automation. As noted in our most recent global <u>Transparency Report</u> , in the period 1 January to 30 June 2023, LinkedIn blocked or removed approximately 58 million fake accounts. Our automated defenses blocked 90.1% of the fake accounts we stopped during that period, with the remaining 9.9% stopped by our manual investigations and	

restrictions. 99.7% of the fake accounts were stopped proactively, before a member report.

LinkedIn also acts vigilantly to maintain the integrity of all accounts and to ward off bot and false account activity (including "deep fakes"). LinkedIn enforces the policies in its<u>User Agreement</u> prohibiting the use of "bots or other automated methods to access the Services, add or download contacts, send or redirect messages" through:

- Having a dedicated Anti-Abuse team to create the tools to enforce this prohibition
- Using automated systems to detect and block automated activity
- Imposing hard limits on certain categories of activity commonly engaged in by bad actors
- Detecting whether members have installed known prohibited automation software
- Conducting manual investigation and restriction of accounts engaged in automated activity
- Partnering with the broader Microsoft organisation to develop technological solutions for protecting content provenance and identification of deep fakes
- Investing in and using AI to detect coordinated inauthentic activity and communities of fake accounts through similarities in their content and behaviour
- Using third party fact checking sites during the human content review process when suspected deepfakes are flagged or found on the platform
- "Hashing" known instances of deepfake content, which can be used to find copies of the same content on our platform

LinkedIn has reported available metrics at SLI 14.2.1 in respect of the following TTPs:	
<ul> <li>TTP 1: Creation of inauthentic accounts or botnets (which may include automated, partially automated, or nonautomated accounts)</li> <li>TTP 2: Use of fake / inauthentic reactions (e.g. likes, up votes, comments)</li> <li>TTP 3: Use of fake followers or subscribers</li> <li>TTP 4: Creation of inauthentic pages, groups, chat groups, fora, or domains</li> </ul>	
LinkedIn has also reported metrics for SLI 14.2.2 in respect of TTP 1 and TTP 4.	
LinkedIn has focused its efforts on TTPs 1-4 because, as a real- identity professional network, the harm on LinkedIn is generally conducted through fake accounts. Our real members know that the content they post is viewed by their colleagues, managers, and potential business partners, and therefore they generally do not knowingly post misinformation.	
With respect to the remaining TTPs, LinkedIn is unable to reasonably ascertain the intent or provenance of such content. As discussed above, disinformation is not prevalent on LinkedIn due to the professional context of the platform. Distribution of such content through fake accounts is further hampered due to the need to create connections between the fake account and the real member. In the rare instances that such misinformation is spread through fake accounts, due to the adversarial nature of this activity, publicly disclosing details regarding the threat actor's TTPs would hurt our ability to fight against this activity. For example, reporting that	

vulnerable recipients were not targeted may incentivize the targeting of such recipients.	
LinkedIn has and will continue to evaluate what additional metrics it	
could potentially include in future reporting in light of how LinkedIn's	
services function and are used.	

	LinkedIn
	SLI 14.2.1 – SLI 14.2.4
TTP 1	The table below addresses TTP 1: "Creation of inauthentic accounts or botnets (which may include automated, partially automated, or non-automated accounts)." SLI 14.2.1 reports the number of fake accounts that LinkedIn prevented from being created or restricted between 1 July - 31 December 2023, broken out by EEA Member State. The fake accounts reported are attributed to EEA Member States based on the IP address used during registration of the account. 'Number of instances of identified TTPs' and 'Number of actions taken by type' are identical given LinkedIn blocked the registration attempt or restricted the account in all instances.
	SLI 14.2.2. reports two metrics. First, the number of EEA accounts that connected to or followed the fake accounts in SLI 14.2.1 between 1 July – 31 December 2023. For example, the 121,503 fake accounts reported for Austria had a total of 4,465 EEA accounts connect to or follow them between 1 July and 31 December 2023. Whether an account qualifies as an EEA account is based on the IP address used during registration of the account. Second, the number of fake accounts in SLI 14.2.1 that posted a feed post between 1 July – 31 December 2023. For example, of the 230,795 fake accounts prevented or restricted for Ireland, 1,113 posted a feed post between 1 July and 31 December 2023.
	Please note that the metrics provided below are total numbers and do not imply that these fake accounts were engaging in posting misinformation or disinformation.
	We have not reported metrics associated with this TTP where, for example, there is no meaningful metric to report (e.g., metrics for after the TTP in question, given LinkedIn removes detected misinformation and fake accounts from our platform) or LinkedIn does not have a reasonable means to compute the requested metrics.

	SLI 14.2.1		SLI 14.2.2		SLI 14.2.3			SLI 14.2.4		
	Nr of instances of Nr of actions taken		Views/ impressions	Views/ impressions	Penetration	Trends on	Trends on	TTPs	Views/	Interaction/
	identified TTPs	by type	before action	before action	and impact	targeted	narratives	related	impressions	engagement
					on genuine	audiences	used	content	of TTP	with TTP
			The number of EEA		users			in relation		related
				fake accounts that					content (in	content (in
	prevented or	LinkedIn prevented		posted a feed post					relation to	relation to
				between 1 July –				on the	overall	overall
	1 July – 31 December			31 December 2023.				service	views/	interaction/
	2023		1 July – 31 December 2023.						impressions on the	engagement
		December 2023	December 2023.						service)	on the service)
									SEIVICE)	the service)
Member States										
Austria	121,503	121,503	4,465	1,673						
Belgium	140,774	140,774	7,561	1,897						
Bulgaria	101,381	101,381	2,997	910						
Croatia	56,081	56,081	2,418	385						
Cyprus	20,812	20,812	1,621	474						
Czech Republic	175,859	175,859	4,470	1,340						
Denmark	71,989	71,989	4,376	3,258						
Estonia	36,847	36,847	876	309						
Finland	75,976	75,976	2,903	988						
France	1,200,626	1,200,626	100,058	22,846						
Germany	1,273,454	1,273,454	104,281	37,606						
Greece	203,827	203,827	5,848	1,191						
Hungary	80,631	80,631	2,517	459						
Ireland	230,795	230,795	4,536	1,113						
Italy	687,784	687,784	45,952	14,888						
Latvia	69,849	69,849	1,101	235						

Total EEA	7,465,115	7,465,115	418,443	132,815			
Total EU	7,314,583	7,314,583	415,892	132,180			
Norway	130,978	130,978	2,207	550			
Liechtenstein	948	948	36	9			
Iceland	18,606	18,606	308	76			
Sweden	137,928	137,928	5,623	1,590			
Spain	831,887	831,887	46,801	17,266			
Slovenia	26,549	26,549	607	155			
Slovakia	36,260	36,260	1,388	256			
Romania	420,112	420,112	5,777	1,827			
Portugal	171,926	171,926	10,198	2,112			
Poland	553,732	553,732	23,808	8,593			
Netherlands	454,150	454,150	22,392	9,805			
Malta	12,916	12,916	470	109			
Luxembourg	17,533	17,533	699	203			
Lithuania	103,402	103,402	2,149	692			

The table below addresses TTP 2: "Use of fake / inauthentic reactions (e.g. likes, up votes, comments)." The table reports the number of fake accounts reported in TTP 1 SLI 14.2.1 that reacted to, commented on, or shared (collectively, "engaged with") a feed post between 1 July – 31 December 2023.

TTP 2

The numbers of fake accounts reported below are a subset of the fake accounts reported in TTP 1 SLI 14.2.1 that engaged with a feed post between 1 July – 31 December 2023. For example, of the 121,503 fake accounts that LinkedIn prevented from being created or restricted between 1 July – 31 December 2023 in Austria (as reported in TTP 1 SLI 14.2.1), 2,544 of those accounts engaged with a feed post between 1 July – 31 December 2023.

Please note that the metrics provided below are total numbers and do not imply that these fake accounts were engaging in posting misinformation or disinformation.

	We have not reported metrics associated with this TTP where, for example, there is no meaningful metric to report (e.g., metrics for after the TTP in questio given LinkedIn removes detected misinformation and fake accounts from our platform) or LinkedIn does not have a reasonable means to compute the requeste metrics.											
	SLI 14.2.1		SLI 14.2.2				SLI 14.2.3			SLI 14.2.4		
	TTPs The number of fake accounts reported in TTP 1 SLI 14.1.1 that engaged with a feed post between 1 July – 31 December	actions taken by type	impressions before action	/ engagement	impressions	Interaction/ engagement after action	Penetration and impact on genuine users		Trends on narratives used	related content in relation to overall	Views/ impressions of TTP related content (in relation to overall views/impressions on the service)	
Member States	2023											
Austria	2,544											
Belgium	2,921											
Bulgaria	1,125											
Croatia	599											
Cyprus	613											
Czech Republic	1,664											
Denmark	1,872											
Estonia	322											
Finland	1,256											
France	29,447											
Germany	42,528											
Greece	1,723											

Hungary	893						
Ireland	1,664						
Italy	18,823						
Latvia	289						
Lithuania	762						
Luxembourg	394						
Malta	142						
Netherlands	11,370						
Poland	8,631						
Portugal	3,309						
Romania	1,762						
Slovakia	428						
Slovenia	148						
Spain	22,743						
Sweden	2,375						
Iceland	163						
Liechtenstein	13						
Norway	1,018						
Total EU	160,347						
Total EEA	161,541						

TTP 3The table below addresses TTP 3: "Use of fake followers or subscribers." The table reports the number of fake accounts reported in TTP 1 SLI 14.2.1 that followed<br/>a LinkedIn profile or page between 1 July – 31 December 2023.The numbers of fake accounts reported below are a subset of the fake accounts reported in TTP 1 SLI 14.2.1 that followed a LinkedIn profile or page between<br/>1 July – 31 December 2023. For example, of the 121,503 fake accounts that LinkedIn profile or page between 1 July – 31 December 2023 in Austria (as reported in TTP 1 SLI 14.2.1), 13,759 of those accounts followed a LinkedIn profile or page between 1 July – 31 December 2023 (as reported below).

	Please note disinformati		metrics provi	ided below a	re total numb	pers and do	not imply tl	hat these fal	ke accounts	were engagi	ng in posting misi	nformation o
	We have no	t reported	metrics asso	ciated with th	is TTP where,	for example,	there is no r	meaningful n	netric to repo	ort (e.g., metr	ics for after the TTF	o in question,
	given Linked	dln remove	es detected m	nisinformatior	and fake acc	ounts from o	ur platform)	or LinkedIn	does not hav	ve a reasonat	ble means to compu	ute the
	requested n	netrics.										
	SLI 14.2.1		SLI 14.2.2				SLI 14.2.3			SLI 14.2.4		
	Nr of	Nr of	Views/	Interaction	Views/	Interaction	Penetration	Trends on	Trends on	TTPs related	Views/	Interaction
	instances of	actions	impression	/	impressions	/	and impact	targeted	narratives	content in	impressions of	/
	identified	taken by	s before	engagement	after action	engagement	on	audiences	used	relation to	TTP related	engagement
	TTPs	type	action	before		after	genuine			overall	content (in	with TTP
				action		action	users			content on	relation to overall	related
	The number									the	views/impressions	content (in
	of fake									service	on the service)	relation to
	accounts											overall
	reported in											interaction
	TTP 1 SLI											/engagement
	14.1.1 that											on the
	followed a											service)
	LinkedIn											
	profile or											
	page											
	between 1											
	July – 31											
	December											
	2023											
Member States	10 == 6											
Austria	13,759											
Belgium	17,233											
Bulgaria	10,978											
Croatia	4,449											
Cyprus	3,692											

Czech Republic	9,814						
Denmark	10,712						
Estonia	5,176						
Finland	9,115						
France	161,699						
Germany	204,986						
Greece	10,260						
Hungary	5,932						
Ireland	9,689						
Italy	116,580						
Latvia	2,288						
Lithuania	7,788						
Luxembourg	2,230						
Malta	877						
Netherlands	71,597						
Poland	89,843						
Portugal	22,644						
Romania	21,779						
Slovakia	3,034						
Slovenia	1,634						
Spain	115,210						
Sweden	17,147						
Iceland	735						
Liechtenstein	75						
Norway	5,168						
Total EU	950,145						
Total EEA	956,123						

TP 4	The table bel	ow addr	esses TTP 4: "(	Creation of in	authentic pag	ges, groups, cl	hat groups, I,	or domains.	' SLI 14.2.1 re	eports the nur	nber of LinkedIn pa	ages or gro	Sup
	that the fake	account	s reported in	TTP 1 SLI 14.	2.1 created b	etween 1 July	– 31 Decemi	ber 2023.					
	The numbers	s of Linke	edIn pages or	groups creat	ed reported b	pelow are base	ed on the po	pulation of f	ake accounts	s reported in	TTP 1 SLI 14.2.1. Fo	r example	, th
	121,503 fake	account	ts that Linked	In prevented	from being o	created or res	tricted betwe	een 1 Januar	y – 30 Decer	mber 2023 in	Austria (as reporte	ed in TTP	1 SL
	14.2.1) create	ed 49 Lin	kedIn pages o	or groups be	tween 1 July -	- 31 Decembe	er 2023 (as re	ported belo	w).				
	SLI 14.2.2 rep	oorts the	number of a	ccounts in th	e EEA that joi	ned or follow	ed the pages	or groups r	eported in T	TP 4 SLI 14.2.	1 between 1 July –	31 Decem	nbe
	2023. For exa	imple, th	e 49 pages an	nd groups rep	orted for Aus	tria in TTP4 SL	l 14.2.1 had a	a total of 114	EEA account	ts join or follo	w between 1 July –	31 Decem	nbe
	2023. Wheth	er an aco	count qualifie	s as an EEA a	ccount is base	ed on the IP a	ddress used	during regis	tration of the	e account.			
	Please note disinformatio		metrics prov	vided below a	are total num	nbers and do	not imply th	hat these fa	ke accounts	were engagi	ng in posting misi	nformatio	n c
	We have not	reporte	d metrics asso	ociated with t	this TTP where	e, for example	e, there is no	meaningful	metric to rep	oort (e.g., met	rics for after the T	FP in ques	tior
	given Linked requested m	lln remo etrics.	ves detected			accounts fro	m our platfo	-		ot have a rea	rics for after the T asonable means to	•	
	given Linked requested m <b>SLI 14.2.1</b>	lln remo etrics.	ves detected	misinformat	ion and fake	accounts fro	m our platfo SLI 14.2.3	orm) or Linke	edIn does no	ot have a rea	asonable means to	compute	e th
	given Linked requested m <b>SLI 14.2.1</b> Nr of	IIn remo etrics. Nr of	ves detected <b>SLI 14.2.2</b> Views/	misinformat	ion and fake Views/	accounts fro	m our platfo <b>SLI 14.2.3</b> Penetration	orm) or Linke	edIn does no	ot have a rea <b>SLI 14.2.4</b> TTPs related	asonable means to Views/	•	e th
	given Linked requested m <b>SLI 14.2.1</b> Nr of instances of	In remo etrics. Nr of actions	ves detected <b>SLI 14.2.2</b> Views/ impressions	misinformat Interaction/ engagement	ion and fake Views/ timpressions	accounts fro Interaction/ engagement	m our platfo <b>SLI 14.2.3</b> Penetration and impact	Trends on targeted	edIn does no Trends on narratives	ot have a rea <b>SLI 14.2.4</b> TTPs related content in	asonable means to Views/ impressions of	Interactic	e th
	given Linked requested m <b>SLI 14.2.1</b> Nr of instances of identified	In remo etrics. Nr of actions taken	ves detected <b>SLI 14.2.2</b> Views/ impressions before	misinformat	ion and fake Views/ timpressions after	accounts fro	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no	SLI 14.2.4 TTPs related content in relation to	views/ impressions of TTP related	Interactic / engagem	on ient
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs	In remo etrics. Nr of actions taken by	ves detected <b>SLI 14.2.2</b> Views/ impressions	misinformat Interaction/ engagement	ion and fake Views/ timpressions	accounts fro Interaction/ engagement	m our platfo <b>SLI 14.2.3</b> Penetration and impact	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall	views/ impressions of TTP related content (in	Interactic / engagem with	e th
	given Linked requested m <b>SLI 14.2.1</b> Nr of instances of identified TTPs	In remo etrics. Nr of actions taken by type	ves detected <b>SLI 14.2.2</b> Views/ impressions before action	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on	Views/ impressions of TTP related content (in relation to overall	Interactic / engagem with related	e th on nent TT
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs The number	In remo etrics. Nr of actions taken by type	<b>SLI 14.2.2</b> Views/ impressions before action The number	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on the	Views/ impressions of TTP related content (in relation to overall views/impressions	Interactic / engagem with related content	e th on TT (i
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs The number of LinkedIn	In remo etrics. Nr of actions taken by type	sLI 14.2.2 Views/ impressions before action The number of accounts	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on	Views/ impressions of TTP related content (in relation to overall	Interactic / engagem with related content relation	e th on TT
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs The number of LinkedIn pages or	In remo etrics. Nr of actions taken by type	sLI 14.2.2 Views/ impressions before action The number of accounts in the EEA	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on the	Views/ impressions of TTP related content (in relation to overall views/impressions	Interactic / engagem with related content relation overall	e th on TT ( 1
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs The number of LinkedIn pages or groups	In remo etrics. Nr of actions taken by type	<b>SLI 14.2.2</b> Views/ impressions before action The number of accounts in the EEA that joined	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on the	Views/ impressions of TTP related content (in relation to overall views/impressions	Interactic / engagem with related content relation overall interactic	e th on Ien TT ( 1
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs The number of LinkedIn pages or groups created	In remo etrics. Nr of actions taken by type	SLI 14.2.2 Views/ impressions before action The number of accounts in the EEA that joined or followed	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on the	Views/ impressions of TTP related content (in relation to overall views/impressions	Interactic / engagem with related content relation overall interactic /engager	e th on Ien TT ( 1
	given Linked requested m SLI 14.2.1 Nr of instances of identified TTPs The number of LinkedIn pages or groups	In remo etrics. Nr of actions taken by type	<b>SLI 14.2.2</b> Views/ impressions before action The number of accounts in the EEA that joined	misinformat	ion and fake Views/ timpressions after	accounts fro Interaction/ engagement	m our platfo SLI 14.2.3 Penetration and impact on genuine	Trends on targeted	edIn does no Trends on narratives	SLI 14.2.4 TTPs related content in relation to overall content on the	Views/ impressions of TTP related content (in relation to overall views/impressions	Interactic / engagem with related content relation overall interactic	e th on Ien TT ( 1

	December	and groups				
	2023 by the	reported in				
	fake	TTP 4 SLI				
	accounts	14.2.1				
	reported in	between 1				
	TTP 1 SLI	July – 31				
	14.1.1.	December				
		2023	 			
Member States						
Austria	49	114				
Belgium	49	225				
Bulgaria	42	76				
Croatia	24	181				
Cyprus	28	192				
Czech Republic	40	242				
Denmark	13	41				
Estonia	20	10				
Finland	20	104				
France	306	1,983				
Germany	420	2,041				
Greece	27	426				
Hungary	9	50				
Ireland	27	55				
Italy	318	1,220				
Latvia	13	34				
Lithuania	15	57				
Luxembourg	4	22				
Malta	5	63				
Netherlands	223	507				
Poland	339	412				

Total EEA	2,965	10,731					
Total EU	2,955	10,687					
Norway	10	37					
Liechtenstein	0	0					
Iceland	0	7					
Sweden	43	243					
Spain	209	1,548					
Slovenia	4	23					
Slovakia	10	28					
Romania	625	432					
Portugal	73	358					

**TTP 5** LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1 is Theoretically fit for purpose, SLI 14.2.2 is not fit for purpose, SLI 14.2.3 is Optional/ Alternative and SLI 14.2.4 is Partially fit for purpose. However, LinkedIn has not reported metrics associated with this TTP 5 (Account hijacking or impersonation) for the reasons explained in our response to QRE 14.2.1.

**TTP 6** LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1 is Theoretically fit for purpose, SLI 14.2.2 is not fit for purpose, SLI 14.2.3 is Optional/ Alternative and SLI 14.2.4 is Partially fit for purpose. However, LinkedIn has not reported metrics associated with this TTP 6 (Deliberately targeting vulnerable recipients (e.g. via personalized advertising, location spoofing or obfuscation)) for the reasons explained in our response to QRE 14.2.1.

**TTP 7** LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1, SLI 14.2.2 and SLI 14.2.4 are Theoretically fit for purpose and SLI 14.2.3 is Optional/ Alternative. However, LinkedIn has not reported metrics associated with this TTP 7 (Deploy deceptive manipulated media ("deep fakes", "cheap fakes"...)) for the reasons explained in our response to QRE 14.2.1.

**TTP 8** LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1 is Theoretically fit for purpose, SLI 14.2.2 is Not fit for purpose, SLI 14.2.3 is Optional/ Alternative and SLI 14.2.4 is Partially fit for purpose. However, LinkedIn has not reported metrics associated with this TTP 8 (Use "hack and leak" operation (which may or may not include doctored content)) for the reasons explained in our response to QRE 14.2.1

- **TTP 9** LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1 is Theoretically fit for purpose, SLI 14.2.2 is not fit for purpose, SLI 14.2.3 is Optional/ Alternative and SLI 14.2.4 is Partially fit for purpose. However, LinkedIn has not reported metrics associated with this TTP 9 (Inauthentic coordination of content creation or amplification, including attempts to deceive/manipulate platforms algorithms (e.g. keyword stuffing or inauthentic posting/reposting designed to mislead people about popularity of content, including by influencers)) for the reasons explained in our response to QRE 14.2.1.
- TTP 10 LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1 is Theoretically fit for purpose, SLI 14.2.2 is not fit for purpose, SLI 14.2.3 is Optional/ Alternative and SLI 14.2.4 is Partially fit for purpose. However, LinkedIn has not reported metrics associated with this TTP 10 (Use of deceptive practices to deceive/manipulate platforms algorithms, such as to create, amplify or hijack hashtags, data voids, filter bubbles, or echo chambers) for the reasons explained in our response to QRE 14.2.1.
- **TTP 11** LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1, SLI 14.2.2 and SLI 14.2.4 are Theoretically fit for purpose, and SLI 14.2.3 is Optional/ Alternative. However, LinkedIn has not reported metrics associated with this TTP 11 (Non-transparent compensated messages or promotions by influencers) for the reasons explained in our response to QRE 14.2.1. LinkedIn also notes that where members share content in an exchange for value (including monetary payment, endorsements, free products or services, or other benefit), they must label the post as a brand partnership and comply with our Advertising Policies and applicable local laws. This Help Center <u>article</u> provides details on how to label a post as a brand partnership.
- TTP 12 LinkedIn notes that per the Integrity of Services Sub-Group, SLI 14.2.1 is Theoretically fit for purpose, SLI 14.2.2 is not fit for purpose, SLI 14.2.3 is Optional/ Alternative and SLI 14.2.4 is Partially fit for purpose. However, LinkedIn has not reported metrics associated with this TTP 12 (Coordinated mass reporting of non-violative opposing content or accounts) for the reasons explained in our response to QRE 14.2.1.

	Bing Search
	SLI 14.2.1 – SLI 14.2.4
ТТР 10	This Section addresses TTP No. 10 ("Use of deceptive practices to deceive/manipulate platform algorithms, such as to create, amplify or hijack hashtags, data voids, filter bubbles, or echo chambers"), which is the TTP primarily applicable to Bing Search (including Copilot in Bing).
	SLI 14.1.2: In response to this SLI, Bing Search is providing data on defensive search interventions employed to counteract threats and TTPs on the Bing platform. This response includes the following data categories:
	<ul> <li>"New DSI" reflects the net new number of queries treated with defensive search interventions during the Reporting Period (across all of Bing) since the preceding reporting period. Although Member State reporting is requested, because Bing Search takes defensive search actions globally (rather than on a per country basis), each defensive search action would necessarily be implemented in an EU member state, and it is not feasible to provide Member State reporting for globally actioned measures. See SLI 14.2.2 for more detailed Member State reporting.</li> <li>"AutoSuggest DSI" reflects the number of search query suggestions that were suppressed for queries entered by users in the EEA (including traditional web search and Copilot in Bing) during the Reporting Period.</li> <li>"Chat DSI" reflects the total number of user prompts Bing applied defensive search interventions to within Copilot in Bing during the Reporting Period (rounded up). In simple terms, this reflects the number of user queries in chat where responses (i.e., chat output) to the queries were treated with defensive search interventions. Although Member State reporting is requested, because Bing Search takes defensive search actions globally (rather than on a per country basis), each defensive search action would necessarily be implemented in an EU member state, and it is not feasible to provide Member State reporting for globally actioned measures.</li> </ul>
	<ul> <li>SLI 14.2.2: Bing cannot provide data on interaction or engagement, as Bing does not allow users to "like" or "share" content within Bing and this SLI metric appears oriented to social or sharing platforms. Bing also cannot provide "before and after" data due to the preventative nature of search interventions and query-driven nature of web search. Nonetheless, below Bing Search has provided user impressions for queries that were treated with "defensive search" interventions across all of Bing Search (inclusive of traditional web search and Copilot in Bing) during the Reporting Period.</li> <li>"Unique Queries DSI" reflects the total number of unique queries searched by users in the EEA that were treated with defensive search interventions during the Reporting Period.</li> <li>"DSI Query Impressions" reflects the number of impressions for unique queries treated with defensive search interventions that appeared to users in the EEA during the Reporting Period.</li> </ul>

	SLI 14.2.3 – This SLI is n	not applicable to search eng	gines, as Bing Search is	not an online platform that	allows for user hosted content o	r public sharing of user
	generated content with	other users. User account	s in the manner contem	plated under this provision	are not available in search (i.e. r	egistered user accounts
	are not capable of amp	lifying creating or amplifyin	ng content as one may t	hrough a social media netw	vork).	
	For SLI 14.2.4 – This SL	l is also not applicable to s	earch engines for the at	ove reasons.		
	SLI 14.2.1			SLI 14.2.2		
	Nr of actions taken by	Nr of actions taken by	Nr of actions taken			Interaction/
	type	type	by type			engagement before
	New DSI	Autosuggest DSI	Chat DSI	Unique Queries DSI	DSI Query Impressions	action <b>N/A</b>
Member						
States						
Austria		1,177,710		8,662	120,015	
Belgium		2,011,531		11,493	172,367	
Bulgaria		2,239		28	31	
Croatia		1,289		11	11	
Cyprus		424		4	6	
Czech Republic		6,887		39	73	
Denmark		611,729		5,626	59,967	
Estonia		698		13	13	
Finland		460,243		3934	43,144	
France		8,319,549		29,977	1,339,047	
Germany		13,165,986		42,656	1,656,488	
Greece		2,755		19	30	
Hungary		4,225		9	16	
Ireland		481,893		15,056	147,099	
Italy		4,280,339		16,673	481,823	
Latvia		798		16	23	
Lithuania		898		24	31	
Luxembourg		350		4	4	
Malta		246		12	16	

Total Global	2.9 million	-	24.78 million	-	-	
Total EEA		43,325,642		215,893	6,141,815	
Total EU		41,889,378		206,911	6,033,414	
Norway		1,436,004		8,978	108,396	
Liechtenstein		6		0	0	
Iceland		254		4	5	
Sweden		1,404,642		13,143	170,350	
Spain		3,590,057		17,967	1,030,091	
Slovenia		680		8	9	
Slovakia		1,957		14	27	
Romania		4,283		40	68	
Portugal		781,191		8,900	132,968	
Poland		2,386,145		13,484	392,135	
Netherlands		3,190,634		19,099	287,562	

Measure 14.3	LinkedIn	Bing Search
QRE 14.3.1	the second half of 2022 (and reported on in Microsoft's previous reports) as being fit for purpose for the current reporting cycle. LinkedIn reiterates the need for flexibility amongst different types of services to address TTPs that are	
	most relevant to their platforms. This list can be consulted below:	However, as noted in QRE 14.1.1 and 14.1.2, many of these TTPs are inapplicable to or irrelevant to search engines. Bing reiterates the need for flexibility amongst different types of services signatories to address TTPs that are most relevant to their platforms.
	<ul> <li>The following TTPs pertain to the creation of assets for the purpose of a disinformation campaign, and to ways to make these assets seem credible: <ul> <li>1. Creation of inauthentic accounts or botnets (which may include automated, partially automated, or non-automated accounts)</li> <li>2. Use of fake / inauthentic reactions (e.g. likes, up votes, comments)</li> <li>3. Use of fake followers or subscribers</li> </ul> </li> </ul>	The following TTPs pertain to the creation of assets for the purpose of a

<ul> <li>4. Creation of inauthentic pages, groups, chat groups, fora, or domains</li> <li>5. Account hijacking or impersonation</li> <li>The following TTPs pertain to the dissemination of content created in the context of a disinformation campaign, which may or may not include some forms of targeting or attempting to silence opposing views. Relevant TTPs include:</li> <li>6. Deliberately targeting vulnerable recipients (e.g. via personalized advertising, location spoofing or obfuscation)</li> <li>7. Deploy deceptive manipulated media (e.g. "deep fakes", "cheap fakes")</li> <li>8. Use "hack and leak" operation (which may or may not include doctored content)</li> <li>9. Inauthentic coordination of content creation or amplification, including attempts to deceive/manipulate platforms algorithms (e.g. keyword stuffing or inauthentic posting/reposting designed to mislead people about popularity of content, including by influencers)</li> <li>10. Use of deceptive practices to deceive/manipulate platform algorithms, such as to create, amplify or hijack hashtags, data voids, filter bubbles, or echo chambers</li> <li>11. Non-transparent compensated messages or promotions by influences</li> <li>12. Coordinated mass reporting of non-violative opposing content or accounts</li> </ul>	<ul> <li>4. Creation of inauthentic pages, groups, chat groups, fora, or domains</li> <li>5. Account hijacking or impersonation</li> </ul> The following TTPs pertain to the dissemination of content created in the context of a disinformation campaign, which may or may not include some forms of targeting or attempting to silence opposing views. Relevant TTPs include: <ul> <li>6. Deliberately targeting vulnerable recipients (e.g. via personalized advertising, location spoofing or obfuscation)</li> <li>7. Deploy deceptive manipulated media (e.g. "deep fakes", "cheap fakes")</li> <li>8. Use "hack and leak" operation (which may or may not include doctored content)</li> <li>9. Inauthentic coordination of content creation or amplification, including attempts to deceive/manipulate platforms algorithms (e.g. keyword stuffing or inauthentic posting/reposting designed to mislead people about popularity of content, including by influencers)</li> <li>10. Use of deceptive practices to deceive/manipulate platform algorithms, such as to create, amplify or hijack hashtags, data voids, filter bubbles, or echo chambers</li> <li>11. Non-transparent compensated messages or promotions by influencers</li> <li>12. Coordinated mass reporting of non-violative opposing content or accounts</li> </ul>
Further, as noted above, the relevant Taskforce Subgroup has considered whether the SLIs for each of these TTPs are fit for purpose and classified each SLI as either Theoretically fit for purpose, Not fit for purpose, Partially fit for purpose or Optional /Alternative.	

IV. Integrity of Services								
	Commitment 15							
Ŭ	velop or operate AI systems and that disseminate AI-generated and manipulated con onsideration the transparency obligations and the list of manipulative practices prohil Intelligence Act.	U						
	C.15 M 15.1 M 15.2							
We signed up to the following LinkedIn LinkedIn LinkedIn								
measures of this commitment:	Bing Search	Bing Search	Bing Search					

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	In February 2024, Microsoft signed
		the Tech Accord to Combat
		Deceptive Use of AI in 2024
		Elections, which commits to a range
		of measures intended to limit the
		deceptive use of AI in connection
		with important 2024 elections.
		Bing has also improved content
		provenance measures for Copilot in
		Bing. The Copilot in Bing feature to
		create images includes C2PA
		Content Credentials, developed by
		the Coalition for Content
		Provenance and Authenticity
		(C2PA), which sets technical

		standards for certifying the source and history (or provenance) of media content.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures LinkedIn will continue to assess its policies and services and to update them as warranted.	Planned Implementation MeasuresMicrosoft's Content Credentials as a Service will be launched in private preview for political campaigns and elections purposes. This tool enables campaigns to add Content Credentials to their digital media. It will be available to select private preview customers at no cost for the 2024 election cycle and will be the first-of-a-kind SaaS application for adding content credentials to images and video.Bing also consistently reviews and evaluates its policies and practices related to existing and new Bing features and adjusts as needed. Bing expects to continue and 

a number of products and Microsoft takes its commitment to responsible AI seriously and has a robust members to disseminate, Responsible AI program. In addition to the safeguards noted earlier in this
safety risks posed by such plicies and procedures to w features are consistent pplicable law. Kesponsible Al program. In addition to the safeguards noted earlier in this report and discussed thoroughly at <u>Copilot in Bing: Our approach to</u> <u>Responsible Al</u> , and Microsoft has implemented a number of measures and policies to help counter attempts to manipulate Al systems that generate content.
Copilot was developed in accordance with Microsoft's AI Principles, Microsoft's Responsible AI Standard, and in partnership with responsible AI experts across the company, including Microsoft's Office of Responsible AI experts across the company, including Microsoft's Office of Responsible AI experts across the company, including Microsoft's Office of Responsible AI experts across the company, including Microsoft's Office of Responsible AI experts across the company, including Copilot in Bing and other Bing search features, must adhere to Microsoft's <u>Responsible AI Standard</u> and undertake impact assessments to help ensure responsible use of AI- influenced algorithms and processes for any new product features. More details on Microsoft's Responsible AI Standard, impact assessments, and resources on Responsible AI are located at Microsoft's <u>Responsible AI Hub</u> . Bing also conducts detailed annual risk assessments that evaluate risks posed by its systems (including generative AI features) and evaluates current and potential risk mitigation measures. In addition to the measures discussed at QREs 14.1.1.2 and 14.1.2 (including pre and post launch testing, the use of classifiers and metaprompting, defensive search interventions, reporting functionality, and increased operations and incident response), Microsoft has incorporated the following safeguards and policies for countering prohibited manipulative practices for AI systems. To help facilitate safe use of Copilot in Bing, Microsoft published updated terms of use (including a user Code of Conduct) and implemented other
isting po such ne es and ap edIn has security ugmente of genera conside il data us are fi t on in ontrol of ot addec ing and e ng proce ves, that h genera s and rat use. Limi ise. We a ed to the s we asse how to d contain in mode rofession

We also engage in reactive content moderation, through provision of member tools to report policy-violating issues with the content. Additional features have been added to these tools that address generative AI-specific issues such as 'hallucinations.'

- 3. Fairness and Inclusion LinkedIn has a cross functional team that designs policy and process to proactively mitigate the risk that AI tools, including generative AI tools, perpetuate societal biases or facilitate discrimination. To promote fairness and inclusion, we target two key areas - content subject and communities. With respect to content subjects, prompts are engineered to reduce the risk of biased content, blocklists are leveraged to replace harmful terms with neutral terms, and member feedback is monitored to learn and improve. With respect to communities, in addition to a focus on problematic content like stereotypes, we are working to expand the member communities that are served by our generative AI tools. Additionally, LinkedIn continues to invest in methodologies and techniques to more broadly ensure algorithmic fairness.
- 4. Transparency LinkedIn is committed to being transparent with members. With respect to generative AI products and features, our goal is to educate members about the technology and our use of it such that they can make their own decisions about how to engage with it. For example, with Collaborative Articles we identify the use of AI in the relevant UI and we provide additional detail in a linked <u>Help Center article.</u>
- 5. Accountability In addition to the privacy, security, and safety processes discussed above, for AI tools we have additional assessments of training data and model cards so we can more appropriately assess risks and develop mitigations for the AI models that support our AI products and initiatives.

Supplemental Terms prohibit users from "engag[ing] in activity that is fraudulent, false, or misleading" and "attempt[ing] to create or share content that could mislead or deceive others, including for example creation of disinformation, content enabling fraud, or deceptive impersonation." Users that violate these terms may be suspended from the service. In addition, Copilot may block certain text prompts that violate or are likely to violate the Code of Conduct. Repeated attempts to produce prohibited content or other violations of the Code of Conduct may result in service or account suspension. In addition, Microsoft maintains social listening pipelines where insights and user feedback (including efforts to "jailbreak" Copilot systems) on Copilot are collected from the open Internet. These insights and user feedback are manually reviewed by humans, analyzed daily, and shared across the Copilot and Bing product teams and with product leadership to identify new areas of concern and implement additional mitigations as needed. Microsoft also has set up a robust user reporting and appeal process to review and respond to user concerns of harmful or misleading content.

Copilot in Bing also provides several touchpoints for meaningful AI disclosures, where users are notified that they are interacting with an AI system and are presented with opportunities to learn more about these features and generative AI, such as through in-product disclaimers, <u>Copilot in Bing: Our Approach to Responsible AI</u>, educational FAQs, and blog posts. Empowering users with this knowledge can help them avoid over-relying on AI and learn about the system's strengths and limitations. For example, the Copilot home page displays the below disclosure to users:

Copilot uses AI. Check for mistakes. Terms Privacy FAQs Learn more

In addition to the measures discussed above, Microsoft has worked to deliver an experience that encourages responsible use of Copilot in Bing

		<ul> <li>and to limit the generation of harmful or unsafe images. When these systems detect that a potentially harmful image could be generated by a prompt, it blocks the prompt and warns the user.</li> <li>Microsoft's Responsible AI systems will continue to improve, and Microsoft regularly incorporates user and third-party feedback reported via Bing and Copilot Feedback buttons and its user reporting tools.</li> <li>Since its last report, Microsoft has undertaken a number of AI-related initiatives to advance safe, secure, and trustworthy AI, including: <ul> <li>Our commitments to advance safe, secure, and trustworthy AI - Microsoft On the Issues</li> <li>Progress with our AI commitments: an update ahead of the UK AI Safety Summit - Microsoft On the Issues</li> <li>Combating abusive AI-generated content: a comprehensive approach - Microsoft On the Issues</li> <li>Meeting the moment: combating AI deepfakes in elections through today's new tech accord - Microsoft On the Issues</li> </ul> </li> </ul>
Measure 15.2	LinkedIn	Bing Search
QRE 15.2.1	<ul> <li>With respect to the algorithms used for detection, moderation, and sanctioning of impermissible conduct and content, please see:</li> <li>QRE 15.1.1 (policies for countering prohibited manipulative practices in Al systems);</li> <li>QRE 18.1.3 (design of recommender systems and related Al);</li> <li>QRE 18.2.1 (policies and procedures to limit spread of harmful false or misleading information);</li> <li>QRE 22.2.1 (actions taken to assist members in identifying trustworthy content); and</li> <li>QRE 23.2.1 (actions taken to ensure integrity of reporting and appeals process).</li> </ul>	As a search engine, Bing Search operates very differently from social media websites and other online platforms that host content. Bing Search does not host user-generated content and does not use algorithms to detect, moderate or sanction user-provided content except for limited circumstances outside the scope of this Code of Practice (e.g., the use of PhotoDNA software to detect and report child sexually explicit imagery uploaded to visual search). As to third party websites indexed by Bing Search, Bing Search does not use algorithms to detect, monitor or sanction such websites, except for limited circumstances outside the scope of this Code of Practice (e.g., the use of PhotoDNA software to detect and report child sexually explicit imagery). Bing users have many legitimate reasons for seeking out content in search that may be harmful or offensive in other

contexts, and so Bing Search works to provide as comprehensive and useful of a collection of results as possible and does not proactively intervene to limit access to legal content. In some limited cases Bing Search may take action to remove or limit access to third party links where quality, safety, user demand, relevant laws, and/or public policy concerns exist, but these interventions are reactive; Bing generally does not engage in proactive algorithmic interventions to remove content. Copilot in Bing features include additional enhanced safety features such as classifiers, filters, and a bespoke metaprompt that further limits the likelihood of harmful content appearing in generative AI features. Microsoft has engaged in extensive Responsible AI reviews regarding generative AI features in order to ensure outputs are not biased or discriminatory. It has also implemented additional filtering and classifiers to prevent Copilot in Bing chat responses from returning what Bing considers "low authority" content as part of an answer and to help address impermissible content and behaviors. Microsoft is also 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 measures and monitors conversation metrics to improve the interventions to balance the harm prevention and provide users with useful information. Lastly, Microsoft has endeavored to provide transparency about how it designed and tested its generative AI features with responsible AI in mind via blog posts, FAQs, presentations, and overview documentation such as Copilot in Bing: Our Approach To Responsible AI.

## IV. Integrity of Services

## Commitment 16

Relevant Signatories commit to operate channels of exchange between their relevant teams in order to proactively share information about cross-platform influence operations, foreign interference in information space and relevant incidents that emerge on their respective services, with the aim of preventing dissemination and resurgence on other services, in full compliance with privacy legislation and with due consideration for security and human rights risks.

	C.16	M 16.1	M 16.2
We signed up to the following	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search	

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	As part of our work to fulfil our	As part of our work to fulfil our
	commitments under the Tech	commitments under the Tech
	Accord, Microsoft has established	Accord, Microsoft has established
	policies for reporting and sharing of	policies for reporting and sharing of
	deceptive AI in elections. This	deceptive AI in elections. This
	includes reporting, detection and	includes reporting, detection and
	coordination across products to	coordination across products to
	ensure action is taken on any	ensure action is taken on any
	violative content.	violative content.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6	Planned Implementation Measures	Planned Implementation Measures
months?		1
	We look forward to continuing to	We look forward to continuing to
	work on this commitment with the	work on this commitment with the

other signatories as we develop	other signatories as we develop
further cross platform information	further cross platform information
sharing.	sharing.

Measure 16.1	LinkedIn	Bing Search
Measure 16.1 QRE 16.1.1	LinkedIn, through Microsoft, is an active participant in and contributor to the Task-force's Crisis Response subgroup, in which it proactively provides analysis and data, related to influence operations, foreign interference in information space and relevant incidents that emerges on its service. Microsoft's internal threat detection and research teams, including Microsoft Threat Analysis Center (MTAC), Microsoft Threat Intelligence Center (MSTIC), Microsoft Research (MSR), and AI For Good, collect and analyse data on actors of disinformation, misinformation and information manipulation across platforms. Moreover, LinkedIn works with numerous partners to facilitate the flow of information to tackle purveyors of disinformation, including disinformation spread by state-sponsored and institutional actors. LinkedIn maintains an internal Trust and Safety team composed of threat investigators and intelligence personnel to address	Bing Search, through Microsoft, is an active participant in and contributor to the Task-force's Crisis Response subgroup, in which it proactively provides analysis and data related to influence operations, foreign interference in information space and relevant incidents that emerges on its service. Microsoft's internal threat detection and research teams, including Microsoft Threat Analysis Center (MTAC), Microsoft Threat Intelligence Center (MSTIC), Microsoft Research (MSR), and AI For Good, collect and analyse data on actors of disinformation, misinformation and information manipulation across platforms. These teams work with external organisations and companies to share and ingest data that help support Microsoft product and service teams effectively respond to issues and threats. Microsoft also 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),
	disinformation spread by state-sponsored and institutional actors. LinkedIn maintains an internal Trust and Safety team composed of	operations targeting democracies across the world and works with trusted

SLI 16.1.1 – Numbers of actions as a result of information sharing	in threat exchanges. T as: 1) regular discussi trends and campaigns peer companies to dis leads to a better under well-funded threat ac those goals, which ass LinkedIn always stand receive from peers and on-one engagement vendors and investig- disclosures. In turn, we violating content on reports and blog post	These threat exchanges to on amongst industry peo- ; and, 2) one-on-one eng- cuss TTPs and IOCs. This erstanding of the incentive tors and how they evolve ists us in their identificat s ready to receive and i d other external stakehol with peers, we also cor- ate any TTPs and IOCs e also regularly release in our platform in publicly s.	agement with individual exchange of information ves of sophisticated and ve their TTPs to achieve ion and removal. nvestigate any leads we ders. In addition to one_ nsume intelligence from made available in peer formation about policy_ y-available transparency	part on information and t sharing with third partie resources noted in QRE approach Microsoft and operations and sources	threat intelligence gatheres, as well as the inte Es 16.1.1 and 14.1.2. Bing take to monitoring of misinformation and	Given the multipronged g and actioning influence I disinformation and the
			multiple internal and ex	ternal sources relied u g on whether an incider	pon, it is challenging to nce of single information	
			Nr of actions taken (total)	Type of detected content	Other relevant metrics	
Measure 16.2	LinkedIn		•	This measure is not releva		
QRE 16.2.1	We look forward to working on this commitment with the other signatories as we develop further cross-platform information sharing.		service does not have use	ers that can share conte	nt across services.	

## V. Empowering Users Commitments 17 - 25

## V. Empowering Users

<u> </u>	1	17
Comm	itment	
0011111		- · ·

In light of the European Commission's initiatives in the area of media literacy, including the new Digital Education Action Plan, Relevant Signatories commit to continue and strengthen their efforts in the area of media literacy and critical thinking, also with the aim to include vulnerable groups.

	C.17	M 17.1	M 17.2	М 17.3
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search	Bing Search	Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Microsoft has strengthened its partnerships with third-party organisations, including the News Literacy Project and The Trust Project to fund media literacy campaigns while continuing introductory calls with new organizations to grow additional campaigns' reach to new markets.	Microsoft has strengthened its partnerships with third-party organisations, including the News Literacy Project and The Trust Project, to fund media literacy campaigns while continuing introductory calls with new organizations to grow additional campaigns' reach to new markets.
		Microsoft provided pro-bono advertising space across Microsoft surfaces to disseminate the literacy campaigns and helped garner millions of impressions per month.

· · · · · · · · · · · · · · · · · · ·	
	To help educators build Al literacy
	and make the most of Al
	capabilities, we introduced a free
	module on Microsoft Learn:
	Enhancing teaching and learning
	with Copilot. It's designed to guide
	educators through available
	features, learn how to create and
	iterate on prompts, and use
	expertise to evaluate responses for
	quality and credibility.
	Microsoft supported the creation of
	The Investigators, a new world for
	Minecraft Education that helps
	students build information and
	media literacy through game-based
	learning. Launched in English, the
	game and support materials are
	being localized to become available
	globally in 28 languages to millions
	of students and teachers in 2024.
	These languages include Bulgarian,
	Czech, Danish, Dutch, Finnish,
	French, German, Greek, Hungarian,
	Italian, Norwegian, Polish, Slovak,
	Spanish, Swedish, Russian, and
	Ukrainian.
	Microsoft Democracy Forward
	decided to continue supporting
	Verifee, which was launched in Fall
	of 2023, with Azure Credits. Verifee

		is a browser extension and digital literacy program that uses AI to transparently rate the credibility of news articles. It will help you stay safe anywhere on the Internet and not get deceived. With capacity to process large amounts of data, Verifee is able to recognize manipulative elements in the text and mark them safely. Since September 2023, the platform has been analyzing more than 100,000 articles per month and has more than 3000 active users on almost any browser. The project is run from the Czech Republic, but due to similarity in language, it can be used by anybody from Slovakia. The future goal is to scale the project and to support also other language versions.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures Microsoft will expand its efforts to fund and promote media literacy campaigns into more EEA markets and languages. Microsoft will also focus on additional media literacy programs that include greater	Planned Implementation Measures Bing Search is regularly reviewing and evaluating its user tools, policies, and practices related to existing and new Bing generative AI features and adjusts and updates policies as needed. While Bing's

language access and inclusion	existing programs are already
throughout 2023 and into 2024.	designed to address these issues,
	Bing regularly evaluates the efficacy
Beginning March 2024, we	of its measures and endeavors to
launched a new "Be Informed, Not	improve and work to respond
Misled" campaign from the News	quickly to new threats or issues as
Literacy Project through November	they arise.
2024 elections.	
	In addition, Microsoft will expand
	its efforts to fund and promote
	media literacy campaigns into more
	EEA markets and languages.
	Microsoft will also focus on
	additional media literacy programs
	that include greater language
	access and inclusion throughout
	2023 and into 2024.
	Beginning March 2024, Microsoft
	launched a new "Be Informed, Not
	Misled" campaign from the News
	Literacy Project.

Measure 17.1	LinkedIn	Bing Search
QRE 17.1.1	As the world around us changes, LinkedIn continues to evolve and adapt	Bing Search offers a number of tools to help users understand the context
	our systems and practices for combating misinformation and other	and trustworthiness of search results. Even in circumstances where a user is
	inauthentic behaviour on our platform, including to respond to the unique	expressly seeking low authority content (or if there is a data void so little to
	challenges presented by world events.	no high authority content exists for a query), Bing Search provides tools to
		users that can help improve their digital literacy and avoid harms resulting
	LinkedIn's Professional Community Policies, which all members agree to	from engaging with misleading or inaccurate content. For example:
	abide by on joining LinkedIn, prohibit misinformation. As described in	
	more detail in our response to QRE 18.1.1, LinkedIn uses a combination of	

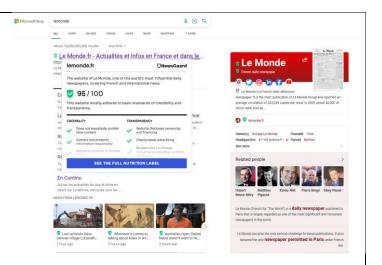
off of Li LinkedI process These of sourced facts a	ated and manual activity to keep content that violates our policies inkedIn. In also aims to educate its members about civic discourse, electoral ses, and public security through its global team of news editors. editors provide each member with relevant, timely information of from credible news sources that educate LinkedIn members about and news events, including global elections, and its content ation teams closely monitor associated conversations in a number	Bing Search often includes answers or public service announcements at the top of search results pointing users to high authority information on a searched topic or warnings on particular URLs known to contain harmful information (such as unaccredited online pharmacies and sites containing malware). Where circumstances warrant (such as public health crises or major elections), Bing Search may provide information hubs for users to easily
of lange	-	access a centralized repository of high authority information.
low aut Invasion append	tion to broader measures, LinkedIn has taken special care to counter hority information in relation to the COVID-19 crisis and the Russian n of Ukraine, as detailed below and further in the Crisis Reporting lices. Some examples of steps we have taken to tackle rmation in connection with unfolding world events include the ng:	Microsoft also partners with NewsGuard to help users evaluate the quality of the news they encounter online. NewsGuard has created trust ratings for 7,500+ news and information sites, which are compiled into a "Nutrition Label" and corresponding Red/Green Reliability Rating to help users understand the reliability of news sources. Within the EU, NewsGuard is currently available in France, Germany, and Italy with plans for future expansion.
	LinkedIn's team of editors covered the <u>most recent developments</u> in the Israel-Hamas war, including promoting updates from key sources to help members stay informed. In response to the COVID-19 pandemic, <u>LinkedIn editors created</u> and promoted trusted content. Further, during the early years of the pandemic, any member that undertook a simple search of the term "coronavirus" was redirected to a link " <i>Know the facts about</i> <i>coronavirus</i> ", which appears first in the list of search results. By clicking on this link, members were directed to <u>LinkedIn's own</u> official page on the coronavirus with information and broadcasts from verified sources, primarily from the World Health Organization. LinkedIn's team of editors covered the most recent developments	Microsoft offers NewsGuard as <u>a free plug-in</u> for the Microsoft Edge web browser (it is also available for other browsers including Chrome and Firefox), and users of the Edge mobile application on both iOS and Android can enable NewsGuard ratings in their app settings. For users with the NewsGuard plug-in, Bing Search results (including in Copilot) include NewsGuard Reliability ratings that lead to a pop-up screen with more site information, an example of which is shown below.
	of Russia's invasion of Ukraine, ranging from the economic impact to major military events that are taking place.	

We also work to identify and remove misinformation and inauthentic behaviour from our platform. As we continue to improve, we are committed to helping our members make informed decisions about content they find on LinkedIn, so we work with Microsoft to provide tools that assist our members in identifying trustworthy, relevant, authentic, and diverse content.

LinkedIn's <u>Professional Community Policies</u> clearly detail the objectionable and harmful content that is not allowed on LinkedIn. Misinformation and inauthentic content is not allowed, and our automated defenses take proactive steps to remove them. LinkedIn's blog provides information regarding our efforts, including <u>How We're Protecting Members From Fake</u> <u>Profiles</u>, <u>Automated Fake Account Detection</u>, and <u>An Update on How We Keep Members Safe</u>.

LinkedIn members can <u>report content</u> that violates our Professional Community Policies, including misinformation and inauthentic content. Our Trust and Safety teams work every day to identify and restrict such activity, and if reported content violates the Professional Community Policies, it will be actioned in accordance with our polices.

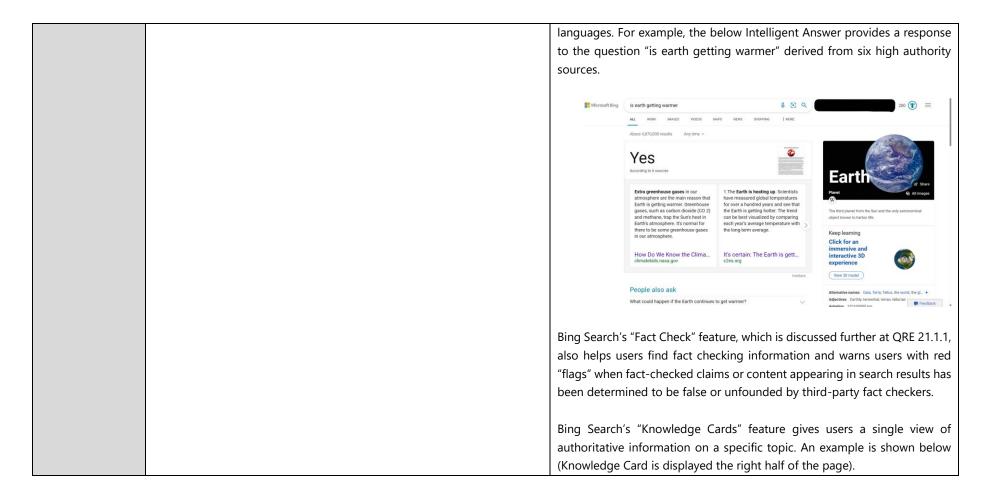
LinkedIn members can identify misinformation and inauthentic behaviour by utilising the <u>News Literacy Project</u>, <u>The Trust Project</u> and <u>Verified</u>, all of which develop information literacy campaigns built on industry research and best practices. The News Literacy Project campaign developed a <u>quiz</u> that tests a person's ability to identify why the information they are seeing is false and inaccurate in less than five minutes. The Trust Project campaign developed the research-backed <u>8 Trust Indicators</u>, which aim to improve consumers ability to identify reliable, ethical journalism. Finally, <u>Verified</u> delivers lifesaving information and fact-based advice to build digital literacy that helps communities protect themselves from misinformation. LinkedIn has also published an <u>article</u> in our Help Center compiling these useful resources on misinformation and inauthentic behaviour.



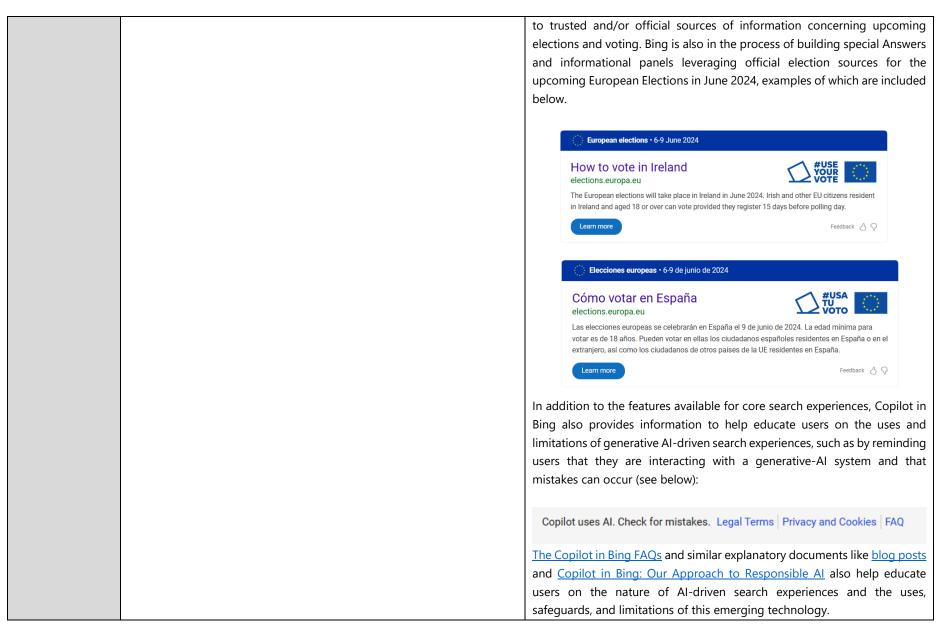
Bing Search's <u>Page Insights</u> feature also helps direct users to authoritative resources and information. The feature, which appears as a light bulb image next to search results, provides users with additional information about the site and its contents from third party information sites such as Wikipedia. An example is shown below.



Bing Search <u>Intelligent Answers</u> also provides users with informative panels and direct answers to certain search queries, and is now available in 100







For example, <u>the Copilot FAQ answer</u> to "Are Copilot's AI-generated responses always factual?" explains: "Copilot aims to base all its responses on reliable sources - but AI can make mistakes, and third-party content on the internet may not always be accurate or reliable. Copilot will sometimes misrepresent the information it finds, and you may see responses that sound convincing but are incomplete, inaccurate, or inappropriate. Use your own judgment and double check the facts before making decisions or taking action based on Copilot's responses."

In another prominent example, another <u>Copilot support page</u> reminds users prominently that while they "lead the way", they need to check facts, review, and avoid simply relying on AI as an "autopilot" (shown below).

You lead the way

Unleash your creativity and get things done with Copilot by your side. Since Al-generated content may be incorrect, here are a few things to remember...



Microsoft also offers meaningful resources for users interested in learning more about generative AI features and tools, including Copilot, though blog posts, articles, information hubs, and support pages. In addition to teaching AI basics and how-tos, these resources reiterate the importance of checking AI-generated materials and understanding the strengths and limitations of AI. See e.g., <u>Microsoft AI help & learning</u>. For example, a recent Copilot support article (<u>Unleash your productivity with AI and Microsoft Copilot – Microsoft Support</u>) has a section specifically directing users to "Be Aware of

				outputs may they were tra AI may not ku interpret sarc Microsoft is o guides so tha and will cor	s", which explain contain inaccura ined on information now about recent casm, irony, or hur committed to pro at users can devel ntinue to explore merative AI topics	cies, "biases, or on from the inter events yet, and s mor." viding resources op literacy when a ways to furth	sensitive materi net, as well as ot struggles to und , educational m interacting with	ials because her sources. erstand and aterials, and n AI systems
SLI 17.1.1 - action enforcing policies above		s the number of visitors to resources on misinforma 3.	•	NewsGuard I NewsGuard to during the Re Knowledge C all types/topic Transparency Microsoft Tran Public Servic announcemer	of data measurer Impressions ("No pol was rendered i porting Period. Cards ("KC)" – Re cs) during the Rep y Hub Viewership nsparency Report e Announcement nt panels (of all porting Period.	GI") – Represen n the Edge brow epresents viewer oorting Period p ("TH") – Repr Hub during the t ("PSA") – Repr	ser discover pan ship of Knowled esents the total Reporting Perio esents views of p	e to EU users ge Cards (of views of the d. public service
	Total count of the tool's impressions Number of visits during the period 1	Interactions/ engagement with the tool	Other relevant metrics	Total count of the tool's impressions	Interactions/ engagement with the tool	Total count of the tool's impressions	Total count of the tool's impressions	Other relevant metrics

	July – 31 December	NGI	КС	ТН	PSA	[
	2023					ļ
Member States						ļ
Austria	0	70	119.100M	256	105,840	
Belgium	0	59	219.657M	387	223,740	
Bulgaria	0	25	41.359M	97	23,700	
Croatia	2	9	26.491M	65	20,700	
Cyprus	0	0	10.969M	42	6,060	
Czech Republic	10	52	133.750M	344	71,760	
Denmark	5	37	82.208M	367	60,880	
Estonia	1	3	14.668M	75	12,760	
Finland	0	14	89.442M	274	50,200	
France	224	542	1.135B	1,452	789,060	
Germany	54	824	1.322B	3,752	997,640	
Greece	8	38	61.293M	151	33,240	
Hungary	0	28	66.335M	207	37,080	
Ireland	0	31	121.054M	371	467,420	
Italy	24	304	740.152M	884	353,000	
Latvia	0	6	16.408M	71	26,980	
Lithuania	0	21	26.358M	48	5,660	
Luxembourg	0	4	11.350M	34	12,060	
Malta	0	3	9.992M	18	8,060	
Netherlands	19	70	360.029M	2,392	403,900	
Poland	27	300	494.769M	613	275,900	
Portugal	2	57	151.180M	301	79,660	
Romania	0	25	79.561M	221	53,260	
Slovakia	5	12	40.591M	103	208,380	
Slovenia	0	4	17.551M	45	16,380	
Spain	59	244	778.205M	960	378,220	
Sweden	7	44	180.822M	747	31,160	

Iceland	0		1	6.588M	22	5,640	
Liechtenstein	0		0	703.389K	9	520	
Norway	1		39	96.328M	390	78,580	
Total EU	447		2,826	6.350B	14,277	4,752,700	
Total EEA	448		2,866	6.454B	14,698	4,837,440	

Measure 17.2	LinkedIn	Bing Search
QRE 17.2.1	Microsoft works with leading media and information literacy partners globally to support the development and promotion of media literacy campaigns.	
	<ul> <li>Microsoft has strengthened its partnerships with third-party organisations, including_the News Literacy Project and The Trust Project, to fund media literacy campaigns while continuing introductory calls with new organizations to grow additional campaigns' reach to new markets. Beginning March 2024, we launched a new "Be Informed, Not Misled" campaign from the News Literacy Project.</li> <li>For the next reporting period, Microsoft is continuing to work with existing and new partners to create, disseminate, and, where available, report on expanded literacy campaigns in EEA markets, including in additional EEA languages. Please also see response to QRE 17.1.1.</li> </ul>	Since its last report, Microsoft continues to grow partnerships to strengthen the company's capacity and ability to combat information operations globally. Microsoft works with EFE to use their EFE Verifica to identify, track and expose emerging information operations spreading across Spanish- speaking communities around the world and Reporters Sans Frontières (RSF) to use their Journalism Trust Initiative (JTI) dataset to proactively promote trusted sources of news around the world that follow transparent and accepted journalistic practices. RSF's JTI data has expanded since our last report and currently includes Bulgaria, France, Greece, Germany, Latvia, Poland, and Romania. Microsoft Democracy Forward and RSF are working to expand the dataset into more countries throughout 2024.
		Microsoft has strengthened its partnerships with third-party organisations, including the News Literacy Project and The Trust Project to fund media literacy campaigns while continuing introductory calls with new organizations to grow additional campaigns' reach to new markets. Microsoft provided pro-bono advertising space across Microsoft surfaces to disseminate the literacy campaigns and helped garner millions of impressions per month. Beginning March 2024, we launched a new "Be Informed, Not Misled" campaign from the News Literacy Project.

Additionally, Search Coach is a free app in Microsoft Teams that helps educators and students to form effective queries and identify reliable resources. It is designed to teach information literacy skills in a safe, secure, and ad-free environment. Search Coach is available throughout the EU.

Microsoft Democracy Forward is supporting Verifee, launched in Fall of 2023, with Azure Credits. Verifee is a browser extension and digital literacy program that uses AI to transparently rate the credibility of news articles. It will help you stay safe anywhere on the Internet and not get deceived. With capacity to process large amounts of data, Verifee is able to recognize manipulative elements in the text and mark them safely. Since September 2023, the platform has been analyzing more than 100,000 articles per month and has more than 3000 active users on almost any browser. The project is run from the Czech Republic, but due to similarity in language, it can be used by anybody from Slovakia. The future goal is to scale the project and to also support other language versions.

In February 2024, Microsoft launched several collaborations with news organizations to adopt generative AI. Through these new programs, we are helping these organizations identify and refine the procedures and policies to use AI responsibly in newsgathering and business practices, helping train a new generation of reporters on best uses of AI and identify ways AI can help create efficient business practices and help build sustainable newsrooms for generations to come. One collaboration is with Semafor to work with us to harness AI tools to assist journalists in their research, source discovery, translation, and more with Semafor Signals, helping journalists provide a diverse array of credible local, national, and global sources to their audience as well as create a Global Elections Hub.

For the next reporting period, Microsoft is continuing to work with existing and new partners to create, disseminate, and where available, report on expanded literacy campaigns in additional EEA markets, including in additional EEA languages. Please also see QRE 17.1.1

SLI 17.2.1 -	Methodology of data measurement:					Methodology of data measurement:				
actions enforcing	Nr of media	Reach of	Nr of	Nr of	Nr of	Nr of media	Reach of	Nr of	Nr of	Nr of
policies above	literacy/	campaigns	participants	interactions	participants	literacy/	campaigns	participants	interactions	participants
	awareness			with online	(etc)	awareness			with online	(etc)
	raising			assets		raising			assets	
	activities					activities				
	organised/					organised/				
	participated					participated				
	in					in				
						3				

t continues to work with multiple organisations to develop and media literacy campaigns, including new campaigns from the News Project and The Trust Project to promote information literacy
s on Microsoft platforms.
last report, Microsoft continues to grow partnerships to strengthen apany's capacity and ability to combat information operations Microsoft works with EFE to use their EFE Verifica to identify, track ose emerging information operations spreading across Spanish- g communities around the world and Reporters Sans Frontières (RSF) heir Journalism Trust Initiative (JTI) dataset to proactively promote sources of news around the world that follow transparent and d journalistic practices. RSF's JTI data has expanded since our last nd currently includes Bulgaria, France, Greece, Germany, Latvia, and Romania. Microsoft Democracy Forward and RSF are working id the dataset into more countries throughout 2024.
l I I I I I I I I I I I I I I I I I I I

will help you stay safe anywhere on the Internet and not get deceived. With capacity to process large amounts of data, Verifee is able to recognize manipulative elements in the text and mark them safely. Since September 2023, the platform has been analyzing more than 100,000 articles per month and has more than 3000 active users on almost any browser. The project is run from the Czech Republic, but due to similarity in language, it can be used by anybody from Slovakia. The future goal is to scale the project and to support also other language versions.
In February 2024, Microsoft launched several collaborations with news organizations to adopt generative AI. Through these new programs, we are helping these organizations identify and refine the procedures and policies to use AI responsibly in newsgathering and business practices, helping train a new generation of reporters on best uses of AI and identify ways AI can help create efficient business practices and help build sustainable newsrooms for generations to come. One collaboration is with Semafor to work with us to harness AI tools to assist journalists in their research, source discovery, translation, and more with Semafor Signals, helping journalists provide a diverse array of credible local, national, and global sources to their audience as well as create a Global Elections Hub.
For the next reporting period, Microsoft is continuing to work with existing and new partners to create, disseminate, and report on expanded literacy campaigns in EEA markets, including in additional EEA languages.

V. Empowering Users				
Commitment 18				
Relevant Signatories commit to minimise the risks of viral propagation of Disinformation by adopting safe design practices as they develop their systems, policies, and features.				

	C.18	M 18.1	M 18.2	M 18.3
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search		Bing Search	Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No	Yes
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	<ul> <li>Bing regularly reviews and implements mitigations, safeguards, and safe design considerations to help proactively address, prevent and mitigate harms arising from potential misuse of generative AI search experiences, including viral propagation of content. However, Copilot in Bing features do not allow users to post or share content within Bing so virality is not possible on the platform itself.</li> <li>In addition, since its last report Bing has implemented additional risk assessment protocols to analyze risks on its platform, has improved internal compliance programs related to Responsible AI and digital safety, and updated user tooling for reporting concerns.</li> </ul>

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes	Microsoft and Bing have also undertaken additional research projects related to these commitments. Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures LinkedIn will continue to assess its policies and services and to update them as warranted.	Planned Implementation Measures Bing is regularly reviewing and evaluating its policies and practices related to features and adjusts and updates policies as needed. Bing continues to explore additional potential research opportunities and partnerships related to the spread of harmful misinformation and/or disinformation. Microsoft Research will share findings on internal research concerning information integrity and elections in the age of generative AI. Findings were presented as part of Microsoft's Research forum on March 7, 2024 and will be published publicly soon.

Measure 18.1	LinkedIn	This measure is not relevant or pertinent to Bing Search as search engines do
QRE 18.1.1	Outline relevant actions	not fuel virality of disinformation via users sharing content.
	With respect to AI design, please see QRE 18.1.3	

	<ul> <li>With respect to additional tools, procedures, or features, please see:</li> <li>QRE 17.1.1 (editorial practices to provide members with trustworthy news);</li> <li>QRE 18.2.1 (policies and procedures to limit spread of harmful false or misleading information);</li> <li>QRE 21.1.1 (action taken when information is identified as misinformation);</li> <li>QRE 22.1.1 (features and systems related to fake and inauthentic or file).</li> </ul>	
	<ul> <li>profiles);</li> <li>QRE 22.2.1 (actions taken to assist members identify trustworthy content);</li> <li>QRE 23.2.1 (actions taken to ensure integrity of reporting and appeals process).</li> </ul>	
QRE 18.1.2	<ul> <li>The main parameters of the LinkedIn feed recommender systems are as follows:</li> <li>Identity: We seek to contextualise content based on who a member is by looking at their profile, for example: Who are you? Where do you work? What are your skills? Who are your connections? Where is your profile location?</li> <li>Content: We aim to match appropriate content to each member by evaluating, for example: How many times was the feed update viewed? How many times was it reacted to? What is the content about? How old is it? Is the update sharing knowledge or professional advice? What language is it written in? Is the conversation constructive and professional? Will engagement on the update lead to future high-quality content? What companies, people, or topics are mentioned in the update?</li> <li>Member Activity: Finally, we look at how a member engages with content and examine, for example: What have you reacted to and shared in the past? Who do you interact with most frequently or recently? Where do you spend the most time in your feed? Which hashtags, people or companies do you follow? What types of topics are you interested in? What other members follow you? What</li> </ul>	

	actions have other members taken on your posts? How long has it	
	been since the foregoing actions took place?	
	Combining these signals, the LinkedIn feed recommender system ranks the	
	content for the member, with the goal of showing the member high quality	
	content that the member will enjoy consuming and can lead to further	
	creation on the platform. To do this, the feed optimizes for content that a	
	member is most likely to find highly valuable, which in turn is likely to lead	
	the member to act on (e.g., react, comment, or reshare).	
QRE 18.1.3	At LinkedIn, our guiding principle is "Members First." It ensures we honour	
QILL IO.I.S	our responsibility to protect our members and maintain their trust in every	
	decision we make, and puts their interests first. A key area where we apply	
	this value in engineering is within our design process. We call this	
	"responsible design," which means that everything we build is intended to	
	work as part of a unified system that delivers the best member experience,	
	provides the right protections for our members and customers, and	
	mitigates any unintended consequences in our products.	
	One of the core pillars of "responsible design" is "responsible AI," which	
	follows the LinkedIn Responsible AI Principles, which are inspired by and	
	aligned with <u>Microsoft's Responsible AI Principles</u> . The LinkedIn Responsible	
	Al Principles are to advance economic opportunity, uphold trust, promote	
	fairness and inclusion, provide transparency, and embrace accountability. In	
	addition to the LinkedIn Responsible AI Principles, responsible AI is also	
	about intent and impact. "Intent" involves evaluating training data, designing	
	systems, and reviewing model performance before the model is ever	
	deployed to production to make sure that our principles are reflected at	
	every step in the process. It includes actively changing our products and	
	algorithms to empower every member. "Impact" covers detecting and	
	monitoring the ways that people interact with products and features after	
	they are deployed. We do this by measuring whether they provide significant	

value and empower individuals to reach their goals. Intent and impact are a cyclical process of refinement that go hand-in-hand towards the broader goal of responsible design.

With respect to safety, we seek to keep content that violates our Professional Community Policies off of LinkedIn. This is done through a combination of automated and manual activity. Our first layer of protection is using AI to proactively filter out bad content and deliver relevant experiences for our members. We use content (like certain key words or images) that has previously been identified as violating our content policies to help inform our AI models so that we can better identify and restrict similar content from being posted in the future. The second layer of protection uses AI to flag content that is likely to be violative for human review. This occurs when the algorithm is not confident enough to warrant automatic removal. The third layer is member led, where members report content and then our team of reviewers evaluates the content and removes it if it is found to be in violation of our policies.

Quantifying the above process to monitor how many content violations are successfully prevented is another important task that our Data Science team prioritises, such that we can continuously refine our processes to improve detection and prevention of violative content.

Please also see:

- QRE 17.1.1 (editorial practices to provide members with trustworthy news);
- QRE 18.2.1 (policies and procedures to limit spread of harmful false or misleading information);
- QRE 21.1.1 (action taken when information is identified as misinformation);
- QRE 22.1.1 (features and systems related to fake and inauthentic profiles);

	QRE 22.2.1 (actions tak	• QRE 22.2.1 (actions taken to assist members identify trustworthy					
	content);						
	QRE 23.2.1 (actions take	n to ensure integri	ty of reporting a	nd appeals			
	process).						
SLI 18.1.1 -	We continue to explore opportunities to effectively report on the measures						
actions	outlined in QRE 18.1 in future	outlined in QRE 18.1 in future reporting periods.					
proving	Reduction of prevalence of	Reduction of	Increase in	Other			
effectiveness	disinformation	views/	visibility of	relevant			
of measures		impressions of	authoritative	metrics			
and policies		disinformation	information				

Measure 18.2	LinkedIn	Bing Search
QRE 18.2.1	LinkedIn is an online professional network. On LinkedIn, the world's	Unlike social media services, for which this Commitment appears primarily
	professionals come together to find jobs, stay informed, learn new skills, and	oriented, search engines do not typically cause or facilitate the viral
	build productive relationships. The content that our members share	propagation of disinformation, as they do not allow users to post or share
	becomes part of their professional identity and can be seen by their boss,	content directly on the service. Please see <u>How Bing delivers search results</u>
	colleagues, and potential business partners. Accordingly, the content on	and Microsoft Bing Webmaster Guidelines for an overview of how Bing
	LinkedIn is professional in nature.	Search designs its algorithms to deliver high authority and highly relevant
		content while minimizing the negative impact of spam and less credible
	To help keep LinkedIn safe, trusted, and professional, our Professional	information sources. Bing Search's ranking algorithms and related policies
	<u>Community Policies</u> clearly detail the range of objectionable and harmful	are intended to address deceptive tactics intended to manipulate the
	content that is not allowed on LinkedIn. Fake accounts, misinformation, and	algorithms and are discussed in more detail at QREs 14.1.1 and 14.1.2. Bing
	inauthentic content are not allowed, and we take active steps to remove it	Search features such as Answers, Page Insights, NewsGuard ratings, and
	from our platform.	ClaimReview fact checks, which are discussed at QRE 17.1.1. and 21.1.1, as
		well as the other features and policies discussed throughout this report
	LinkedIn removes "specific claims, presented as fact, that are demonstrably	further help minimize the risk of viral propagation of misinformation through
	false or substantially misleading and likely to cause harm." This approach	Bing Search.
	applies globally and is used for purposes of content moderation and for	
	publicly reporting figures on misinformation. Specific examples of what	Bing's suggestions features offer possible search queries to users to facilitate
	might constitute misinformation can be found here in our Help Center. As	a more efficient search experience. While search suggestions are not directly
	part of our User Agreement, our Professional Community Policies are	tied to virality of content, Bing Search also undertakes measures to help

accepted by every member when joining LinkedIn and are easily available to every member. ensure it does not inadvertently lead users to misleading or other harmful content through suggestions. Specifically, Bing Search uses a combination of

LinkedIn creates value and preserves trust by fostering a safe, trusted, and professional platform, while honouring members' professional expression and speech. LinkedIn enables healthy on-platform conversations by facilitating the removal of misinformation that threatens its members' safety. And when content doesn't conclusively violate LinkedIn policies, LinkedIn gives the speaker the benefit of the doubt and favours speech (i.e., leaves the content up on platform).

Additionally, as described in greater detail below, human review plays a significant role in our content moderation process. Additionally, Members who post content and members who report content can appeal our content moderation decisions.

Our content policies are clear and we apply them equally for all members. Within our Professional Community Policies we provide granular information and examples on what is and what is not allowed on LinkedIn.

Furthermore, LinkedIn has automated defences to identify and prevent abuse, including inauthentic behaviour, such as spam, phishing and scams, duplicate accounts, fake accounts, and misinformation. Our Trust and Safety teams work every day to identify and restrict inauthentic activity. We're regularly rolling out scalable <u>technologies</u> like machine learning models to keep our platform safe.

As inauthentic behaviour gets more sophisticated, we're improving our detection. Here are some of the latest actions we've taken on fake profiles to help keep our members safe while engaging in our community.

ensure it does not inadvertently lead users to misleading or other harmful content through suggestions. Specifically, Bing Search uses a combination of proactive and reactive algorithmic and manual interventions to prevent the display of search suggestions that could lead to low authority content.

Bing Search also provides a tool for users to provide feedback on suggestions they encounter. The feedback tool is shown below.

Microsoft Bing	why is the sky blue	× \$ © Q
	Q why is the sky	Related Searches
	Q why is the sky blue	Q why the skies are blue
	Q why is the sky <b>yellow</b>	Q why is the sky green
om spaceplace	Q why is the sky blue on earth	Q why is the sky blue for kids
ontent	Q why is the sky orange	People Also Ask
It's Easy to See	Q why is the sky red	Q is the sky really blue and why?
What Makes A	Q why is the sky purple	Q Why is the sky used to be more blue?
Is The Sky Blue	Search Microsoft for "why is the sky"	Why is the sky used to be more blue?
is the sky blue	Manage Your Search History	
	it took a long time to figure it out! The light	Feedback on these suggestions

Clicking "Feedback on these suggestions" allows users to provide specific feedback on individual suggestions on the below bases:

## These suggestions are:

Not relevant
 Not correct
 Misspelled or garbled
 Inappropriate or offensive
 Hateful to groups or individuals
 About illegal or dangerous activity
 Other

In Copilot in Bing, there is no risk of viral spread of generated content through Bing because Bing still does not allow users to directly post or otherwise share content on the platform, including from Copilot. However, Copilot in Bing has taken steps to prevent the service from being used to create misinformation that might be shared on other platforms through a

		multipronged approach to mitigations that helps prevent users from creating misinformative content through Copilot in Bing. This system includes terms of use and a code of conduct, classifiers, filters, a bespoke metaprompt, and robust reporting mechanisms designed to mitigate the risk of potential misuse of the platform. The Supplemental Terms governing usage of Copilot in Bing prohibit users from using the service to generate fraudulent or misleading information, including the creation of disinformation. Bing's ranking and relevance systems for search, which are an essential component to answering user questions in Copilot, work to ensure that high authority content is returned first in search results in traditional search and in chat. Where Copilot in Bing's systems flag that a user's prompt or generated output may result in low authority or misleading information, the system will take steps to mitigate that possible harm through solutions, such as not returning generated content to the user, diverting the user to a different topic, or redirecting the user to traditional search. Users who encounter problematic content can report concerns via Feedback or Report a Concern tools for both Bing and Copilot.
SLI 18.2.1 -	Methodology of data measurement:	Methodology of data measurement:
actions taken in response to	The table below reports metrics concerning content LinkedIn removed from	Bing Search does not have data relevant to this SLI. Users come to Bing
policy	its platform as Misinformation, pursuant to the policy outlined in QRE 18.2.1	Search with a specific research topic in mind and expect Bing Search to
violations	above. The metrics include:	provide links to the most relevant and authoritative third-party websites on
	- the number of pieces of content removed as Misinformation	the Internet that are responsive to their search terms. Bing Search does not
	<ul> <li>between 1 July - December 2023, broken out by EEA Member State;</li> <li>the number of those content removals that were appealed by the content author;</li> <li>the number of those appeals that were granted;</li> <li>the median time from appeal-to-appeal decision for those appeals. The metrics are assigned to EEA Member State based on the IP address of the of the content author.</li> </ul>	have a news feed for users of user content, allow users to post and share content within Bing, or otherwise enable content to go "viral" on Bing. See response to SLI 14.2.1 for relevant metrics.

Mambag Strive	The number of pieces of content removed as Misinformation between 1 July – 31 December 2023	The number of removals that were appealed by the content author	The number of appeals that were granted	The median time from appeal- to- appeal decision in hours	Total no of violations	Metric 1: indicating the impact of the action taken	Metric 2: indicating the impact of the action taken	Metric 3: indicating the impact of the action taken
Member States	255	2	2	0.5 have				
Austria Belgium	255 652	3 8	2	8.5 hours				
Bulgaria	110	0 1	0	_				
Croatia	101	3	0					
Cyprus	38	2	0	_				
Czech Republic	113	1	0	_				
Denmark	305	5	0					
Estonia	14	0	0					
Finland	52	0	0					
France	4,796	24	0					
Germany	2,219	28	2	-				
Greece	209	1	0	-				
Hungary	75	1	0	1				
Ireland	192	3	0					
Italy	3,627	12	0					
Latvia	19	0	0					
Lithuania	27	0	0					
Luxembourg	85	0	0					
Malta	24	0	0					
Netherlands	4,137	50	3					
Poland	280	4	0					
Portugal	344	4	0					

Total EEA	21,206	164	8
Total EU	21,130	164	8
Norway	66	0	0
Liechtenstein	1	0	0
Iceland	9	0	0
Sweden	286	2	0
Spain	2,953	11	1
Slovenia	21	0	0
Slovakia	24	1	0
Romania	172	0	0

Measure 18.3	LinkedIn	Bing Search
QRE 18.3.1	Through the acquisition of Miburo in July 2022, Microsoft now has an	Bing Search regularly reviews and considers safe design practices and
	internal research team — the Microsoft Threat Analysis Center (MTAC) —	research and conducts user studies as part of its product and new feature
	that conducts research on information influence operations and publishes	development processes. Bing Search employees have actively partnered with
	both internal and public reports on its findings across world regions.	Microsoft Research and third-party research organizations to contribute to
		novel research and internal studies concerning safe design practices,
	Microsoft also funds and works with Princeton University on the creation of	responsible AI, and disinformation.
	hub for researchers to access data from social media companies to improve	
	the identification and tracking of information operations. This accelerator	During the reporting period, Bing engaged third party experts to support
	will be available to researchers around the world including in Europe.	internal evaluations of misinformation risks related to image generation
		features. Such research helps to inform Bing and Copilot safety mitigations.
		During the reporting period, Bing also prepared the launch of its Bing
		Qualified Researcher Program to enable EU researchers to request access for
		publicly accessible Bing data from a singular landing page.
		As part of preparations for global elections, Microsoft Research recently
		conducted internal research concerning information integrity and elections
		in the age of generative AI. One related segment "Generative AI and Plural
		Governance: Mitigating Challenges and Surfacing Opportunities" was
		presented publicly as part of Microsoft's Research forum on March 7, 2024

and will be shared to the research community at https://researchforum.microsoft.com/.
Microsoft also funds and works with Princeton University on the creation of hub for researchers to access data from social media companies to improve the identification and tracking of information operations. This accelerator will be available to researchers around the world including in Europe.
Microsoft Research and Microsoft's AI for Good Lab regularly undertake and publish research that addresses or can be used in understanding online misinformation and disinformation. Microsoft researchers are currently engaged in research leveraging search data to explore how medical hoaxes went viral during the COVID-19 pandemic and research concerning the detection of bias in mainstream news in connection with elections.
Microsoft maintains an internal research team—the Microsoft Threat Analysis Center (MTAC)—that conducts research on information influence operations and publishes both internal and public reports on its findings. MTAC maintains global hubs and conducts intelligence analysis in over 13 languages. Additionally, Microsoft funds and works with external organizations, the Global Disinformation Index, NewsGuard, EFE, and the Alliance for Securing Democracy, to ingest data and research that they conduct into Microsoft products, including Bing Search and LinkedIn. Microsoft also currently funds and works with Carnegie Foundation on the creation of hub for researchers to access data from social media companies to improve the identification and tracking of information operations.
Bing Search looks forward to continued opportunities to contribute to and collaborate with the research community on future research and is in active discussions with third party organizations and the research community on best practices and mitigations for core web search and new generative AI experiences.

V. Empowering Users								
Commitment 19								
<b>3 3</b>	Relevant Signatories using recommender systems commit to make them transparent to the recipients regarding the main criteria and parameters used for prioritising or deprioritising information, and provide options to users about recommender systems, and make available information on those options.							
C.19 M 19.1 M 19.2								
We signed up to the following LinkedIn LinkedIn LinkedIn								
measures of this commitment:	Bing Search	Bing Search	Bing Search					

	LinkedIn	Bing Search
In line with this commitment, did you deploy new	Yes	Yes
implementation measures (e.g. changes to your terms of		
service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet	New Implementation Measures	New Implementation Measures
points].		
	As mentioned in our prior report, LinkedIn	Bing has published updates to "How Bing Delivers
	launched in the EU a revised and expanded	Search Results" for users to better understand how our
	experience to enable Members to change how	search engine delivers the search results and content
	their Feed experience is presented to them.	presented to users, including options for users to control
	The choice is presented in the Feed and also	how that content is delivered or personalized.
	points members to the setting referenced	
	above where members can change the default	
	sort of their Feed.	

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	Not applicable	Bing regularly updates its policies and terms and conditions to account for product changes, user feedback, and evolving legal considerations.

Measure 19.1	LinkedIn	Bing Search				
QRE 19.1.1	LinkedIn has published a variety of articles to explain to users how our	Bing's search engine ranking algorithms are not a traditional				
	recommender systems work, including: <u>"Mythbusting the Feed: How the</u>	"recommender system" in that Bing Search only provides content to users				
	Algorithm Works"; "Mythbusting the Feed: Helping our members better	as a result of their express request, rather than pushing content to users				
	understand LinkedIn"; "Keeping your feed relevant and productive";	who were not expressly seeking it. That said, the main parameters of Bing				
	LinkedIn Safety Series: Using AI to Protect Member Data; "Guide: Features	Search's ranking algorithms are published in the "How Bing Ranks Search				
	to Help You Control Your Feed and Conversations"; Our approach to	Results" section of <u>How Bing Delivers Search Results</u> , which is available to				
	building transparent and explainable AI systems. During the prior	Bing Search users in the EU. Bing Search also provides information on				
	reporting period LinkedIn collated and expanded upon existing resources	how it ranks and returns search suggestions in the Enhanced Search				
	to further explain the main parameters of LinkedIn recommender systems	Experiences section of <u>How Bing Delivers Search Results</u> . Bing's ranking				
	and options provided to users to influence and control these	algorithms apply equally to traditional search results and generative AI				
	recommender systems.	features that extend to Bing.				
	Additionally, LinkedIn addresses automated processing and relevancy in	More information about how Microsoft moderates generative AI features				
	the LinkedIn <u>User Agreement</u> and it includes a link to the above referenced	to prevent users from being presented with harmful information is				
	Help Centre article at the end of Section 3.6 of the LinkedIn User	available at: https://aka.ms/responsibleAI-CopilotinBing.				
	Agreement. During the prior reporting period, LinkedIn launched a new					
	setting for members to control the default for how their LinkedIn feed is	Please also see QREs 14.1.1, and 22.2.1.				
	presented to them. Members can now change their preferred feed view					
	from "most relevant first" to "most recent first". "Most relevant first" means					
	that LinkedIn will use data from the member's profile and LinkedIn activity					
	data to rank feed content based on the member's interests. "Most recent					

	first" means that LinkedIn will not use the member's profile and LinkedIn	
	activity data to rank feed content and will instead show content in reverse	
	chronological order.	
	<ul> <li>As reported in our prior report, in August 2023, LinkedIn launched two new experiences in the EU. Additional detail is included below-</li> <li>LinkedIn launched a revised and expanded experience to enable Members to change how their Feed experience is presented to them. The choice is presented in the Feed (on desktop, mobile app, and mobile web) and it also points members to the setting referenced above where members can change the default sort of their Feed. Members can toggle between the following two choices: "most relevant first" or "most recent first." The default sort option is "most relevant first." If the Member toggles to "most recent first," that choice will only persist for the current feed view on that particular device. LinkedIn also launched a new setting within a Member's Account Preferences settings so Members can change the default sort option from "most relevant first" to "most recent first." Changing that setting will persist across sessions and</li> </ul>	
	devices. Members can learn more about this experience and the	
	setting in our Help Center.	
Measure 19.2	LinkedIn	Bing Search
SLI 19.2.1 – user	Methodology of data measurement:	Methodology of data measurement:
settings		
	Members that do not wish to have their LinkedIn feed experience sorted by relevance can elect to have their LinkedIn feed sorted chronologically using the "Sort by" functionality above the feed. This functionality only controls the current session and will revert to the "relevant" feed experience when the member next visits LinkedIn.	Bing Search allows users to turn off search suggestions (including auto- suggest and related search suggestions) ("AS/RS") in its user setting page, as shown below: Search suggestions
	In addition to the above functionality, a new setting was introduced to Members to enable them to change the default of how their LinkedIn feed experience is sorted.	In the Bing image experience, users can turn off personalized search suggestions through the Settings pane. Bing anticipates providing reporting on utilization of this new measure in forthcoming reports.

	The tables below			Users may also access, view, and delete their previous search queries in their Microsoft Account Privacy dashboard or clear their search history in			
	Table A: (1) the r	number of EEA mer	nbers who used th	e feed "Sort by"	Bing Search settings, which in turn will remove that content from any		
	functionality betw	veen 1 July - 31 Dec	ember 2023; and (2	2) the number of	personalized search suggestions.		
	times those mem	pers used the feed "	Sort by" functionalit				
	- 31 December 20	23;		Bing is currently building out expanded data retention and reporting			
					functionalities.		
	(1) the number of	EEA members who u	used the "preferred f	eed view" setting			
	between 1 July -	31 December 2023;	and (2) the number	er of times those	Users who are Microsoft Account users may also add or remove a topic		
	members used th	ne "preferred feed	view" setting betw	veen 1 July - 31	from their Ad interests through the Microsoft Account Settings page at		
	December 2023 .		-	-	account.microsoft.com/privacy/ad-settings ("Ad Interest"). Figures on		
					the amount of users that have globally added or removed Ad Interests		
	The metrics are assigned to EEA Member State based on the self-reported profile location of the member.				are below. This information cannot be provided at an EU member state		
					level and therefore the below Ad Interest figures represent global		
	· ·				numbers.		
	Table A: No of ti	Table A: No of times users actively		nes users actively			
	engaged with th	ne feed "Sort by"	engaged with the "preferred feed				
		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		etting			
	The number of	The number of	The number of	The number of	No of times users actively engaged with these settings		
	EEA members	times the	EEA members	times the			
	who used the	members used	who used the	members used	(Ad Interest)		
	feed "Sort by"	the feed "Sort	"preferred feed	the "preferred			
Data	functionality	by" functionality	view" setting	feed view"			
	between 1 July -	between 1 July -	between 1 July -	setting			
	31 December	31 December	31 December	between 1 July			
	2023	2023	2023	- 31 December			
				2023			
Member States							
Austria	4,768	43,901	5,879	3,281			
Belgium	9,270	97,030	4,456	5,137			
Bulgaria	1,142	14,311	2,335	840			

Croatia	958	11,935	551	919		
Cyprus	704	6,831	338	461		
Czech Republic	2,938	29,248	15,382	1,717		
Denmark	8,241	87,467	7,460	3,727		
Estonia	675	7,192	5,879	416		
Finland	6,191	65,577	4,456	4,225	1	
France	54,041	499,593	2,335	38,098		
Germany	50,965	480,906	551	35,538		
Greece	2,713	33,566	338	1,796		
Hungary	1,483	14,083	15,382	967		
Ireland	7,407	91,409	2,988	3,697		
Italy	19,728	199,768	9,546	11,938	1	
Latvia	537	6,612	311	386	1	
Lithuania	1,024	11,461	501	702		
Luxembourg	1,053	14,133	605	745	1	
Malta	499	5,797	201	242		
Netherlands	36,222	398,189	17,539	21,596		
Poland	12,592	139,045	4,747	5,879		
Portugal	7,824	78,543	3,619	4,456		
Romania	2,018	-21,656	1,749	2,335		
Slovakia	891	8,570	450	551		
Slovenia	589	5,357	284	338		
Spain	24,532	218,611	12,193	15,382		
Sweden	8,311	75,336	6,087	7,460		
Iceland	267	2,908	43	56		
Liechtenstein	181	4,083	38	48		
Norway	8,865	121,501	1063	1,453		
Total EU	267,316	2,666,127	137,896	172,829		
Total EEA	276,629	2,794,619	139,040	174,386		

Global		71,609

V. Empowering Users			
Commitment 20			
Relevant Signatories commit t	o empower users with tools to assess the provenance and edit history or authenticity o	r accuracy of digita	l content.
	C.20	M 20.1	M 20.2
We signed up to the following	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search	Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	<ul> <li>Microsoft has continued to improve content provenance measures on its AI image generation features, including Copilot in Bing.</li> <li>As part of Microsoft's commitments in the <u>Tech Accord to Combat</u></li> <li>Deceptive Use of AI in 2024</li> <li><u>Elections</u>, it is working to fulfill its commitments to:</li> </ul>
		<ul> <li>Developing and implementing technology to mitigate risks related to Deceptive AI Election</li> </ul>

		<ul> <li>content, including open- source tools where appropriate</li> <li>Assessing models in scope to understand the risks they may present regarding Deceptive Al Election Content</li> <li>Seeking to detect the distribution of this content on their platforms</li> <li>Seeking to appropriately address this content detected on their platforms</li> <li>Fostering cross-industry resilience to Deceptive Al Election Content</li> <li>Providing transparency to the public regarding how the company addresses it</li> <li>Continuing to engage with a diverse set of global civil society organizations, academics</li> <li>Supporting efforts to foster public awareness, media literacy, and all-of- society resilience.</li> </ul>
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No	Yes

If yes, which further implementation measures do you plan to put in place in the next 6	Planned Implementation Measures	Planned Implementation Measures
months?		
	Not applicable	Microsoft expects to continue its
		important work in content
		provenance tools and ways to help
		counter harmful AI-generated
		content.

Measure 20.1	LinkedIn	Bing Search
QRE 20.1.1	Microsoft is a founding and active member of the Coalition for	In July 2023, Microsoft agreed to make a number of voluntary commitments
	Content Provenance and Authenticity (C2PA) and is currently a co-	related to furthering safe and trustworthy AI systems, including a
	chair. In January 2022, the C2PA released version 1.0 of its technical	commitment to deploy new state-of-the-art provenance tools to help the
	specification for digital provenance.	public identify AI generated audio-visual content and understand its
		provenance. See more at Our commitments to advance safe, secure, and
	Microsoft is also partnering with Truepic to develop a prototype	trustworthy AI – Microsoft On the Issues.
	provenance mobile application, Project Providence. This prototype	
	application provides a proof of concept for safe, flexible, and easy to	By endorsing all of the voluntary commitments presented by U.S. President
	use capture, storage and viewing of authenticated visual media	Biden and independently committing to several others that support these
	anywhere in the world. Project Providence is currently in a proof-of-	critical goals, Microsoft is expanding its global, safe and responsible AI
	concept phase for limited use cases, including to support	practices, working alongside other industry leaders. Guided by the enduring
	documentation of cultural heritage destruction and recovery activities	principles of safety, security, and trust, the voluntary commitments address
	in Ukraine.	the risks presented by advanced AI models and promote the adoption of
		specific practices – such as red-team testing – that will propel the whole
	Further, Microsoft has partnered with NewsGuard to provide a free	ecosystem forward.
	plug-in for the Microsoft Edge web browser (also available for other	
	browsers including Chrome and Firefox), as well as an opt-in news	Microsoft and key members of the Bing Search team are also involved in the
	rating feature for the Edge mobile application on both iOS and	Partnership on AI ("PAI") to identify possible countermeasures against
	Android. This empowers Edge users to benefit from the	deepfakes and has participated in the drafting and refinement of PAI's
	comprehensive analysis done by NewsGuard and to better identify	proposed Synthetic Media Code of Conduct. The proposed Code of Conduct
	the most reliable news and information sites. LinkedIn users are also	provides guidelines for the ethical and responsible development, creation,
	able to benefit from NewsGuard via this plug-in. This plug-in enables	and sharing of synthetic media (such as AI-generated artwork).
	LinkedIn users to benefit from NewsGuard's reliability rating, where	

	available, when browsing news posts from news and information sites	Microsoft is also a founding and active member of the Coalition for Content
	rated by NewsGuard.	Provenance and Authenticity (C2PA) and is currently a co-chair. In January
		2022, the C2PA released version 1.0 of its technical specification for digital provenance.
		Microsoft is furthermore partnering with TruePic to develop a prototype provenance mobile application, Project Providence. This prototype application will provide a proof of concept for safe, flexible, and easy to use capture, storage and viewing of authenticated visual media anywhere in the world. Project Providence is an interoperable image authentication platform that collects images and video with authentication information embedded. This could assist any organization or person to document events, places, or objects anywhere in the world and provide proof that images were unaltered and authentic. Project Providence is currently in a proof-of-concept phase for limited use cases, including to support documentation of cultural heritage destruction and recovery activities in Ukraine.
		In addition, images generated by Copilot in Bing include a watermark in each image to indicate that the image was generated by AI as well as content provenance data that is embedded in image metadata.
Measure 20.2	LinkedIn	Bing Search
QRE 20.2.1	Microsoft and LinkedIn were two of 20 companies that announced a new Tech Accord to Combat Deceptive Use of AI in 2024 Elections, to combat video, audio, and images that fake or alter the appearance, voice, or actions of political candidates, election officials, and other	Microsoft is one of 20 companies that announced a new Tech Accord to Combat Deceptive Use of AI in 2024 Elections, to combat video, audio, and images that fake or alter the appearance, voice, or actions of political candidates, election officials, and other key election stakeholders.
	key election stakeholders. The Tech Accord's commitments will make it more difficult for bad	The Tech Accord's commitments will make it more difficult for bad actors to use legitimate tools to create deepfakes and easier for users to identify
	actors to use legitimate tools to create deepfakes and easier for users to identify authentic content. This focuses on the work of companies that generate AI content as well as those that distribute it and calls	authentic content. This focuses on the work of companies that generate AI content as well as those that distribute it and calls on them to strengthen the safety architecture in AI services by assessing risks and strengthening
	on them to strengthen the safety architecture in AI services by	controls to help prevent abuse. For its part Microsoft has already taken steps

assessing risks and strengthening controls to help prevent abuse. For	to meet the commitments in the Tech Accord by further implementing
its part Microsoft has already taken steps to meet the commitments	content provenance, establishment of reporting channels, and improved
in the Tech Accord by further implementing content provenance,	detection capability. For example, Microsoft launched a new web page –
establishment of reporting channels and improved detection	Microsoft-2024 Elections – where a political candidate can report a concern
capability. For example, Microsoft launched a new web page -	about a deepfake of themselves.
Microsoft-2024 Elections – where a political candidate can report a concern about a deepfake of themselves.	The accord's commitments oblige Microsoft and the tech sector to continue to engage with a diverse set of global civil society organizations, academics,
The accord's commitments oblige Microsoft and the tech sector to continue to engage with a diverse set of global civil society	and other subject matter experts. These groups and individuals play an indispensable role in the promotion and protection of the world's
organizations, academics, and other subject matter experts. These	democracies.
groups and individuals play an indispensable role in the promotion and protection of the world's democracies.	Microsoft is a founding member of the Coalition for Content Provenance and Authenticity (C2PA), and Bing Search leadership regularly participates in
Microsoft is a founding member of the Coalition for Content Provenance and Authenticity (C2PA). The C2PA Coalition aims to address the prevalence of disinformation, misinformation, and online	discussions related to the development of provenance tools. The C2PA Coalition aims to address the prevalence of disinformation, misinformation, and online content fraud through developing technical standards for
content fraud through developing technical standards for certifying the source and history or provenance of media content.	certifying the source and history or provenance of media content. The Copilot in Bing feature to create images includes C2PA Content Credentials, developed by the C2PA Coalition.
Standards are still in development, as member organisations work together to develop content provenance specifications for common asset types and formats to enable publishers, creators, and consumers to trace the origin and evolution of a piece of media, including images, videos, audio, and documents. These technical specifications include defining what information is associated with each type of asset, how that information is presented and stored, and	Microsoft has also developed Project Origin, which aims to create a measure of accountability for media to reduce the spread of disinformation through synthetic media. This technology helps provide evidence of the source of the content, like a watermark, which can help identify if alterations have been made.
how evidence of tampering can be identified.	<ul> <li>Project Origin gives publishers and consumers a tool to identify when media has been altered from the original source</li> <li>While it is ultimately up to viewers to believe in the accuracy or truth of a publisher, knowing if media has been tampered with gives us more power to determine what we can trust</li> <li>As part of the publication process, tools will register media items by creating a digital fingerprint</li> </ul>

In return, the content creator will receive a certification of
authentication, which will be stored in a tamper-proof distributed
ledger with no single controlling entity
The certification can be embedded into a piece of media before
distribution
When a user consumes a piece of media, the web browser or dedicated
application will automatically compare the embedded digital fingerprint of
the file being viewed with the original stored in the distributed ledger. Based
on that comparison, it will display a clear indicator of authentication,
allowing users to understand if what they are viewing is as published or has
been tampered with. See also QRE 20.1.1.

V. En	npowering	Users
-------	-----------	-------

Commitment 21

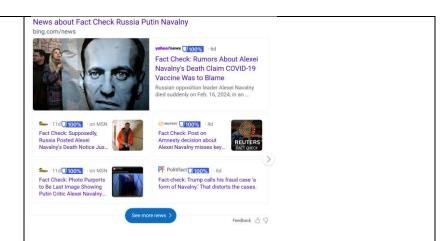
Relevant Signatories commit to strengthen their efforts to better equip users to identify Disinformation. In particular, in order to enable users to navigate services in an informed way, Relevant Signatories commit to facilitate, across all Member States languages in which their services are provided, user access to tools for assessing the factual accuracy of sources through fact-checks from fact-checking organisations that have flagged potential Disinformation, as well as warning labels from other authoritative sources.

	C.21	M 21.1	M 21.2	M 21.3
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search		Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures

	Not applicable	Bing has highlighted additional sources of fact checks included in its services and has been in active, productive conversations with fact- checking organizations to discuss possible means of expanding fact- checking coverage in the EU and the most impactful means of doing so. Bing hopes to have additional updates in its next report.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	Not applicable	Bing expects to finalize an additional agreement for enhanced fact-checking resources shortly and hopes to begin piloting those additional resources in the next Reporting Period. Bing continues to explore possibilities for agreements with additional fact checking organizations and has been in active dialogue with multiple fact checking organizations. Bing is also evaluating additional tools and resources that support the spirit of this commitment.

Measure 21.1	LinkedIn	Bing Search
QRE 21.1.1	As detailed above in QRE1.1.1 and 17.1.1, LinkedIn prohibits misinformation and disinformation on its platform, whether in the form of organic content or in the form of advertising content. LinkedIn's <u>Professional Community Policies</u> , which apply to all content on LinkedIn's platform, expressly prohibit the sharing of false or misleading content, including misinformation and disinformation.	Bing Search provides tools to help users determine the reliability of content displayed in Bing search results through its "Fact Check" feature. Bing Search's Fact Check feature offers users credible ways to assess the reliability of content displayed in its search results by providing fact-check flags and warnings on certain search results and by directing users to fact-checking articles.
	<ul> <li>Where content is identified as misinformation (whether as a result of a report or proactively detected), we do not label it, rather it is removed from LinkedIn. This includes situations where LinkedIn personnel leverage the conclusions of fact checkers to determine whether the content at issue violates LinkedIn's Professional Community Policies.</li> <li>Further, Microsoft has partnered with NewsGuard to provide a free plug-in for the Microsoft Edge web browser (also available for other browsers including Chrome and Firefox), as well as an opt-in news</li> </ul>	<u>ClaimReview</u> functions as a tagging system that individual fact-checking organizations or media can use to identify their articles for search engines and social media platforms. Since 2017, Bing Search has ingested content from the public ClaimReview protocol made available through Schema.org and displayed fact check tags within search results for websites that have incorporated ClaimReview tags. Content tagged with a ClaimReview tag is "read" by Bing Search's search engine crawlers when ingesting websites for their index and is used by Bing Search to help promote and/or highlight fact-checked content in search results. See additional information at <u>Bing</u> adds Fact Check label in SERP to support the ClaimReview markup
	rating feature for the Edge mobile application on both iOS and Android. This empowers Edge users to benefit from the comprehensive analysis done by NewsGuard and to better identify the most reliable news and information sites. LinkedIn users are also able to benefit from NewsGuard via this plug in. This plug-in enables LinkedIn users to benefit from NewsGuard's reliability rating, where available, when browsing news posts from news and information sites rated by NewsGuard. Please also see our response to QRE 17.1.1 which details how our	<ul> <li>Webmaster Blog. Bing Search also utilizes ClaimReview and fact check data feeds from the from DataCommons.org.</li> <li>Fact-checks and articles from news and fact checking organizations, including Deutsche Welle, AFP, AP, Verify, and Reuters may also appear as part of Bing Answers or similar features directed to users. In addition, news and fact-check articles can appear in Bing News verticals, which are often presented at the top of search results pages, depending on the user query, an example of which is below.</li> </ul>
	internal team of experienced news editors provides trustworthy news about current events from verified sources and other steps we take to tackle disinformation.	



Bing also leverages NewsGuard "Misinformation Fingerprints" comprised of datasets and tooling intended to highlight and track provably false narratives appearing online. Bing currently leverages Misinformation Fingerprints and localized reporting in English, Italian, Spanish, French, and German languages. In addition to global narratives, Misinformation Fingerprints cover EU markets including France, Germany, Spain, Italy, and Austria as well as Ukraine and the United Kingdom. Bing also integrates misinformation and state-sponsored disinformation risk briefings and new source ratings (including ratings for publications in the above jurisdictions) from NewsGuard that help inform defensive search strategies.

Since its last report, Microsoft continues to grow partnerships to strengthen the company's capacity and ability to combat information operations globally. Microsoft works with EFE to use their EFE Verifica to identify, track and expose emerging information operations spreading across Spanishspeaking communities around the world and Reporters Sans Frontières (RSF) to use their Journalism Trust Initiative (JTI) dataset to proactively promote trusted sources of news around the world that follow transparent and accepted journalistic practices. RSF's JTI data has expanded since our last report and currently includes Bulgaria, France, Greece, Germany, Latvia,

		Poland, and Romania. Microsoft Democracy Forward and RSF are working to expand the dataset into more countries throughout 2024.
		Bing's generative AI features are subject to Copilot in Bing's robust safety classifiers (which are informed by data shared by information integrity
		partners and other authority signals) to prevent the creation of low authority
		materials or disinformation. As Copilot in Bing prohibits the creation of
		misleading content, and Bing does not allow users to publish or share
		content created by the system on the platform, in-product fact checks are
		not relevant to Bing's generative AI features. Moreover, Copilot in Bing
		responses are grounded in web search results, and we provide links to
		websites so that users can learn more and evaluate the content for
		themselves.
		For more information on other media literacy tools in Bing Search, please
		see QRE 17.1.1.
SLI 21.1.1 - actions	Methodology of data measurement:	Methodology of data measurement:
taken under measure		
21.1	As detailed in our response to 21.1.1, LinkedIn removes, rather than	Bing Search uses the open schema.org ClaimReview protocol, and fact check
	labels, content that violates policy on <u>false and misleading content</u> .	articles incorporating ClaimReview tags are ingested in the Bing index. Bing cannot feasibly assess the location of the website or its operator for the
	Accordingly, the metrics for this SLI for the period 1 July - 31	purposes of this metric. Therefore, when calculating the location of fact
	December 2023 are zero.	checked content for the purposes of SLI, Bing has compiled the data as follows:
		<b>"FCURL"</b> represents the number of distinct URLs containing a ClaimReview
		tag (i.e. fact-check content) that appeared on the first page of Bing search
		results for any number of users located in the EU Member States and EEA.
		<b>"FCI"</b> represents number of times the above-mentioned URLs appeared on
		the first page of Bing search results to a user located in EU Member States
		and the EEA. Bing cannot provide data on the ratio of impressions of fact- checks to original impressions of fact-checked content, as Bing ingests

			1	cannot reas its search ir checks dep	th "fact check" tags conably quantify the ndex beyond what ends entirely on ea	e "reach" of labels is provided below, ch user's unique se	or fact checks a , as the appeara earch query.	vailable on ince of fact
	Total impressions of fact-checks	Ratio of impressions of fact-checks to original impressions of fact-checked content	Reach of labels/ fact- checkers and other authoritative sources	Other pertinent metric	Total impressions of fact-checks	Ratio of impressions of fact-checks to original impressions of fact-checked content	Reach of labels/ fact- checkers and other authoritative sources	Other pertinent metrics
_				(N/A)	(FCI)	(N/A)	(FCURLs)	(N/A)
Member States								
Austria					18,913		2,685	
Belgium					54,272		4,184	
Bulgaria					1		1	
Croatia					1		1	
Cyprus					1		1	
Czech Republic					1		1	
Denmark					5,147		1,057	
Estonia					1		1	
Finland					3,332		790	
France					170,314		5,079	
Germany					5,691,220		10,328	
Greece					2		2	
Hungary					1		1	
Ireland					17,508		3,156	
Italy					37,630		3,367	
Latvia					3		2	
Lithuania					1		1	
Luxembourg					1		1	

Malta		0	0
Netherlands		56,449	5,751
Poland		21,824	2,404
Portugal		20,624	2,427
Romania		10	8
Slovakia		1	1
Slovenia		0	0
Spain		135,093	5,251
Sweden		15,927	2,814
Iceland		0	0
Liechtenstein		0	0
Norway		8,710	1,960
Total EU		6,248,277	49,314
Total EEA		6,256,987	51,274

SLI 21.1.2 - actions taken	Methodology of data n	neasurement:		Methodology of data me	asurement:	
under measure 21.1						
	As detailed in response	to QRE 21.1.1, LinkedIn	removes, rather than	Metrics concerning "user	interactions" and "label	ling of content" are not
	labels, content that vi	olates our policy on <u>f</u>	alse and misleading	applicable to Bing, as it	is a search engine. Bing	g does not apply fact-
	content. Accordingly, th	he metrics for this SLI fo	or the period 1 July –	check labels to websites; i	rather websites must em	bed a ClaimReview tag
	31 December 2023 are	zero.		in order for Bing to inge	est and display fact-ch	eck tags for a website
				within search results. Bi	ing cannot quantify th	ne number of articles
				published by independer	nt fact-checkers that are	e available on the Bing
				Search index given the im	nmense scale of search	engine indexes. See SLI
				21.1.1.		
	Nr of articles	Nr of labels applied	Meaningful	Nr of articles published	Nr of labels applied	Meaningful metrics
	published by	to content, such as	metrics such as	by independent fact-	to content, such as	such as the impact of
	independent fact-	on the basis of such	the impact of 21.1.	checkers	on the basis of such	21.1. measures on
	checkers	articles	measures on user		articles	user interactions
			interactions with,			with, or user re-
			or user re-shares			shares of, content
			of, content fact-			

	checked as false		fact-checked as false
	or misleading		or misleading

Measure 21.2	LinkedIn	This Measure is not relevant or pertinent to Bing Search as it does not
QRE 21.2.1	LinkedIn has to date not undertaken and/or supported separate research and testing on the potential efficacy of warnings or updates targeted to users that have interacted with content that was later actioned upon for violation of our Professional Community Policies.	host the content and cannot assess the extent to which content or indexed landing pages has reached its users.
	Given LinkedIn currently removes, rather than labels, content that violates our policy on <u>false and misleading content</u> , LinkedIn may be unable to provide meaningful context to users as to the specific content that they had viewed which was later actioned.	
	To the extent others have conducted such research and/or testing, we are happy to discuss findings within the relevant Task-force Subgroups in view of identifying relevant follow-up actions.	
Measure 21.3	LinkedIn	Bing Search
QRE 21.3.1	As detailed in response to QRE 21.1.1, LinkedIn removes, rather than labels, content that violates our policy on <u>false and</u> <u>misleading content</u> .	Bing Search regularly consults research and evidence, including from internal Microsoft research and data science teams, related to safe design practices, labeling, and user experience and considers such research as part of its product design and testing. Bing Search also conducts internal research and user studies for product features, such as by analyzing impressions, engagement, or clicks of various features. Bing Search also has a "feedback" button easily accessible from any page of Bing. Bing Search reviews and may make improvements based on user feedback.

Bing Search also regularly consults with third party organizations to hear
feedback about product design and related safety considerations.
As to fact check labels, Bing Search participated in the W3C organization that helped to design and promote Schema.org and ClaimReview and regularly meets with stakeholders to discuss common issues, including whether updates to these common schemas are necessary.
As part of Copilot in Bing, Microsoft's Responsible AI team and product teams have worked to develop labeling and warning systems – as well as robust support and educational resources on AI generally - to help ensure users are informed that AI-powered answers are not always accurate and to encourage users to consult the source links provided. Microsoft's
Responsible AI team is staffed by a cross-disciplinary team of experts in AI, who consult regularly with external experts in the field to ensure our labels and warnings are designed in accordance with best practices.

	۷.	Empower	ring Users					
		Commitm	nent 22					
false or misleading, and	nmit to provide users with tools to help th I to facilitate user access to tools and ir ned online navigation, particularly relating	nformation to	o assess the t	rustworthiness o	of informa			
	C.22	M 22.1	M 22.2	M 22.3	M 22.4	M 22.5	M 22.6	M 22.7
We signed up to the	LinkedIn	LinkedIn	LinkedIn	LinkedIn				LinkedIn
following measures of this commitment:	Bing Search		Bing Search	Bing Search				Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	LinkedIn enhanced its "About this profile" feature	Microsoft decided to continue supporting Verifee with Azure Credits. Verifee is a browser extension and digital literacy program that uses AI to transparently rate the credibility of news articles.
Do you plan to put further implementation measures in place in the next 6 months to	No	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	Not applicable.	Bing regularly evaluates opportunities to improve its product and educate users on the trustworthiness and limitations of Al. In the next six months, Microsoft plans to continue implementing expanded data from RSF using their Journalism Trust Initiative data
		In February 2024, Microsoft launched several collaborations with news organizations to adopt generative AI, including Semafor, the GroundTruth Project, educational institutions, and others <u>.</u>

Measure 22.1	LinkedIn	This Measure is not relevant or pertinent to Bing Search as it does not allow
QRE 22.1.1	Microsoft has partnered with NewsGuard to provide a free plug-in for	users to share content. However, Bing will, where appropriate, provide users
	the Microsoft Edge web browser (also available for other browsers	with information on the availability of credibility extensions for the Microsoft
	including Chrome and Firefox), as well as an opt-in news rating feature	Edge browser, such as the extension currently offered by NewsGuard, and
	for the Edge mobile application on both iOS and Android. This empowers	provide, in appropriate cases, links to access those extensions. Bing wil
	Edge users to benefit from the comprehensive analysis done by	continue to explore whether and how other indicators of trustworthiness may
	NewsGuard and to better identify the most reliable news and	be useful to users in the context of search.
	information sites. LinkedIn users are also able to benefit from	
	NewsGuard via this plug in. This plug-in enables LinkedIn users to benefit	
	from NewsGuard's reliability rating, where available, when browsing	
	news posts from news and information sites rated by NewsGuard.	
	Further, in October 2022, LinkedIn began <u>offering</u> an "About this profile"	
	feature that shows users when a profile was created and last updated	
	along with whether the member has verified a phone number and/or	
	work email associated with their account. Over the past year, LinkedIr	
	also has been rolling out a range of free verifications, which allow our	
	members to verify certain information about themselves, like their	
	association with a particular company or educational institution or their	
	identity (through one of LinkedIn's verification partners).	
	The above features can be strong user empowerment tools. Specifically	
	they can provide our members valuable authenticity signals to help them	
	make more informed decisions about what content and individuals they	
	engage with online.	
SLI 22.1.1 - actions	Methodology of data measurement:	
enforcing policies		
above	The table below reports metrics concerning EEA member use of the	
	"About this profile" feature described above in QRE 22.1.1. The metrics	

	include: (1) the number of members feature between 1 July – 31 Decer number of times those members us December 2023. The metrics are assigned to EEA 1	mber 2023; and (2) the aggregate and the feature between 1 July - 31 Member State based on the self-
	reported profile location of the mem	ıber.
	Percentage of users that have	The aggregate number of times
	enabled the trustworthiness	those members used the feature
	indicator	between 1 July – 31 December 2023
	The number of members who used	
	the "About this profile" feature	
	between 1 July – 31 December	
	2023	
Member States		
Austria	140,404	288,080
Belgium	290,032	609,549
Bulgaria	48,989	120,924
Croatia	36,676	77,638
Cyprus	24,665	66,118
Czech Republic	104,539	223,326
Denmark	227,660	468,526
Estonia	19,567	48,829
Finland	98,411	192,298
France	1,945,584	4,369,452
Germany	1,255,147	2,757,240
Greece	114,169	254,883
Hungary	66,423	135,552
Ireland	172,092	383,330
Italy	882,345	1,725,304
Latvia	20,213	43,592

Lithuania	39,729	96,992	
Luxembourg	33,809	77,314	
Malta	17,218	41,775	
Netherlands	852,695	1,770,388	
Poland	372,432	837,818	
Portugal	250,751	514,691	
Romania	145,468	318,185	
Slovakia	31,245	70,173	
Slovenia	19,157	37,432	
Spain	855,389	1,829,190	
Sweden	318,367	651,386	
Iceland	4,883	9,995	
Liechtenstein	2,126	4,345	
Norway	114,655	221,373	
Total EU	8,383,176	18,009,985	
Total EEA	8,504,840	18,245,698	

Measure 22.2	LinkedIn	Bing Search
	LinkedIn does not prioritise any new sources in our feed, but in crisis	Bing Search utilizes a variety of signals – including trustworthiness indicators
	situations, (e.g., Covid-19 or Ukraine), we will use search banners to	from trusted third parties such as NewsGuard, GDI, EFE, Reporters Sans
	point members to reputable sources of information (e.g., when	Frontiers, and others – as one of several means to help determine the
	members searched for COVID, we pointed members to "trusted	authority score of a given website and rank it accordingly in search results.
storylines" where we provided trustworthy information about tho		
	topics, including links to global health organisations) (please also see	Bing Search also relies upon trustworthiness signals to help ensure that its
QRE 22.2.1	QRE 17.1.1).	recommender systems, such as auto-suggest and related search, direct users
		to high authority, trustworthy results and do not inadvertently suggest
	LinkedIn's focus, in addition to pointing members to trustworthy	problematic or misleading content.
	content, has been to prohibit members from sharing harmful content	
	on the platform. As a real identity online professional networking	Microsoft's NewsGuard partnership provides users with trustworthiness
	platform, content posted by members is seen by that member's	signals to better evaluate the quality of the news they encounter online.
	colleagues, employer, and potential business partners. Consequently,	NewsGuard has created trust ratings for 7,500+ news and information sites,

	which are compiled into a "Nutrition Label" and corresponding Red/Green
-	Reliability Rating to help users understand the reliability of news sources
for the same reasons. Nonetheless, where misinformation is removed	Within the EU, NewsGuard is currently available in France, Austria, Germany,
from LinkedIn, it is ineligible to be included in our recommender	and Italy with plans for future expansion. NewsGuard ratings and fingerprints
systems.	cover the news and information sources that account for 95%+ of online
	engagement with news in those countries. Microsoft offers NewsGuard as a
	free plug-in for the Microsoft Edge web browser (it is also available for othe
	browsers including Chrome and Firefox), and users of the Edge mobile
	application on both iOS and Android can enable NewsGuard ratings in their
	app settings. For users with the NewsGuard plug-in, Bing Search results (both
	in traditional web search and Copilot in Bing) include NewsGuard Reliability
	ratings that lead to a pop-up screen with more site information.
	Microsoft also maintains additional partnerships to strengthen the company'
	capacity and understanding of global threats to disinformation: first, with EFI
	Verifica (EFE), to track false narratives spreading in Spain and Latin America
	and second, with Reporters Sans Frontières (RSF), to use their Journalism Trus
	Initiative (JTI) data set in Microsoft products. These partnerships are part of a
	broader effort to empower Microsoft users to better understand the
	information they consume across our platforms and products. RSF's JTI data
	has expanded since our last report and currently includes Bulgaria, France
	Greece, Germany, Latvia, Poland, and Romania. Microsoft Democracy Forward
	and RSF are working to expand the dataset into more countries throughour
	2024.
	The above mechanisms and the Bing algorithm's emphasis on promoting high
	authority content are applied equally to the new Bing generative AI features
	to help ensure that users are protected from misleading information in the
	new Bing experience. See, for example, <u>How Microsoft Uses NewsGuard to</u>
	Support its Trusted, Transparent Technology.
	Support its Trusteu, Transparent Technology.

Measure 22.3	LinkedIn	Bing Search
	In addition to the LinkedIn <u>User Agreement</u> , LinkedIn has established and published (a) the LinkedIn <u>Professional Community Policies</u> to set out and elaborate on LinkedIn's requirements and expectations	The above-mentioned policies and measures are available to EU users in EU languages through the <u>Microsoft Bing Webmaster Guidelines</u> and <u>How Bing</u> <u>delivers search results.</u>
	for its member base; and (b) help center content that collates and	
	expands upon existing resources to further explain the main parameters of LinkedIn recommender systems and options provided	In addition, Copilot FAQs and support pages provide guidance to users on how to make informed decisions and evaluate content generated by AI
	to users to influence and control these recommender systems.	features. See e.g., <u>Unleash your productivity with AI and Microsoft Copilot –</u> <u>Microsoft Support</u> .
	The Professional Community Policies and help center content are	
		More information about how Microsoft helps prevent users from being
		presented with harmful information in Copilot in Bing is available at:
	languages: English, Czech, Danish, Dutch, French, German, Italian,	
	Norwegian, Polish, Portuguese, Romanian, Spanish and Swedish.	
	Additionally, we have extended this language coverage in accordance	
QRE 22.3.1	with the Digital Services Act.	
	LinkedIn seeks to reflect the best version of professional life through	
	a community where we treat each other with respect and help one another succeed.	
	The Professional Community Policies have three main elements: (1)	
	Be Safe, (2) Be Trustworthy and (3) Be Professional. Additionally, the	
	Professional Community Policies set out how members can report	
	content that may violate our policies and that a violation of our	
	Professional Community Policies can result in action taken against	
	that member's account or content.	
	(1) Be Safe: do not post harassing content; do not threaten,	
	incite, or promote violence; do not share material depicting	
	the exploitation of children; do not promote, sell or attempt	
	to purchase illegal or dangerous goods or services; do not	

	share content promoting dangerous organisations or				
	individuals.				
	(2) Be Trustworthy: do not share false or misleading content; do				
	not create a fake profile or falsify information about yourself;				
	do not scam, defraud, deceive others.				
	(3) Be Professional: do not be hateful, do not engage in sexual				
	innuendos or unwanted advances; do not share harmful or				
	shocking material; do not spam members or the platform.				
Measure 22.4	This measure is aimed specifically at "providers of trustworthiness	This measure is aimed specifically at "providers of trustworthiness			
QRE 22.4.1	indicators" and is therefore not relevant or pertinent to Microsoft	indicators" and is therefore not relevant or pertinent to Microsoft and its			
	and its services.	services.			
SLI 22.4.1 – actions					
enforcing policies above					

Measure 22.5	This measure is aimed specifically at "providers of trustworthiness	This measure is aimed specifically at "providers of trustworthiness
QRE 22.5.1	indicators" and is therefore not relevant or pertinent to Microsoft	indicators" and is therefore not relevant or pertinent to Microsoft and its
	and its services.	services.
SLI 22.5.1 – actions		
enforcing policies above		

SLI 22.5.2 – actions enforcing policies above		
	indicators" and is therefore not relevant or pertinent to Microsoft and its services.	

Measure 22.7	LinkedIn	Bing Search
QRE 22.7.1	adapt our systems and practices for combating misinformation and	In addition to broader measures described in this report, Bing Search has taken special care to counter low authority information and misinformation in relation to elections, the Russian invasion of Ukraine and the Israel-Hamas conflict, as detailed below and further in the Crisis Reporting appendices.
	agree to abide by on joining LinkedIn, prohibit misinformation. As described in more detail in our response to QRE 18.1.1, LinkedIn	Microsoft is also an active participant in the elections working group established by the Crisis Response Working Group. In addition, Microsoft works with election authorities responsible for running elections to promote trusted information regarding elections and monitors for foreign information operations targeting elections. Bing is also working to launch special features concerning upcoming 2024 European elections, such as info panels and specialized answers directing
	LinkedIn also aims to educate its members about civic discourse, electoral processes, and public security through its global team of news editors. These editors provide each member with relevant, timely information sourced from credible news sources that educate LinkedIn members about facts and news events, including global elections, and its content moderation teams closely monitor associated conversations in a number of languages.	examples for upcoming EU Parliament elections).
	In addition to broader measures, LinkedIn has taken special care to counter low authority information in relation to the COVID-19 crisis and the Russian Invasion of Ukraine, as detailed below and further in the Crisis Reporting appendices. Some examples of steps we have taken to tackle disinformation in connection with unfolding world events include the following:	Cómo votar en España elections.europa.eu
	<ul> <li>LinkedIn's team of editors covered the most recent developments in the Israel-Hamas war, including promoting updates from key sources to help members stay informed.</li> <li>In response to the COVID-19 pandemic, LinkedIn editors created and promoted trusted content. Further, during the</li> </ul>	Additional detail is provided in the Crisis Reporting appendix.

		• Bing Search has taken steps to algorithmically boost authority signals and has downgraded less authoritative information ( <i>see</i> SLI 22.7.1). These queries are translated automatically into other languages
	<ul> <li>Organization.</li> <li>LinkedIn's team of editors covered the most recent developments of Russia's invasion of Ukraine, ranging from the economic impact to major military events that are taking place.</li> </ul>	<ul> <li>supported by Bing Search and integrated into Copilot in Bing.</li> <li>Bing search works with Microsoft's Democracy Forward, Threat Assessment Center (MTAC) and Threat Intelligence Center (MSTIC) to ensure access to signals regarding Russian cyber and information operations targeting Ukraine to inform potential algorithmic interventions both for traditional and generative AI search tools.</li> <li>Bing Search regularly partners with independent research organizations and nonprofit organizations to maintain threat intelligence and inform potential algorithmic interventions both for traditional and generative AI search tools.</li> <li>Bing Search also takes action to remove autosuggest and related search terms that have been found likely to lead users to low authority content. Bing Search continues to partner with fact checking organizations and apply fact check labels to help users understand the context and trustworthiness of certain content in the EU, including an agreement with AFP covering Ukraine.</li> <li>These measures have helped ensure that Bing Search is promoting authoritative news sources, timelines, and other factual information at the top of algorithmic search results and in Copilot in Bing.</li> <li>Bing Search has also complied with EU sanctions orders requiring the</li> </ul>
		<ul> <li>Bing Search has also complied with EU sanctions orders requiring the removal of certain Russian sources, such as Russia Today and Sputnik.</li> </ul>
enforcing policies above	Methodology of data measurement:	Methodology of data measurement:
	<i>Ne</i> look forward to providing reports reflecting any applicable actions under this SLI during future reporting periods.	Bing has revised the SLI action descriptions below for accuracy with respect to the metrics provided pursuant to this Section. Please see below key.

			applied in every following data for e <b>"Defen</b> queries interver Period. e <b>"Impre</b> queries	Member State), Bing or this SLI: sive Interventions entered by users th ntions related to the ssions (RU/UA)" ref searched by users wi	has endeavored ( <b>RU/UA)</b> " refers lat were address Ukraine/Russia o flects the numbe nere defensive se	obal level (and therefore to provide the additional s to the total number of ed with defensive search risis during the Reporting er of user impressions for earch interventions related ing the Reporting Period.		
	Impressions	Clicks	CTR	Shares	Defensive	Impressions	CTR	Shares
					Interventions	(RU/UA)		
Member States					(RU/UA)		(N/A)	(N/A)
Austria					3,419	43,690		
					4,855	56,789		
Belgium								
Bulgaria					13	15		
Croatia					3	3		
Cyprus					2	2		
Czech Republic					12	14		
Denmark					1,580	13,148		
Estonia					6	6		
Finland					1,473	9,156		
France					11,773	247,458		
Germany					14,174	674,440		
Greece					6	15		
Hungary					3	3		
Ireland					3,332	39,438		
Italy					6,681	186,700		
Latvia					7	10		
Lithuania					11	15		
Luxembourg					1	1		

Malta	2	2	
Netherlands	6,057	53,755	
Poland	5,777	94,040	
Portugal	3,274	26,179	
Romania	11	16	
Slovakia	4	4	
Slovenia	2	2	
Spain	10,979	941,947	
Sweden	4,316	37,747	
Iceland	0	0	
Liechtenstein	0	0	
Norway	2,716	29,006	
Total EU	77,773	2,424,595	
Total EEA	80,489	2,453,601	

	V. Empowering users					
	Commitment 23					
Relevant Signatories commit to	Relevant Signatories commit to provide users with the functionality to flag harmful false and/or misleading information that violates Signatories policies or terms of service.					
	C.23 M 23.1 M 23.2					
We signed up to the following	LinkedIn	LinkedIn	LinkedIn			
measures of this commitment:	Bing Search	Bing Search	Bing Search			

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		

If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	Bing has updated its "Report a
		Concern" and "Feedback" tools
		regularly, and Copilot in Bing has
		distinct reporting options.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6	Planned Implementation Measures	Planned Implementation Measures
months?		
	As part of our work to implement	Bing monitors user feedback and
	the Digital Services Act, we are	regularly evolves its product
	working to introduce onboarding	reporting process and forms in
	flows for entities designated as	response to user feedback, new
	Trusted Flaggers .	legal obligations, or product
		developments.

Measure 23.1	LinkedIn	Bing Search
QRE 23.1.1	If LinkedIn users locate content they believe violates our Professional	As a search engine that does not host or display user generated content,
	Community Policies, we encourage them to report it using the in-	Bing Search does not have a reporting function for user generated content.
	product reporting mechanism represented by the three dots in the	
	upper right-hand corner of the content itself on LinkedIn.	Bing Search's Report a Concern Form permits users to report third-party
		websites for a variety of reasons including disclosure of private information,
	Misinformation is specifically called out as one of the reporting	spam and malicious pages, and illegal materials.
	options.	
		Bing Search's "Feedback" tool, which is accessible on the lower right corner
	The reporting feature is available through, and largely identical	on a search results page, allows users to provide feedback on search results
	across, LinkedIn's website and mobile app, although reporting	(including a screenshot of the results page) to Bing Search. Depending on
	reasons and their visual presentation may vary slightly for certain	the nature of the feedback, Bing Search may take appropriate action, such
	types of content. In most instances, the reporting process is located	as to engage in algorithmic interventions to ensure high authority content
	just one click away from the content being reported and, depending	appears above low authority content in search results, remove links that

QRE 23.2.1	As noted in our response to QRE 23.1.1, content that is flagged as misinformation (whether reported or automatically detected) is removed from LinkedIn. LinkedIn has a quality assurance team that is dedicated to ensure the quality of our content review processes and decisions. For example, the quality assurance team performs quality checks, on a routine basis, the content moderation decisions that have previously been made. This also allows us to improve our processes and further strengthen our platform as a trusted source of information.	See QRE 23.1.1. Bing Search generally does not experience issues with mass flagging of content or abuse of its reporting features. This concern appears more applicable to other types of services (e.g., social media and online media websites) or content outside the scope of this regulation that is more prone to mass flagging, such as copyright infringement. Bing Search engages in human review of reports submitted through its reporting functionality and evaluates each report consistent with its policies and procedures
Measure 23.2	LinkedIn	Bing Search
	Members also receive an email notifying them in in the event their content actioned in accordance with our policies. The email includes a link to a notice page for additional details and resources. If the member believes that their content complies with our Professional Community Policies, they can ask us to revisit our decision by submitting an appeal by clicking on the link in the notice page.	
	When members use the above reporting process, they will receive an email acknowledging receipt of the report. The email includes a link to the report status page, which we update when we make a decision, including providing the opportunity to appeal. Logged-out users receive updates on their report by email and are also provided with the opportunity to appeal.	
	Reported content generally is reviewed by trained content reviewers. In addition, LinkedIn uses automation to flag potentially violative content to our content moderation teams. If reported or flagged content violates the Professional Community Policies, it will be actioned in accordance with our policies.	As discussed in QRE 14.1.2, these tools have also been updated to make it easy for users to report problematic content they encounter while using Copilot in Bing by including the same "Feedback" button with direct links to the respective service's "Report a Concern" tool on the footer of each page of Copilot.
	on whether content is reported in the LinkedIn App or on desktop, the reporting process takes between four or five clicks to complete.	violate local law or Bing policies, add answers, warnings or other media literacy interventions on certain topics, or remove autosuggest terms.

Furthermore, as a real identity professional network, LinkedIn acts
vigilantly to maintain the integrity of all accounts and to ward off bot
and false account activity. LinkedIn enforces the policies in its User
Agreement prohibiting the use of "bots or other automated methods
to access the Services, add or download contacts, send or redirect
messages" through:
Having a dedicated Anti-Abuse team to create the tools to
enforce this prohibition
Using automated systems detect and block
automated activity
Imposing hard limits on certain categories of activity
commonly engaged in by bad actors
Detecting whether members have installed
known prohibited automation software
Conducting manual investigation and restriction of
accounts engaged in automated activity
Partnering with the broader Microsoft organisation to
develop technological solutions for detecting "deep
fakes"
Investing in and using AI to detect coordinated inauthentic
activity and communities of fake accounts through
similarities in their content and behaviour.

#### V. Empowering users

#### Commitment 24

Relevant Signatories commit to inform users whose content or accounts has been subject to enforcement actions (content/accounts labelled, demoted or otherwise enforced on) taken on the basis of violation of policies relevant to this section (as outlined in Measure 18.2), and provide them with the possibility to appeal against the enforcement action at issue and to handle complaints in a timely, diligent, transparent, and objective manner and to reverse the action without undue delay where the complaint is deemed to be founded.

	C.24	M 24.1
We signed up to the	LinkedIn	LinkedIn
following measures of		
this commitment:		

	LinkedIn	Commitment 24 is not relevant or pertinent to Bing Search as it does
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	not allow the posting of user
your terms of service, new tools, new policies, etc)? [Yes/No]		content that would require user
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	appeals.
	Not applicable	
Do you plan to put further implementation measures in place in the next 6 months to	No	
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	
	Not applicable	

Measure 24.1	LinkedIn	Measure 24.1 is not relevant or pertinent to Bing Search as it does not
QRE 24.1.1	<ul> <li>When a post, comment, reply, or article, is reported and found to go against our Professional Community Policies, we take appropriate actions to remove it and/or restrict accounts depending on the severity of violation.</li> <li>The author whose content we've actioned will typically receive an email notifying them that their content does not comply with our policies. If the author believes LinkedIn has made a mistake in actioning their content, the author can ask LinkedIn to take a second look by submitting an appeal. To begin that appeal process, members can log into their account and follow the onscreen messaging or reply to the message they received notifying them of the content removal.</li> <li>Appeals made by members are treated the same regardless of whether they use LinkedIn's premium services. After our review is completed, the author will receive one of the following status updates: <ul> <li>If your post complies with our Professional Community Policies, we'll put it back on LinkedIn.</li> <li>If we find that your post doesn't comply with our Professional Community Policies, it will be removed from LinkedIn and only you will be able to access your post.</li> </ul> </li> </ul>	allow the posting of user content that would require user appeals
SLI 24.1.1 - enforcement actions Data	<ul> <li>Methodology of data measurement (where appropriate insights on data provided)</li> <li>The table below reports metrics concerning content LinkedIn removed from its platform as Misinformation, pursuant to the policy outlined in QRE 18.2.1.</li> <li>The metrics include: <ul> <li>(1) the number of pieces of content removed as Misinformation between 1 July - 31 December 2023, broken out by EEA Member State;</li> <li>(2) the number of those content removals that were appealed by the content author;</li> </ul> </li> </ul>	

			appeals that were g	
			appeal to appeal o	
	appeals. The metrics are assigned to EEA Member State based on the self-reported profile location of the content author.			
	The number of pieces of content removed as Misinformation between 1 July - 31 December 2023	-reported profile The number of removals that were appealed by the content author	The number of appeals that were granted	The median time from appeal-to- appeal decision in hours
Member States		-	-	
Austria	255	3	2	8.5 hours
Belgium	652	8	0	_
Bulgaria	110	1	0	_
Croatia	101	3	0	_
Cyprus	38	2	0	_
Czech Republic	113	1	0	_
Denmark	305	5	0	_
Estonia	14	0	0	-
Finland	52	0 24		-
France Germany	4,796	24	0	-
Greece	2,219	1	0	-
Hungary	75	1	0	-
Ireland	192	3	0	-
Italy	3,627	12	0	-
Latvia	19	0	0	-
Lithuania	27	0	0	-
Luxembourg	85	0	0	-

Malta	24	0	0
Netherlands	4,137	50	3
Poland	280	4	0
Portugal	344	4	0
Romania	172	0	0
Slovakia	24	1	0
Slovenia	21	0	0
Spain	2,953	11	1
Sweden	286	2	0
Iceland	9	0	0
Liechtenstein	1	0	0
Norway	66	0	0
Total EU	21,130	164	8
Total EEA	21,206	164	8

	V. Empowering users				
	Commitment 25				
In order to help users of private messaging services to identify possible disinformation disseminated through such services, Relevant Signatories that provide messaging applications commit to continue to build and implement features or initiatives that empower users to think critically about information they receive and help them to determine whether it is accurate, without any weakening of encryption and with due regard to the protection of privacy.					
C.25 M 25.1 M 25.2					
We signed up to the following       Commitment 25 is not relevant to Bing Search and LinkedIn.       N/A       N/A         measures of this commitment:       N/A       N/A       N/A					

## VI. Empowering the research community Commitments 26 - 29

### **VI. Empowering the Research Community**

Commitment 26

Relevant Signatories commit to provide access, wherever safe and practicable, to continuous, real-time or near real-time, searchable stable access to nonpersonal data and anonymised, aggregated, or manifestly-made public data for research purposes on Disinformation through automated means such as APIs or other open and accessible technical solutions allowing the analysis of said data.

	C.26	M 26.1	M 26.2	M 26.3
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search	Bing Search	Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Beta program to allow bonafide	Bing is working to provide
	researchers to access public data	additional open datasets and
	for research on impact of	resources that may be used by the
	misinformation and other online	research community, and Microsoft
	harms impacting the Union.	continues to provide open access
		to a range of research publications,
		tools, and data.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6	Planned Implementation Measures	Planned Implementation Measures
months?		
	Continued to review Beta program	Bing is actively exploring additional
	and researcher feedback for	mechanisms to meet this

improvements of program to better	commitment and welcomes
enable clarity and address obstacles	feedback from the research
to rightful access to public data.	community and Commission on the
	types of data that would be most
Update our program in light of	useful to the research community.
regulatory guidance that may be	
provided pursuant to Art. 40 of the	
DSA.	

Measure 26.1	LinkedIn	Bing Search
QRE 26.1.1	LinkedIn supports the aims of the research community and regularly	Bing Search and Microsoft are dedicated to supporting the research community and
	provides information and data to the research community in a	regularly provide information and data to the research community in a variety of ways.
	variety of ways.	
		Bing Search already provides researchers and the public with access to $\underline{MS}$ MARCO, a
	To date, we have made non-personal, aggregated data publicly	collection of datasets focused on deep learning in search that are derived from Bing
	available (data on gender equity in the workplace, data on green	Search queries and related data. Research organizations can gain access to the MS
	skills and jobs, data on industry and job skills, and data on	MARCO datasets instantaneously via the MS MARCO homepage. The MS MARCO
	engagement with labor markets and employment trends). Our goal	dataset has been cited in over 1400 research papers since its release and has been
	with this action to enable researchers to understand the rapidly	utilized for a range of research issues, including in connection with misinformation and
	changing world of work through access to and use of LinkedIn data.	disinformation related issues. Because the dataset is provided open source, the extent
	Because much of our data is publicly available, the extent to which	to which it has been used for disinformation related research purposes cannot easily
	such data has been used for disinformation related research	be ascertained. However, the dataset has been relied upon and/or cited in various
	purposes cannot easily be ascertained.	academic papers concerning misinformation and fact-checking, including the
		following research released during the Reporting Period:
	Additionally, LinkedIn is expanding its <u>API access</u> for public data for	
	disinformation related research purposes. Information about the	<ul> <li>"<u>Retrieving Supporting Evidence for Generative Question Answering</u>", SIGIR-</li> </ul>
	LinkedIn APIs are available to the public and researcher access is	AP '23: Proceedings of the Annual International ACM SIGIR Conference on
	provided here.	Research and Development in Information Retrieval in the Asia Pacific Region,
		November 2023.
	Finally, Microsoft is also a leader in research in Responsible AI and	
	provides <u>a range of tools and resources</u> dedicated to promoting	

responsible usage of artificial intelligence to allow practitioners and researchers to maximize the benefits of AI systems while mitigating harms. For example, as part of its <u>Responsible AI Toolbox</u>, Microsoft provides a Responsible AI Mitigations Library, which enables practitioners to more easily experiment with different techniques for addressing failure (which could include inaccurate outputs), and the Responsible AI Tracker, which uses visualizations to show the effectiveness of the different techniques for more informed decision-making. These tools are available to the public and research community for free.

- "<u>Cross-Genre Retrieval for Information Integrity: A COVID-19 Case Study</u>", In: Yang, X., *et al.* Advanced Data Mining and Applications. ADMA 2023. Lecture Notes in Computer Science), vol 14180. Springer, Cham, November 2023.
- "<u>Personas as a Way to Model Truthfulness in Language Models</u>" New York University, ETH Zurich, et al. <u>arXiv:2310.18168</u>, October 2023.

Bing Search also provides researchers with access to <u>ORCAS: Open Resource for Click</u> <u>Analysis in Search | msmarco (microsoft.github.io)</u>, a click-based dataset associated with the <u>TREC Deep Learning Track</u>, which provides 18 million connections to 10 million distinct queries and is available to researchers.

In 2020, Bing Search also shared a search dataset for Coronavirus Intent comprised of queries from all over the world that had an intent related to the Coronavirus or Covid-19 (e.g., searches for "Coronavirus updates Seattle" or "Shelter in place") for use by researchers and the public. This data, which is divisible by country, is particularly relevant to misinformation research on public health issues and the COVID-19 pandemic, as it provides insights into how users sought information related to the coronavirus during the pandemic. The dataset was also posted to Azure Open datasets for Machine Learning, Tensorflow.org and Kaggle. See additional information on the dataset at Extracting Covid-19 insights from Bing Search data | Bing Search Blog. This dataset has already been used by the disinformation research community. For example, Researchers at Aalborg University used the dataset as part of their research paper "Does Vinegar Kill Coronavirus?' - Using Search Log Analysis to Estimate the Extent of COVID-19-Related Misinformation Searching Behaviour in the United States ". The Aalborg University study used Bing search log analysis to investigate the extent and characteristics of misinformation seeking behaviour in the US. Harvard University's Kennedy School (HKS) Misinformation Review published the peer reviewed work "Less reliable media drive interest in anti-vaccine information" leveraging this dataset and Microsoft Research to study anti-vaccine discourse and patterns of misinformation.

Additionally, researchers who are registered webmasters may utilize Bing Search's <u>Keyword Tools</u> and <u>Backlinks</u> Webmaster Tools to provide insights into search usage

and keywords. Bing is also working on ways to provide deeper research access to the tool across the research community and hopes to provide updates in its next report.
Bing Search also offers use of <u>Bing APIs</u> to the public, which include Bing Image Search,
Bing News Search, Bing Video Search, Bing Visual Search, Bing Web Search, Bing Entity
Search, Bing Autosuggest, and Bing Spell Check. Bing Search provides free access to
these APIs for up to 1,000 transactions per month, which may be leveraged by the
research community.
In addition to the above datasets, Microsoft Research maintains a public portal of
codes, APIs, software development kits, and datasets that are available to the Research
Community at Researcher tools: code & datasets - Microsoft Research. These public
research tools can be accessed by researchers and downloaded instantaneously
without formal applications or login credentials.
Bing recently formalized its <u>Qualified Researcher Program</u> to enable EU researchers to
easily request access for publicly accessible Bing data from a singular landing
page. However, because these datasets are already available open-source (see below),
we expect some researchers may elect to obtain datasets via the above means to avoid
the burden of an application and credentialing process.
Microsoft is also a leader in research in Responsible AI and provides a range of tools
and resources dedicated to promoting responsible usage of artificial intelligence to
allow practitioners and researchers to maximize the benefits of AI systems while
mitigating harms. For example, as part of its Responsible Al Toolbox, Microsoft
provides a Responsible AI Mitigations Library, which enables practitioners to more
easily experiment with different techniques for addressing failure (which could include
inaccurate outputs), and the Responsible AI Tracker, which uses visualizations to show
the effectiveness of the different techniques for more informed decision-making.
These tools are available to the public and research community for free.
Lastly, given the open nature of the Bing Search index and public nature of search
results, researchers can utilize Bing Search or Copilot in Bing to run specific queries

				nedia which may require private accounts or	
QRE 26.1.2	LinkedIn <u>published</u> information on Program researcher access to public		connections between users to access certain materials). Bing Search will publish information as it continues to build further data research infrastructure pertinent to these commitments.		
SLI 26.1.1 - uptake of the	Methodology of data measurement: LinkedIn will publish information as research program pertinent to these	; it continues to build its data	Methodology of data measurement: Because the above-mentioned tools discussed in QRE 26.1.2 predate the Code of Practice and were provided open source without tracking mechanisms, Microsoft is working on developing adequate usage tracking for its many publicly accessible researcher tools and datasets.		
tools and processes described in Measure 26.1	Nr of users of public access 4 applications with no approvals under our Beta Art 40 process. Note: Unknown number of researchers who use our broadly available service to conduct research (e.g. Trustlab)	Other quantitative information on public access	Nr of users of public access	Other quantitative information on public access	
Data					
Measure 26.2	LinkedIn		Bing Search		
QRE 26.2.1	See QRE 26.1.1		Unlike social media platforms, Bing Search does not have private user account other personal data provided by users as contemplated by Measure 26.2. How Bing does enable researcher access to data on the platform through a num mechanisms, as described in QRE 26.1.1 and the research partnerships descril QRE 18.3.1. Researchers can also submit real-time queries in Bing Search and Co Bing Search welcomes feedback on the types of real-time data that would be useful to the misinformation and disinformation research community in light above considerations; noting however that providing access to an entire "real search engine dataset would be too large to feasibly process or share.		

QRE 26.2.2	See QRE 26.1.1	See QRE 26.2.1 and 26.1.1.
	Access to Data: For access to LinkedIn APIs, a researcher needs to	Currently, there is not an application process to access the MS MARCO, ORCAS, or
	submit an application, meet criteria for approval and provide	Bing Coronavirus Query datasets in their original download locations, as Microsoft
	additional information necessary for us to assess their project. APIs	intended to allow open and easy access to the public and research community without
	including non-public data may be made available for research	arduous credentialing or account creation processes. Users may freely access the
	purposes based on special requests and the ability of the researcher	datasets instantaneously through the <u>MS_MARCO_</u> and <u>ORCAS</u> websites and Bing
	to protect personal data pursuant to GDPR and relevant intellectual	Coronavirus Query page on <u>Github</u> . No application or credentialing is required, but
	property rights. Upon approval, the researcher's application with be	unfortunately this open source model makes tracking usage more challenging and
	provisioned with the relevant APIs. In addition, access is available	requires investment in additional tooling.
	to anyone who visits the relevant LinkedIn site.	
		For Bing Search's Keyword Research tool, Researchers may add their website (new or
	For access to other data, researchers may be provided with datasets	existing) to Bing Search Webmaster Tools and can freely use <u>Keyword Research – Bing</u>
	and information as part of research inquiries and research	Webmaster Tools or Backlinks – Bing Webmaster Tools. Bing Search APIs are accessible
	partnerships with LinkedIn. Researchers may contact LinkedIn to	at <u>Bing APIs</u> and may be accessed by signing up for an account or requesting access
	discuss research opportunities.	through the Qualified Researcher Program discussed below.
		For the <b>Bing Qualified Researcher Program</b> , eligible EU researchers may request access
QRE 26.2.3		for publicly accessible Bing data and APIs, including the resources mentioned above,
		through an application form. If their request meets the criteria highlighted on the
		application page, data will be made available for the approved research purpose. More
		details are available at Bing Qualified Researcher Program - Microsoft Support.
		Other Microsoft datasets, including datasets derived by Bing Search, are accessible at
		Researcher tools: code & datasets - Microsoft Research. These also do not require
		credentials and datasets may be freely downloaded by researchers.
		For other research data, researchers may be provided datasets and information as part
		of research partnerships with Microsoft Research. Researchers may contact Microsoft
		Research to discuss research opportunities.
		And of course, the Bing Search service, including Copilot in Bing, is also public and
		may be used for a variety of research purposes without login or credentials.

							Microsoft is continuing to explore possibilities to streamline data access consistent	
							with this provision and in accordance with Microsoft Research's longstanding data	
							sharing and collaboration with the research community. Bing is also in the process of	
						exploring additional research partnerships specifically related to Copilot and exploring		
						additional red team testing of Copilot within Microsoft.		
	Methodolo	ogy of data r	neasurement:				Methodology of data measurement:	
SLI 26.2.1 -	LinkodIn	hac provide	ad motrics	relating to t	ha numbr	or of	Bing does not gate access for most of its datasets and therefore this metric is	
		•		2				
meaningful			a Researcher	Access Progra	m in the p	erioa	inapplicable, as any individual may freely access the tools. Bing is working on	
metrics on	covered by	/ this report.					approaches for better tracking of usage of publicly released datasets and APIs.	
the uptake,								
swiftness,							Bing tracks applications to its Qualified Researcher Program and will provide additional	
and							reporting in its next report.	
acceptance								
level of the							"MSMARCO" under "Other Metrics" provides the total global number of download	
tools and			1	1			of the MS Marco database	
processes in	No of	No of	No of	No of	Averag	Oth	Other metrics	
Measure 26.2	monthly	applicati	applicatio	application	е	er		
	users	ons	ns	s accepted	respons	met	MSMARCO	
		received	rejected		e time	rics		
Data		4	4	0			6,617	
Measure 26.3	LinkedIn						Bing Search	
QRE 26.3.1	LinkedIn p	rovides a co	mprehensive	<u>Help Center</u> fo	r assistance	e with	Users can report issues accessing MS MARCO and ORCAS datasets to $\underline{ms}$ -	
	other matt	ers. LinkedIn	endeavors to	restore access	and addres	ss any	marco@microsoft.com. Microsoft endeavors to restore access and address any issues	
	issues expeditiously.			with dataset access expeditiously.				
							For Bing Search's Keyword Research tool, users can report issues by creating a support	
							request ticket at <u>Webmaster Support - Bing Webmaster Tools.</u>	
							request ticket at webmaster support - bing webmaster roois.	
							Users may report issues with Bing API access at https://aka.ms/bingapisupport.	

#### **VI. Empowering the research community**

Commitment 27

Relevant Signatories commit to provide vetted researchers with access to data necessary to undertake research on Disinformation by developing, funding, and cooperating with an independent, third-party body that can vet researchers and research proposals.

	C.27	M 27.1	M 27.2	M 27.3	M 27.4
We signed up to the	LinkedIn	LinkedIn	LinkedIn	LinkedIn	LinkedIn
following measures of	Bing Search				
this commitment:					

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Microsoft, is an active participant in	Microsoft is an active participant in
	the EDMO Working Group for the	the EDMO Working Group for the
	Creation of an Independent	Creation of an Independent
	Intermediary Body to Support	Intermediary Body to Support
	Research on Digital Platforms.	Research on Digital Platforms.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	Microsoft plans to continue to	Microsoft plans to continue to
	engage in the work of above-	engage in the work of above-
	mentioned EDMO Working Group.	mentioned EDMO Working Group.

Measure 27.1	LinkedIn	Bing Search
QRE 27.1.1	Microsoft is a member of the Working Group for the Creation of an	Microsoft is a member of the Working Group for the Creation of an
	Independent Intermediary Body to Support Research on Digital	Independent Intermediary Body to Support Research on Digital Platforms.
	Platforms. The Working Group has started its work on 10 May 2023	The Working Group has started its work on 10 May 2023 under the
	under the coordination of the European Digital Media Observatory	coordination of the European Digital Media Observatory (EDMO). Its main
	(EDMO). Its main task is to develop an organizational model for a new	task is to develop an organizational model for a new independent
	independent intermediary body that will facilitate data sharing	intermediary body that will facilitate data sharing between digital platforms
	between digital platforms and independent, external researchers. The	and independent, external researchers. The Working Group plans to identify
	Working Group plans to identify a set of governing principles for the	a set of governing principles for the new intermediary body, lay out its core
	new intermediary body, lay out its core functions, outline an	functions, outline an organizational structure, staffing and budgetary needs,
	organizational structure, staffing and budgetary needs, identify a form	identify a form and place of establishment, and provide a timeline for the
	and place of establishment, and provide a timeline for the body's	body's initial phases of work. The practical outcome of the Working Group
	initial phases of work. The practical outcome of the Working Group is	is a report laying out a roadmap, guidance on timeline and roles, as well as
	a report laying out a roadmap, guidance on timeline and roles, as well	other plans for establishing the independent, third-party body.
	as other plans for establishing the independent, third-party body.	
Measure 27.2	LinkedIn	Bing Search
	As the development of the independent third-party body has not yet	As the development of the independent third-party body has not yet been
QRE 27.2.1	been finalized, there has not yet been any funding allocated to the	finalized, there has not yet been any funding allocated to the
	implementation of Measure 27.2.	implementation of Measure 27.2.
Measure 27.3	LinkedIn	Bing Search
	As the development of the independent third-party body has not yet	As the development of the independent third-party body has not yet been
QRE 27.3.1	been finalized, no data has been shared with this body for the	finalized, no data has been shared with this body for the purposes of
	purposes of research as outlined under Measure 27.3.	research as outlined under Measure 27.3.
	Methodology of data measurement:	Methodology of data measurement:
SLI 27.3.1 – research	As the development of the independent third-party body has not yet	As the development of the independent third-party body has not yet been
projects vetted by the	been finalized, no research projects have yet been vetted by this body,	finalized, no research projects have yet been vetted by this body, as set out
independent third-	as set out under Measure 27.3.	under Measure 27.3.
party body		
	Nr of research projects for which they provided access to data	Nr of research projects for which they provided access to data

	Not applicable.	Not applicable.
Data		
Measure 27.4	LinkedIn	Bing Search
QRE 27.4.1	While the vetted researcher process is still being formalized, Microsoft continues to engage in ongoing research partnerships outside the formal process. For example, Microsoft is partnering with Princeton University and the Carnegie Endowment for International Peace to fund and provide data to the Institute for Research on the Information Environment (IRIE), an international resource to study information ecosystems that can spur evidence-based policy solutions. Microsoft will continue investing in its data-driven partnership with Princeton until at least mid-2024 to continue bolstering this resource. This project includes recommendations and best practices based on academic research for preparing for generative Al's role in 2024 elections.	While the vetted researcher process is still being formalized, Microsoft continues to engage in ongoing research partnerships outside the formal process. For example, Microsoft is partnering with Princeton University and the Carnegie Endowment for International Peace to fund and provide data to the Institute for Research on the Information Environment (IRIE), an international resource to study information ecosystems that can spur evidence-based policy solutions. Microsoft will continue investing in its data-driven partnership with Princeton until at least mid-2024 to continue bolstering this resource. This project includes recommendations and best practices based on academic research for preparing for generative AI's role in 2024 elections.

VI. Empowering the research community							
	Commitment 28						
Releva	Relevant Signatories commit to support good faith research into Disinformation that involves their services.						
	C.28	M 28.1	M 28.2	M 28.3	M 28.4		
We signed up to the following LinkedIn LinkedIn LinkedIn LinkedIn LinkedIn LinkedIn							
measures of this commitment:	Bing Search	Bing Search	Bing Search	Bing Search	Bing Search		

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	Yes	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		

If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Published details on our [Beta] <u>Researcher Access Program</u>	Bing and Microsoft have initiated new research projects and undertaken internal research on misinformation risks in image generation services, generative AI in elections, and other important topics.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures         Ongoing review of researcher         feedback and needs may result in         additional measures and resources         being made available.         Update our program in light of         regulatory guidance that may be         provided pursuant to Art. 40 of the         DSA.	Planned Implementation MeasuresBing recently formalized its Bing Qualified Research program.Microsoft Research regularly explores potential partnerships with third party research institutions and is actively in discussions with several research institutions on potential misinformation and disinformation related research that may leverage Bing Search data. Microsoft's internal research divisions also regularly initiate and support research relevant to misinformation and disinformation and further research may be released in the next reporting period.

	Microsoft Research will share
	findings on internal research
	concerning information integrity
	and elections in the age of
	generative AI. Findings were
	presented as part of Microsoft's
	Research forum on March 7, 2024
	and will be published publicly soon.

Measure 28.1	LinkedIn	Bing Search
QRE 28.1.1	LinkedIn facilitates research, engages with the research community,	Bing Search facilitates research, engages with the research community, and
	and provides data to the research community in a variety of ways, as	provides data to the research community in a variety of ways, as described
	described below and in QRE 26.1-2.	below and in QRE 26.1-2 and 18.3.1.
	Historically, LinkedIn's work with external stakeholders, including, for	More broadly, Microsoft dedicates significant resources to supporting,
	example, research institutes, and academia, to understand the	promoting, and developing research on emerging issues including
	rapidly changing world of work through access to and use of	responsible AI, safe design, search and information retrieval, language
	LinkedIn data. Additionally, LinkedIn employs academics to gain	learning models, and algorithms. Microsoft Research and other research
	practical experience combining industry knowledge with academic	groups within the company, such as the <u>AI for Good Research Lab</u> , employ
	expertise to solve complex business problems spanning all areas of	robust teams of researchers and data scientists and regularly utilize Bing
	engineering, with an initial focus on artificial intelligence (including	Search datasets and web search as part of important research efforts,
	work related to large recommender systems and deep learning	including research focused on misinformation and/or disinformation. A
	algorithms) and data science.	summary of Microsoft Research developments and projects for the year 2023
		is available at: Research at Microsoft 2023: A year of groundbreaking Al
	While the foregoing work remains critical to our mission, we are	advances and discoveries - Microsoft Research.
	working to expand access to data for research purposes consistent	
	with the goals of the CoP as well as the applicable requirements of	Microsoft Research and the AI for Good Lab regularly explore potential
	the DSA and look forward to providing further information on this in	partnerships with third party research institutions and are actively in
	future reports.	discussions with research institutions on potential misinformation and
		disinformation related research that may leverage Bing Search data.
		Microsoft also plans to build upon its partnership with Princeton University

Additional the solution of the	
Additionally, LinkedIn regularly explores potential partnerships with	and Carnegie Foundation and is looking at ways to increase researcher access
non-governmental and research institutions and is actively in	to data.
discussions with one research institution to conduct a data and	
recommender system pilot project leveraging LinkedIn data.	During the Reporting Period, Microsoft announced meaningful contributions
LinkedIn hopes to publicly announce this partnership in its next	to support the National Al Research Resource Pilot to broaden Al research
report.	and spur innovation by providing greater computing resources to AI
	researchers and engineers in academia and non-profit sectors. As part of
Finally, LinkedIn has in place the needed teams and tools to make	
data available to researchers in a variety of ways, including via Excel	Microsoft's commitment to the NAIRR pilot, Microsoft will contribute: \$20
or XML files, GitHub repositories, sandboxed laptops, and APIs.	million in Microsoft Azure compute credits, advanced resources for
of AME files, GILHUD repositories, sandboxed laptops, and APIs.	developing trustworthy AI, including tools for research in AI fairness,
	accuracy, reliability, transparency, privacy, security, and model orchestration,
	innovative tools for scientific discovery via Azure Quantum Elements, and
	collaborative opportunities with Microsoft's scientists and engineers.
	In Fall 2023, Microsoft also opened applications for the Microsoft Research
	Al & Society fellows program. The global program aims to catalyze research
	collaboration between Microsoft Research and eminent scholars and experts
	across a range of disciplines core to discussions at the intersection of AI and
	its impact on society, and scholars and researchers from around the world
	can apply to be a fellow. See more at <u>AI &amp; Society Fellows - Microsoft</u>
	Research.
	Microsoft Research also established the Accelerate Foundation Models
	Research (AFMR) global network and resource platform to assemble an
	interdisciplinary research community around solving some of today's
	greatest technical and societal challenges through aligning AI with shared
	human goals, values, and preferences; improving human-Al interactions; and
	accelerating scientific discoveries. In line with Microsoft support of the White
	House's responsible AI voluntary commitments, grants provide access to
	state-of-the-art foundation models to make sure researchers outside the
	company can appropriately examine cutting-edge model applications and
	their impact. Interested researchers may learn more or reach out about

		collaboration opportunities at the <u>Accelerating Foundation Models Research</u> <u>hub.</u> Microsoft is currently undertaking additional research on how users interact with content provenance tools and the use of content provenance tools for Al and elections. Microsoft researchers are also currently engaged in research leveraging search data to explore how medical hoaxes went viral during the COVID-19 pandemic and research concerning the detection of bias in mainstream news in connection with elections using LLMs
		Microsoft Research has also undertaken research related to information integrity and elections in the age of generative AI, which was discussed as part of Microsoft's Research forum on March 7, 2024. Microsoft expects to publish additional information soon. Lastly, Bing Search regularly partners with third party nonprofits and research
		organizations and NGO partners to review and evaluate emerging trends, techniques, tactics, and threat intelligence in misinformation and/or disinformation and related topics.
Measure 28.2	LinkedIn	Bing Search
QRE 28.2.1	Please see QRE 26.1.1 and QRE 26.2.3.	See QRE 26.2.3 and 26.1.1.
Measure 28.3	LinkedIn	Bing Search
QRE 28.3.1	We look forward to partnering with other relevant signatories on this project and will provide further reporting as the annual consultation is established.	We look forward to partnering with other relevant signatories on this project and will provide further reporting as the annual consultation is established.
Measure 28.4	LinkedIn	Bing Search
QRE 28.4.1	We will provide information on relevant financial commitments once such requests are received and the EDMO process for providing research funds is established.	We will provide information on relevant financial commitments once such requests are received and the EDMO process for providing research funds is established.

	VI. Empowering the research community				
	Commitment 29				
J. J	Relevant Signatories commit to conduct research based on transparent methodology and ethical standards, as well as to share datasets, research findings and methodologies with relevant audiences.				
C.29 M 29.1 M 29.2 M 29.3					
We signed up to the following	Commitment 29 is aimed specifically at the research community and is therefore not				
measures of this commitment:	asures of this commitment: relevant or pertinent to Microsoft and its services.				

# VII. Empowering the fact-checking community Commitments 30 - 33

### **VII. Empowering the fact-checking community**

Commitment 30

Relevant Signatories commit to establish a framework for transparent, structured, open, financially sustainable, and non-discriminatory cooperation between them and the EU fact-checking community regarding resources and support made available to fact-checkers

	C.30	M 30.1	M 30.2	M 30.3	M 30.4
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search				

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No	Yes
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	<ul> <li>Bing is actively exploring additional fact-checking support to provide enhanced coverage across EEA Member States and languages.</li> <li>Bing has also identified additional areas where fact-checks are surfaced to Bing users and has provided additional information below.</li> </ul>
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures

LinkedIn is working on increasing	Bing is in the process of entering an
the language coverage of its fact-	additional agreement for enhanced
checking arrangement.	fact-checking in the EEA and will
	soon begin piloting those
	additional resources.
	Bing is also continuing to work to
	secure further agreements with
	additional fact checking
	organizations.

Measure 30.1	LinkedIn	Bing Search
QRE 30.1.1	LinkedIn has entered into a fact-checking arrangement with an external, independent global news agency. This relationship helps our internal content reviewers determine if user generated content violates LinkedIn's policy on false and misleading content.	Bing Search supports the schema.org ClaimReview fact-check protocol as part of its search ingestion, as discussed further in QRE 21.1.1. Because ClaimReview is an open protocol available for all websites and search engines to use, Bing Search currently does not maintain formal agreements with any individual fact-checking organization for the use of ClaimReview tags. Nonetheless, Bing Search is in active discussions with additional fact checking organizations to evaluate potential agreements to provide expanded fact- checking coverage across EEA Member States and languages.
		In addition to ClaimReview fact-checks that can surface in search results, Bing utilizes a number of other fact-checking and trustworthiness indicators that inform the Bing algorithm and defensive search measures and may be integrated into AI features like Copilot that are grounded in Bing search results.
		Bing works with trusted third-party partners, including NewsGuard, Global Democracy Index (GDI), and Spanish-language news agency EFE, to provide early indicators of narratives, hashtags, or information operations that can be leveraged to inform early detection and defensive search strategies for Bing

QRE 30.1.2	Reuters, a global news organization with 2,500 journalists in about	<ul> <li>(including in EEA member states. In addition, Bing leverages NewsGuard "misinformation fingerprints" comprised of datasets and tooling intended to highlight and track fact checked articles online. Bing currently leverages Misinformation Fingerprints and localized reporting in English, Italian, Spanish, French, and German languages. In addition to global or non- geographically specific narratives, these Misinformation Fingerprints cover EU markets including France, Germany, Spain, Italy, and Austria as well as Ukraine, among other territories. These signals and indicators of mis/disinformation narratives, hashtags, and information integrity risks inform Bing defensive search interventions and helps Bing algorithms present users with high authority, trustworthy sources.</li> <li>Fact-checks and articles from news and fact checking organizations, including Deutsche Welle, AP, Verify, AFP, Reuters, and many others may also appear as part of specialized Bing Answers. In addition, news and fact-check articles can appear in Bing News carousels, which are often presented at the top of search results pages, depending on the nature of the user query. Microsoft maintains agreements with news publishers to surface high authority content, including articles from well-regarded fact checking organizations and journalist-driven fact-checks, high in relevant search results.</li> <li>See QRE 30.1.1 and QRE 21.1.1. Bing Search continues to evaluate additional</li> </ul>
	200 locations worldwide. Reuters is one of the largest news agencies in the world.	fact-checking organizations and tools for use by search engines, and is currently working to secure an additional agreement for enhanced fact- checking resources to cover additional EEA member states and languages. We expect this agreement to result in further coverage during the next reporting period.
QRE 30.1.3	LinkedIn has implemented internal processes empowering our hundreds of global internal content reviewers to be able to obtain a fact-check from an external fact-checker partnership. Fact-checker conclusions are reviewed by internal content reviewers to determine whether the content at issue violates LinkedIn's <u>policy on false and</u>	See QREs 30.1.1-2. Bing Search is currently working to secure an additional agreement for enhanced fact-checking resources to cover additional EEA member states and languages. We expect this agreement to result in further coverage during the next reporting period.

	misleading content and if so, the content is removed from the platform.	As noted above, any authorized fact-checking organization can leverage the ClaimReview protocol to provide fact-checks to Bing Search. Bing Search would welcome additional usage of the ClaimReview protocols in EU Member States and is actively exploring proposals and partnerships with additional fact-checking organizations to help further expand this type of fact-checking coverage.
SLI 30.1.1 - Member States and languages covered by agreements with the fact-checking organisations	Methodology of data measurement: LinkedIn receives fact checking services for content in English, Spanish, Portuguese, French, German, and Italian. LinkedIn sends content to external fact checkers regardless of the location of the member posting the content, the viewers of the content, or the topic at issue. Content that violates LinkedIn's <u>policy</u> on false and misleading content is removed.	Methodology of data measurement: As set out in QRE 30.1.2 and SLI 31.1.1, any fact-checking organisation can leverage the ClaimReview protocol to embed fact-check tags into their website (thereby adding fact-check tags or flags into indexed results) and there is no limitation in terms of languages and Member States covered. Because ClaimReview is an open protocol available for all websites and search engines to use, Bing currently does not maintain formal agreements with individual fact-checking organisations for the use of Claim Review tags. As described in QRE 30.1.1, Microsoft has a number of news agreements that include fact checking coverage and provide remuneration to fact checkers and news organizations for news that is surfaced on Bing. These agreements cover a range of languages and markets, including EU member states. While certain agreements include fact checking coverage, because these arrangements are not strictly for fact-checking services, we do not reflect these agreements in this SLI. Third party providers including NewsGuard, Global Democracy Index (GDI), and Spanish-language news agency EFE, provide early indicators of narratives, hashtags, or information operations that can be leveraged to inform early detection and defensive search strategies in the following languages and member states: Croatian, Dutch, Estonian, English, French, German, Italian, Portuguese, Polish, Romanian, Spanish, and Swedish. Bing also leverages translation resources to employ defensive measures globally.

	Nr of agreements with fact-checking organisations	However, Bing Search is currently working to secure an additional agreement for enhanced fact-checking resources to cover additional EEA member states and languages. Bing expects to formalize the agreement shortly which would provide further coverage during the next reporting period for additional EU countries and languages, including Bulgarian, Catalan, Croatian, Czech, Dutch, English, Finnish, French, German, Greek, Hungarian, Polish, Portuguese, Romanian, Slovak, Spanish, and Swedish. Nr of agreements with fact-checking organisations
Member States	1	3

Measure 30.2	LinkedIn	Bing Search
QRE 30.2.1	LinkedIn has engaged in arms-length negotiations with a large global news organisation that follows the highest ethical standards in news reporting, including those related to accuracy, independence, integrity, and freedom from bias. See Our agreements give the fact-checkers complete discretion in providing their factchecking conclusions, and LinkedIn personnel leverage these conclusions to determine whether the content at issue violates LinkedIn's policy on false and misleading content and if so, the content is removed from the platform.	Bing Search is actively working to secure additional fact-checking partners and has engaged in efforts to discuss potential relationships with multiple fact-checking organizations. Bing Search supports the free and open ClaimReview protocol, which may be utilized by all fact-checking organizations to embed fact-check tags into content, articles, and websites that appear in search indexes. Bing Search welcomes the opportunity to partner with additional fact checking organizations. See also QREs 30.1.1-3. In addition, Microsoft provides fair compensation and has engaged in arms- length negotiations with news and fact-checking organizations through news partnership arrangements that support Bing news product features, such as specialized answers and news carousels. This includes European news and media organizations as well as fact-checking organizations such as AFP Fact Check, Verify, Maldita.es, dpa Germany, and Deutsche Welle.
QRE 30.2.2	LinkedIn regularly meets with its fact-checking partner to discuss improvements in process.	As noted above, Bing Search ingests ClaimReview tags embedded in fact- check content posted on websites that are indexed in the Bing Search index.

		<ul> <li>Webmasters for fact-checking organizations have self-help tools available that allow them to review website analytics and search effectiveness (including insights into keywords or search queries used in Bing Search to reach their website) for websites containing ClaimReview tags as part of Bing Search's Webmaster Tools, discussed further at QRE 32.1.1. This dashboard provides website operators with a range of data and analytics that can be used by fact-checking organizations to assess how users found their fact-checked content, website traffic patterns, and the effectiveness of their fact-check tags.</li> <li>See QREs 30.1.1-2 for additional information on Bing Search's ClaimReview fact check program.</li> <li>Bing has also engaged in conversations with members of the fact-checking tools and discuss potential agreements for enhanced fact-checking coverage.</li> </ul>
QRE 30.2.3	This QRE is not relevant or pertinent as LinkedIn is not a fact- checking organisation.	This QRE is not applicable to Bing Search, as it is not a fact-checking organization.
Measure 30.3	LinkedIn	Bing Search
QRE 30.3.1	LinkedIn regularly meets with its fact-checking partner to discuss improvements in process.	Bing has engaged in conversations with members of the fact-checking community and signatories to solicit feedback on its fact-checking tools and discuss potential agreements for enhanced fact-checking coverage in EEA member states and languages.
Measure 30.4	LinkedIn	Bing Search
QRE 30.4.1	LinkedIn stands ready to cooperate in accordance with this QRE at the appropriate time.	Bing stands ready to cooperate in accordance with this QRE at the appropriate time.

VII. Empowering the fact-checking community							
Commitment 31							
Relevant Signatories commit to integrate, showcase, or otherwise consistently use fact-checkers' work in their platforms' services, processes, and contents; with full coverage of all Member States and languages.							
	C.31		M 31.1	M 31.2	M 31.3	M 31.4	
We signed up to the	We signed up to the LinkedIn LinkedIn LinkedIn LinkedIn LinkedIn LinkedIn						
following measures of	Bing Search			Bing Search	Bing Search	Bing Search	
this commitment:							

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	Bing is actively exploring additional
		fact-checking support to provide
		enhanced coverage across EEA
		Member States and languages.
		Bing has also provided additional
		information on how fact check
		information is surfaced to users.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures

As noted in Commitment 30,	Bing plans to enter an additional
LinkedIn is working on increasing	agreement for enhanced fact-
the language coverage of its fact-	checking resources in the EU and
checking arrangement.	expects to begin piloting those
	additional resources soon.
	Bing is also continuing to workto
	secure agreements with additional
	fact checking organizations.

Measure 31.1	LinkedIn	This Measure is not relevant or pertinent to Bing Search as it does not showcase User Generated Content.
Measure 31.2	LinkedIn	Bing Search
QRE 31.1.1	LinkedIn leverages its fact-checker to review user generated content that may violate its Professional Community Policies, which prohibit misinformation. Content that violates LinkedIn's Professional Community Policies is removed from LinkedIn.	Bing Search uses ClaimReview tags embedded in websites with fact-checked content to help inform its algorithms (i.e., by leading users to more authoritative sources of information) and to provide useful context and indications of trustworthiness to its users in search results.
		See QREs 21.1.1 and 30.2.1-2 for an overview of Bing Search's use of Schema.org ClaimReview fact check labels and ways news and fact check articles are surfaced to users.
	Methodology of data measurement:	Methodology of data measurement:
SLI 31.1.1 - use of fact-checks	The figure for the number of content pieces reviewed by fact- checkers represents the number of pieces sent to our external fact checkers during the period 1 July 2023 – 31 December 2023. See also SLI 21.1.2	
		<b>Fact Check Impressions ("FCI")</b> – The number of times the above-mentioned URLs appeared on the first page of Bing search results to a user located in the EU Member States.

	Nr of fact-checked	Reach of	Nr of content	Other	Nr of fact-checked	Reach of fact-	Nr of content	Other (FC URL)
	articles published	fact-checked	pieces		articles published	checked (FCI)	pieces reviewed	
			reviewed by				by fact-checkers	
			fact-checkers					
Member States								
Austria						18,913		2,685
Belgium						54,272		4,184
Bulgaria						1		1
Croatia						1		1
Cyprus						1		1
Czech Republic						1		1
Denmark						5,147		1,057
Estonia						1		1
Finland						3,332		790
France						170,314		5,079
Germany						5,691,220		10,328
Greece						2		2
Hungary						1		1
Ireland						17,508		3,156
Italy						37,630		3,367
Latvia						3		2
Lithuania						1		1
Luxembourg						1		1
Malta						0		0
Netherlands						56,449		5,751
Poland						21,824		2,404
Portugal						20,624		2,427
Romania						10		8
Slovakia						1		1
Slovenia						0		0
Spain						135,093		5,251

Sweden					15,927	2,814
Iceland					0	0
Liechtenstein					0	0
Norway					8,710	1,960
Total EU				N/A	6,248,277	49,314
Total EEA				N/A	6,256,987	51,274
Total Global	0	N/A	449			

	Methodology of data measurement:			Methodology of data measurement:		
	violates our policy on <u>false and misleading content</u> .			Bing Search is not able to provide metrics concerning the "impact" of individual fact check websites appearing in its search results in response to specific queries.		
SLI 31.1.2 - impact of actions taken	Nr of pieces of content labelled	Impact of said measures on user interactions with information labelled as false or misleading	Other	Nr of pieces of content labelled	Impact of said measures on user interactions with information labelled as false or misleading	Other
Total EEA	0	Not applicable		0	Not applicable	

	Methodology of data measurement:	Methodology of data measurement:
SLI 31.1.3 – Quantitative information used for	LinkedIn looks forward to continue working with other signatories on contextualisation for SLIs 31.1.1 and 31.1.2, taking into account the specificities of our platform.	Bing Search looks forward to continue working with other signatories on the contextualisation for SLIs 31.1.1 and 31.1.2, taking into account the specificities of our service.
contextualisation for the SLIs 31.1.1 / 31.1.2	Denominator to be decided within the TF ahead of the baseline report Not applicable	
Total EEA	Not applicable	Not applicable

Measure 31.3	LinkedIn	Bing Search
QRE 31.3.1	As the work in the Permanent Task-force on the development of the repository of fact-checking content had not yet started during this reporting period, no contributions as listed under Measure 31.3 can currently be reported on.	As the work in the Permanent Task-force on the development of the repository of fact-checking content had not yet started during this reporting period, no contributions as listed under Measure 31.3 can currently be reported on.
Measure 31.4	LinkedIn	Bing Search
QRE 31.4.1	As the work in the Permanent Taskforce on the development of the repository of fact-checking content had not yet started during this reporting period, no technical solutions as referred to under Measure 31.4 can currently be reported on.	As the work in the Permanent Taskforce on the development of the repository of fact-checking content had not yet started during this reporting period, no technical solutions as referred to under Measure 31.4 can currently be reported on.

#### VII. Empowering the fact-checking community

Commitment 32

Relevant Signatories commit to provide fact-checkers with prompt, and whenever possible automated, access to information that is pertinent to help them to maximise the quality and impact of fact-checking, as defined in a framework to be designed in coordination with EDMO and an elected body representative of the independent European fact-checking organisations.

	C.32	M 32.1	M 32.2	M 32.3
We signed up to the following	LinkedIn	LinkedIn	LinkedIn	LinkedIn
measures of this commitment:	Bing Search	Bing Search		Bing Search

	LinkedIn	Bing Search
In line with this commitment, did you deploy new implementation measures (e.g. changes to	No	Yes
your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	New Implementation Measures	New Implementation Measures
	Not applicable	

		Bing has engaged in conversations with members of the fact-checking community and signatories to solicit feedback on its fact-checking tools and discuss potential
		agreements for enhanced fact- checking coverage. Bing Search
		looks forward to partnering with other relevant signatories and the
		fact-checking community on this
		Measure moving forward.
Do you plan to put further implementation measures in place in the next 6 months to	Yes	Yes
substantially improve the maturity of the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned Implementation Measures	Planned Implementation Measures
	As noted in Commitment 30,	Bing is working to secure
	LinkedIn is working on increasing	agreements with additional fact
	the language coverage of its fact-	checking organizations and
	checking arrangement.	exploring additional opportunities
		for furthering support information
		exchange with the fact-checking
		community.

Measure 32.1	LinkedIn	Bing Search
Measure 32.2	LinkedIn	This Measure is not relevant to Bing Search as it does not showcase User
		Generated Content.
QRE 32.1.1	LinkedIn's Professional Community Policies prohibit misinformation,	See QREs 21.1.1 and 30.2.1-2.
	and LinkedIn will remove policy-violating content, after consulting	
	with its fact-checking partners (if needed). LinkedIn provides its fact-	
	checking partners with live links to content, which allows partners to	

	easily determine whether that content was thereafter removed by LinkedIn.			
SLI 32.1.1 - use of the interfaces and other			Methodology of data measurement:	
tools			See SLI 31.1.1.	
10013	Monthly users	Other	Monthly users	Other
Total EEA	Not applicable.			
Measure 32.3	LinkedIn		Bing Search	
	LinkedIn currently uses the Code's Task-force, in particular the Crisis		Bing Search currently uses the Code	e's Task-force, in particular the Crisis
	Response and Empowerment of Fact-checkers subgroups, as a channel of communication with the fact-checking community			
QRE 32.3.1				
	represented by the signatories to the Code.		signatories to the Code. We continue to explore ways in which we can further	
			support information exchange with th	e fact-checking community.

VII. Empowering the fact-checking community		
	Commitment 33	
Relevant Signatories (i.e. fact-checking organisations) commit to operate on the basis of strict ethical and transparency rules, and to protect their independence.		
	C.33	M 33.1
We signed up to the	Commitment 33 is aimed specifically at fact-checking organisations and therefore not relevant or pertinent to	N/A
following measures of	Microsoft and its services.	
this commitment:		

# VIII. Transparency Centre Commitments 34 - 36

VIII. Transparency Centre		
Commitment 34		
To ensure transparency and accountability around the implementation of this Code, Relevant Signatories commit to set up and maintain a publicly available common Transparency Centre website		
In line with this commitment, did you deploy new	Yes	
implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet points].	Signatories changed vendor for the Transparency Center.	
Do you plan to put further implementation measures in place	Yes	
in the next 6 months to substantially improve the maturity of		
the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to	Planned Implementation Measures	
put in place in the next 6 months?		
	Microsoft is committed to the proper functioning of the Transparency Center website and will therefore	
	continue its engagement in the Transparency Center subgroup in order to assess the necessity of technical adjustments and new actions to improve the website. Microsoft will thereby contribute to, where necessary,	
	making the website more user-friendly and easily accessible for users ahead of the next reporting period.	
Measure 34.1	Microsoft Corporation	
Measure 34.2	Microsoft Corporation	
Measure 34.3	Microsoft Corporation	
Measure 34.4	Microsoft Corporation	
Measure 34.5	Microsoft Corporation	

#### VIII. Transparency Centre

# Commitment 35

Signatories commit to ensure that the Transparency Centre contains all the relevant information related to the implementation of the Code's Commitments and Measures and that this information is presented in an easy-to-understand manner, per service, and is easily searchable.

Yes
New implementation Measures
Microsoft will upload its March 2024 Report to the Transparency Centre website in a timely manner, which
includes clear and simple information on the new or existing policies and actions that each service has
implemented based on our Subscription document.
Yes
Planned Implementation Measures
Within the context of the work of the Transparency Center subgroup, Microsoft will assess the necessity of
technical adjustments and contribute to actions where necessary, aiming to make the website more user-
friendly and easily accessible for users ahead of the next reporting period.
Microsoft Corporation

# VIII. Transparency Centre

# Commitment 36

#### Signatories commit to updating the relevant information contained in the Transparency Centre in a timely and complete manner.

In line with this commitment, did you deploy new	Yes
implementation measures (e.g. changes to your terms of	
service, new tools, new policies, etc)? [Yes/No]	
If yes, list these implementation measures here [short bullet	New implementation Measures
points].	
	By uploading this report, Microsoft updated the Transparency Centre with relevant information related to its
	new policies and implementation actions.
Do you plan to put further implementation measures in place	Yes
in the next 6 months to substantially improve the maturity of	
the implementation of this commitment? [Yes/No]	
If yes, which further implementation measures do you plan to	Planned Implementation Measures
put in place in the next 6 months?	
	Microsoft is committed to providing regular updates as set out under Measures 36.1, 36.2 and 36.3.
Measure 36.1	Microsoft is committed to providing regular updates as set out under Measures 36.1, 36.2 and 36.3. Microsoft Corporation
Measure 36.1 Measure 36.2	
	Microsoft Corporation
Measure 36.2	Microsoft Corporation Microsoft Corporation
Measure 36.2 Measure 36.3	Microsoft Corporation Microsoft Corporation Microsoft Corporation
Measure 36.2 Measure 36.3	Microsoft Corporation         Microsoft Corporation         Microsoft Corporation         The administration of the Transparency Centre website has been transferred fully to the community of the
Measure 36.2 Measure 36.3 QRE 36.1.1 (for the Commitments 34-36)	Microsoft Corporation         Microsoft Corporation         Microsoft Corporation         The administration of the Transparency Centre website has been transferred fully to the community of the Code's signatories, with VOST Europe taking the role of developer.
Measure 36.2 Measure 36.3 QRE 36.1.1 (for the Commitments 34-36)	Microsoft Corporation         Microsoft Corporation         Microsoft Corporation         The administration of the Transparency Centre website has been transferred fully to the community of the Code's signatories, with VOST Europe taking the role of developer.         The Transparency Centre was successfully launched in February 2023. Microsoft will outline any material
Measure 36.2 Measure 36.3 QRE 36.1.1 (for the Commitments 34-36)	Microsoft Corporation         Microsoft Corporation         Microsoft Corporation         The administration of the Transparency Centre website has been transferred fully to the community of the Code's signatories, with VOST Europe taking the role of developer.         The Transparency Centre was successfully launched in February 2023. Microsoft will outline any material changes made to the Transparency Centre's content, operations, or functioning in upcoming reports, as
Measure 36.2 Measure 36.3 QRE 36.1.1 (for the Commitments 34-36) QRE 36.1.2 (for the Commitments 34-36)	Microsoft Corporation         Microsoft Corporation         Microsoft Corporation         The administration of the Transparency Centre website has been transferred fully to the community of the Code's signatories, with VOST Europe taking the role of developer.         The Transparency Centre was successfully launched in February 2023. Microsoft will outline any material changes made to the Transparency Centre's content, operations, or functioning in upcoming reports, as relevant.
Measure 36.2 Measure 36.3 QRE 36.1.1 (for the Commitments 34-36) QRE 36.1.2 (for the Commitments 34-36) SLI 36.1.1 - (for Measures 34 and 36) meaningful	Microsoft Corporation         Microsoft Corporation         Microsoft Corporation         The administration of the Transparency Centre website has been transferred fully to the community of the Code's signatories, with VOST Europe taking the role of developer.         The Transparency Centre was successfully launched in February 2023. Microsoft will outline any material changes made to the Transparency Centre's content, operations, or functioning in upcoming reports, as relevant.

Transparency Centre, such as the average monthly visits of the webpage.	
Data	The common Transparency Center has received around 35.000 views between July 1 <sup>st</sup> 2023 and December 31 <sup>st</sup> 2023. The average engagement time on the website is 1m11s and reports have been downloaded more
	than 9.000 times.

# IX. Permanent Task-Force Commitment 37

#### IX. Permanent Task-Force

# Commitment 37

Signatories commit to participate in the permanent Task-force. The Task-force includes the Signatories of the Code and representatives from EDMO and ERGA. It is chaired by the European Commission, and includes representatives of the European External Action Service (EEAS). The Task-force can also invite relevant experts as observers to support its work. Decisions of the Task-force are made by consensus.

In line with this commitment, did you deploy new	Yes
implementation measures (e.g. changes to your terms of	
service, new tools, new policies, etc)? [Yes/No]	
If yes, list these implementation measures here [short bullet	New implementation Measures
points].	
	Microsoft has actively engaged in and contributed to the work of the Task-force and all of its Subgroups and
	Working Groups. Moreover, Microsoft has joined the newly established Subgroup on Generative AI and
	Disinformation as a co-coordinator.
Do you plan to put further implementation measures in place	Yes
in the next 6 months to substantially improve the maturity of	
the implementation of this commitment? [Yes/No]	
If yes, which further implementation measures do you plan to	Planned Implementation Measures
put in place in the next 6 months?	
	Microsoft is committed to continuing its active engagement in and contribution to the Task-force and all of
	its Subgroups and Working Groups in the upcoming six-month period.

Measure 37.1	Microsoft Corporation
Measure 37.2	Microsoft Corporation
Measure 37.3	Microsoft Corporation
Measure 37.4	Microsoft Corporation
Measure 37.5	Microsoft Corporation
Measure 37.6	Microsoft Corporation
QRE 37.6.1	<ul> <li>Microsoft has actively engaged in and contributed to all the Task-force Plenary meetings as well as to the meetings of all Subgroups and Working Groups active in the current reporting cycle under the Task-force: <ul> <li>Subgroup on Scrutiny of Advertising</li> <li>Subgroup on Integrity of Services</li> <li>Subgroup on Monitoring and Reporting</li> <li>Subgroup on Crisis Response</li> <li>Working Group on Elections</li> <li>Crisis-Elections Steering Committee</li> <li>Working Group on Structural Indicators</li> <li>Subgroup on Outreach and Integration of New Signatories</li> <li>Subgroup on Empowerment of Fact-checkers</li> <li>Mini-board on Fact-checking Repository</li> <li>Subgroup on Generative AI and Disinformation</li> </ul> </li> <li>As part of each Subgroup and Working Group, Microsoft has actively contributed to the development of the deliverables that were collectively agreed leading up to this reporting cycle.</li> <li>Microsoft has continuously engaged with all Signatories of the Code, offering its perspectives on issues unique to its subscribed services and responding to ad-hoc inquiries related to various actions taken by its subscribed services. Microsoft appreciates the added value and insights that the Task-Force has created for each Signatory individually as well as the collective community of Signatories. Microsoft looks forward to continue its constructive cooperation within the Code of Practice's governance framework.</li> </ul>

# X. Monitoring of Code Commitment 38 - 44

Commitment 38

The Signatories commit to dedicate adequate financial and human resources and put in place appropriate internal processes to ensure the implementation of their commitments under the Code.

Yes
New implementation Measures
A dedicated cross-company team continues to ensure proper tracking and compliance with the Code of
Practice across all applicable geographical areas, consisting of relevant product members from all subscribed
services, attorneys, members of the European Government Affairs team and Democracy Forward Team. Budget
items from across Microsoft teams have been used to ensure compliance including ongoing investment in
trusted third parties.
Yes
Planned Implementation Measures
Microsoft will keep its dedicated cross-company team in place to ensure continued progress towards
compliance of relevant measures.
Microsoft Corporation
Microsoft implemented a dedicated cross-company team to ensure proper tracking and compliance with the
Code of Practice across all applicable geographical areas, consisting of relevant product members from all
subscribed services as well as relevant lawyers, members of the European Government Affairs team and
Democracy Forward Team. In addition, we implemented an internal tracking process that captured all relevant
commitments, responsible entity and persons responsible for compliance with the Code of Practice.
Moverover, regular review of new product features take place to assess potential impacts and compliance
under the Code.

X. Monitoring of Code		
	Commitment 39	
Signatories commit to provide to the European Com	mission, within 1 month after the end of the implementation period (6 months after this Code's	
signatu	re) the baseline reports as set out in the Preamble.	
In line with this commitment, did you deploy new	No	
implementation measures (e.g. changes to your terms of		
service, new tools, new policies, etc)? [Yes/No]		
If yes, list these implementation measures here [short bullet	New implementation Measures	
points].		
	This Commitment is only relevant for the Baseline Reports, which were provided to the European Commission	
	in January 2023.	
Do you plan to put further implementation measures in place	No	
in the next 6 months to substantially improve the maturity of		
the implementation of this commitment? [Yes/No]		
If yes, which further implementation measures do you plan to	Planned Implementation Measures	
put in place in the next 6 months?		
	Not applicable.	

# Commitment 40

Signatories commit to provide regular reporting on Service Level Indicators (SLIs) and Qualitative Reporting Elements (QREs). The reports and data provided should allow for a thorough assessment of the extent of the implementation of the Code's Commitments and Measures by each Signatory, service and at Member State level.

Yes
New implementation Measures
This Report sets out in detail the measures that Microsoft and its subscribing services have taken in view of
the regular reporting that they have subscribed to under the Code of Practice.
Yes
Planned Implementation Measures
Microsoft has worked intensively over the past months to ensure it is in a good position for the delivery of the
March 2024 Report and looks forward to further expanding and deepening its implementation work ahead of
the next reporting cycle.
Microsoft Corporation

### Commitment 41

Signatories commit to work within the Task-force towards developing Structural Indicators, and publish a first set of them within 9 months from the signature of this Code; and to publish an initial measurement alongside their first full report. To achieve this goal, Signatories commit to support their implementation, including the testing and adapting of the initial set of Structural Indicators agreed in this Code. This, in order to assess the effectiveness of the Code in reducing the spread of online disinformation for each of the relevant Signatories, and for the entire online ecosystem in the EU and at Member State level. Signatories will collaborate with relevant actors in that regard, including ERGA and EDMO.

Measure 41.3	Microsoft Corporation
Measure 41.2	Microsoft Corporation
Measure 41.1	Microsoft Corporation
	dedicated to developing Structural Indicators on a long-term approach.
	We continue to support the publication of Structural Indicators, and will collaborate with the working group
put in place in the next 6 months?	
If yes, which further implementation measures do you plan to	Planned Implementation Measures
the implementation of this commitment? [Yes/No]	
in the next 6 months to substantially improve the maturity of	
Do you plan to put further implementation measures in place	Yes
	We support the publication of a second report on Structural Indicators by TrustLab.
	We continue to be an active participant in the working group dedicated to developing Structural Indicators.
points].	
If yes, list these implementation measures here [short bullet	New implementation Measures
service, new tools, new policies, etc)? [Yes/No]	
implementation measures (e.g. changes to your terms of	
In line with this commitment, did you deploy new	Yes

### Commitment 42

Relevant Signatories commit to provide, in special situations like elections or crisis, upon request of the European Commission, proportionate and appropriate information and data, including ad-hoc specific reports and specific chapters within the regular monitoring, in accordance with the rapid response system established by the Taskforce.

In the south this second to set all loss a sheet or a	Ver
In line with this commitment, did you deploy new	Yes
implementation measures (e.g. changes to your terms of	
service, new tools, new policies, etc)? [Yes/No]	
If yes, list these implementation measures here [short bullet	New implementation Measures
points].	
	Microsoft has been an active participant in and contributor to the Task-force's Crisis Response Subgroup, in which it proactively provides analysis and data related to influence operations, foreign interference in information space and relevant incidents that emerges on its service, including through responding to ad-hoc requests for information by the European Commission and contributing to the development of the Crisis Response templates as part of the regular monitoring cycle of the Code. Microsoft's internal threat detection and research teams, including the Microsoft Threat Analysis Center (MTAC), Microsoft Threat Intelligence Center (MSTIC), Microsoft Research (MSR), and AI For Good Lab, collect and analyse data on actors of disinformation, misinformation and information manipulation across platforms. These teams work with external organisations and companies to share and ingest data that help support Microsoft product and service teams effectively respond to issues and threats.
Do you plan to put further implementation measures in place	Yes
in the next 6 months to substantially improve the maturity of	
the implementation of this commitment? [Yes/No]	
If yes, which further implementation measures do you plan to	Planned Implementation Measures
put in place in the next 6 months?	
	Microsoft looks forward to continuing the cooperation between Signatories in the context of the Task-force's
	Crisis Response Subgroup and Elections Working Group.

# Commitment 43

Signatories commit to produce reports and provide data following the harmonised reporting templates and refined methodology for reporting and data disclosure, as agreed in the Task-force.

In line with this commitment, did you deploy new	Yes
implementation measures (e.g. changes to your terms of	
service, new tools, new policies, etc)? [Yes/No]	
If yes, list these implementation measures here [short bullet	New implementation Measures
points].	
	Microsoft has provided its March 2024 Report in accordance with the revised Harmonised Reporting Template
	and underlying methodologies as jointly developed by Signatories in the Monitoring and Reporting Subgroup
	under the Code's Task-force.
Do you plan to put further implementation measures in place	Yes
in the next 6 months to substantially improve the maturity of	
the implementation of this commitment? [Yes/No]	
If yes, which further implementation measures do you plan to	Planned Implementation Measures
put in place in the next 6 months?	
	Microsoft will continue its active engagement in the respective Task-force Subgroups to further develop and
	improve the Harmonised Reporting Template and underlying methodologies where necessary in view of its
	experience with reporting.

#### Commitment 44

Relevant Signatories that are providers of Very Large Online Platforms commit, seeking alignment with the DSA, to be audited at their own expense, for their compliance with the commitments undertaken pursuant to this Code. Audits should be performed by organisations, independent from, and without conflict of interest with, the provider of the Very Large Online Platform concerned. Such organisations shall have proven expertise in the area of disinformation, appropriate technical competence and capabilities and have proven objectivity and professional ethics, based in particular on adherence to auditing standards and guidelines.

In line with this commitment did you deploy new	No
In line with this commitment, did you deploy new	
implementation measures (e.g. changes to your terms of	
service, new tools, new policies, etc)? [Yes/No]	
If yes, list these implementation measures here [short bullet	New implementation Measures
points].	
	N/A
Do you plan to put further implementation measures in place	Yes
in the next 6 months to substantially improve the maturity of	
the implementation of this commitment? [Yes/No]	
If yes, which further implementation measures do you plan to	Planned Implementation Measures
put in place in the next 6 months?	
	Prepare for conducting audits of relevant services.

# Reporting on the services' response during a period of crisis

# Reporting on the services' response during a crisis

#### War of aggression by Russia on Ukraine

Threats observed or anticipated at time of reporting:

**LinkedIn** is an online professional networking site with a real identity requirement, which means that content posted by our members is visible to that member's professional network, including colleagues, managers, and potential future employers. As a result of LinkedIn's professional context, our members do not tend to post misinformation, nor does misinformation content gain traction on LinkedIn. Nonetheless, LinkedIn may be subject to certain members posting misinformation during crisis situations.

**Microsoft Advertising**, in its role as an online advertising network, may be subject to the malicious use of its advertising services through either the spreading of misleading or deceptive advertising content or the funneling of advertising revenue to sites spreading disinformation.

**Bing Search** has observed instances of data void manipulation to show low-authority content to unsuspecting users related to Russia's war against Ukraine. This type of search algorithm manipulation could potentially lead to the spread of misinformation/disinformation.

Mitigations in place at time of reporting:

**Microsoft** has been actively involved in identifying and helping counter Russia's cyber and influence operations aimed against Ukraine. In addition to supporting nonprofits, journalists, and academics within Ukraine, Microsoft's Threat Analysis Center (MTAC) team closely tracks cyber-enabled influence operations. MTAC analysts focused on Europe/Eurasia report on a wide range of Russian influence tactics used to malign or diminish support for Ukraine: propaganda and disinformation published across different languages; people-to-people and party-to-party engagement; real-world provocations; and those that blend cyber and influence activity, like hack-and-leak campaigns. MTAC's work includes analysing the ways these methods are leveraged to target audiences in Central and Eastern Europe.

In June of 2022, Microsoft issued its "<u>Defending Ukraine</u>" report and a follow up report issued in December 2022, both of which detailed the relentless and destructive Russian cyberattacks and influence operations, that we have directly observed in the hybrid war Russia is waging against Ukraine. Microsoft followed those reports with a report in March of 2023, outlining how Russia was regrouping for additional offensive measures against Ukraine including cyber and influence operations and a <u>report</u> in December of 2023 assessing Russian influence and cyber operations, including Russia's anti-Ukraine messaging to Israel and elsewhere. In February of 2024, Microsoft and OpenAI Issued a threat <u>report on activity by adversaries utilizing AI capability</u>. This report identified recent Russian activity including activities targeting Ukraine.

**LinkedIn**'s Professional Community Policies expressly prohibit false and misleading content, including <u>misinformation and disinformation</u>, and its in-house Editorial team provides members with trustworthy content regarding global events, including the conflict in Ukraine. LinkedIn has an internal team of hundreds of content reviewers located all over the world (for 24/7 coverage), approximately 235 of which are located in the EMEA region, and includes specialists in a number of languages including English, German, French, Russian, and Ukrainian. These reviewers use policies and guidance developed by a dedicated content policy team and experienced lawyers, and work with external fact checkers as needed. When LinkedIn sees content or behaviour that violates its Professional Community Policies, it takes action, including the removal of content or the restriction

of an account for repeated abusive behaviour. LinkedIn has been banned in Russia since 2016 and has implemented the European bans on Russian state media. In addition to not operating in Russia, political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue, including the current Russia-Ukraine war. LinkedIn also does not provide a mechanism for content creators to monetise the content they post on LinkedIn.

**LinkedIn** continues to mature its crisis response processes. In addition to the increase in resource allocation and process improvements, best practices include: 1) quickly coordinating with industry peers regarding the exchange of threat indicators; 2) engaging with external stakeholders regarding trends and TTPs; 3) continuously providing updated policy guidance to internal teams to assist with the removal of misinformation; and 4) continuing to proactively provide localised trustworthy information to our members.

LinkedIn has continued to mature its crisis response playbook by continually monitoring crisis situations globally, expanding internal teams that work on crisis response, and maturing our processes to respond more efficiently and effectively to crisis situations. **LinkedIn** will continue to follow its processes related to the removal of misinformation, and continually increase investments in resource allocation and process improvements where necessary to respond to the demands of the crisis.

**Bing Search** has implemented the following measures: (1) Defensive search interventions; (2) regular direction of users to high authority, high quality sources; (3) removed auto suggest and related search terms considered likely to lead users to low authority content; and (4) partnered with research and nonprofit organizations to maintain threat intelligence and inform potential algorithmic interventions. These measures are also integrated into Bing generative AI experiences, along with the additional safeguards discussed at QRE 14.1.1 and QRE 14.1.2 and other measures discussed throughout this report.

[Note: Signatories are requested to provide information relevant to their particular response to the threats and challenges they observed on their service(s). They ensure that the information below provides an accurate and complete report of their relevant actions. As operational responses to crisis/election situations can vary from service to service, an absence of information should not be considered a priori a shortfall in the way a particular service has responded. Impact metrics are accurate to the best of signatories' abilities to measure them].

#### Policies and Terms and Conditions

Outline any changes to your policies

Policy	Changes (such as newly introduced policies, edits, adaptation in scope or implementation)	Rationale
LinkedIn		
False or misleading content	LinkedIn continually updates its policies as appropriate during any crisis, including the Ukraine crisis.	

		foreign state actors, are all harms that existed prior to the Ukraine crisis, and therefore LinkedIn already had policies in place to address these harms.
<u>Graphic content</u>	LinkedIn now provides sensitive content warnings as a way to notify members when a photo or video on LinkedIn contains sensitive or graphic content. This is particularly relevant to graphic content related to war.	This warning gives our members control over whether they view this content, which can serve a newsworthy or educational purpose while complying with our policies. More information about these warnings is available <u>here</u> .
Microsoft Advertising		
Microsoft Advertising has not implemented changes to its policies or introduced further measures specific to this crisis since the last reporting period.	Advertising is preventing serving advertising related to the Ukrainian crisis pursuant to its Sensitive Advertising Policies. Under this policy, Microsoft Advertising reserves the right to remove or limit advertising in response to a sensitive or high-profile news event to	
Bing Search		
Bing has not implemented additional policies specific to this crisis since its last report.	approach to addressing this crisis and continues to reporting period. As part of its work with RSA JTI Uk	tablished measures to address this crisis. Although Bing continues to refine its o evolve mitigation measures, it has not introduced specific policy changes this rainian news organizations have been JTI certified with the support of NewsGuard form mitigation measures and promotion of authoritative content.
Scrutiny of Ads Placements		
Outline approaches pertinent to this cha	apter, highlighting similarities/commonalities and diff	erences with regular enforcement.

LinkedIn			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue. LinkedIn does not pay its members for creating content, other than in rare instances where we individually select and manage a small group of content creators to create professionally relevant content. Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a		
Microsoft Advertising			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention No specific interventions to report related to this crisis. Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a		
	Political Advertising		
Outline approaches pertinent to this ch	Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.		
LinkedIn			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Not applicable. Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue. Indication of impact (at beginning of action: expected impact) including relevant metrics when available		
Microsoft Advertising			

<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention         Not applicable. As described in Section 3 of the report, Microsoft Advertising does not support political advertising.         Indication of impact (at beginning of action: expected impact) including relevant metrics when available         n/a
	Integrity of Services
Outline approaches pertinent to this	chapter, highlighting similarities/commonalities and differences with regular enforcement.
LinkedIn	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	<ul> <li>Description of intervention</li> <li>LinkedIn's Professional Community Policies prohibit misinformation, and misinformation is removed from the LinkedIn platform.</li> <li>Members that post misinformation are notified of LinkedIn's removal of their content. Members that repeatedly post misinformation are permanently restricted. State-sponsored attempts to post misinformation, if any, are removed.</li> <li>LinkedIn also maintains a robust ad-review process to prevent malicious advertising wherein ads related to crises, including the Ukraine war, are manually reviewed and approved by our internal team.</li> <li>LinkedIn regularly meets with and exchanges information with industry peers to identify and share granular information related to manipulative behaviours, coordinated influence operations, and TTPs. All such content that violates LinkedIn's Professional Community Policies is removed.</li> <li>Indication of impact (at beginning of action: expected impact) including relevant metrics when available</li> <li>LinkedIn does not track misinformation content removal by subject matter, and a comprehensive listing of its misinformation content takedowns is provided in this disclosure and in its Transparency Report. See also SLI 18.2.1.</li> </ul>
	Impact metrics for malicious advertising removals are available in SLI 2.1.1 and 2.3.1, whereas metrics related to manipulative behaviors and TTPs are available in SLI 12.1.1 and 14.2.1.

Bing Search	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing Search has implemented the following measures: (1) Defensive search interventions; (2) regular direction of users to high authority, high quality sources; (3) removed auto suggest and related search terms considered likely to lead users to low authority content; and (4) partnered with independent fact checkers, research, and nonprofit organizations to maintain threat intelligence and inform potential algorithmic interventions. These measures are also integrated into Bing generative AI experiences, along with the additional safeguards discussed at QRE 14.1.1 and QRE 14.1.2 and other measures discussed throughout this report. Indication of impact (at beginning of action: expected impact) including relevant metrics when available Globally, more than 798,000 queries searched by users related to the Russia/Ukraine conflict have been treated with defensive search, resulting in over 19.2 million impressions. In the EEA, this resulted in 2,453,601 impressions. For member state level reporting, see SLI 22.7.1.
Outline approaches pertinent to this	Empowering Users chapter, highlighting similarities/commonalities and differences with regular enforcement.
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention         LinkedIn has an internal team of global news editors that provides trustworthy and authoritative content to its member-base at all times. During important events in the Ukraine crisis, this team provides manually curated and localised storylines.         Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Bing Search Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing utilises the ClaimReview open protocol to ingest fact checks into search results and is exploring further fact checking partnerships in the EU. In addition, Bing utilizes a number of other fact-checking and trustworthiness indicators that inform the Bing algorithm and defensive search measures that power Bing and generative AI features like Copilot that are grounded in Bing search results. Bing works

	with trusted third-party partners, including NewsGuard, Global Democracy Index (GDI), and Spanish-language news agency EFE, to provide early indicators of narratives, hashtags, or information operations that can be leveraged to inform early detection and defensive search strategies for Bing (including in EEA member states).
	Articles from news and fact checking organizations, including Deutsche Welle, AP, Verify, AFP, Reuters, and many others may also appear as part of specialized Bing Answers. In addition, news and fact-check articles can appear in Bing News carousels, which are often presented at the top of search results pages, depending on the nature of user query. Microsoft maintains agreements with news publishers to surface high authority content, including articles from well-regarded fact checking organizations and journalist-driven fact-checks, high in relevant search results.
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
	See QRE 22.7.1 and QRE 21.1.1 for further information.
Microsoft	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Microsoft supported the creation of <u>The Investigators</u> , a new world for Minecraft Education that helps students build information and media literacy through game-based learning. Launched in English, the game and support materials are being localized to become available globally in 28 languages to millions of students and teachers in 2024. These languages include Bulgarian, Czech, Danish, Dutch, Finnish, French, German, Greek, Hungarian, Italian, Norwegian, Polish, Slovak, Spanish, Swedish, Russian, and Ukrainian. Microsoft also developed <u>Search Progress and Search Coach</u> , free tools built-into Teams for Education that provide real-time coaching to students on how to form effective queries and identify and review reliable sources.
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available We will be able to track the engagement of the additional languages upon their launch to indicate impact and reach of the game- based information and media literacy learning.
	Empowering the Research Community
Outline approaches pertinent to this	chapter, highlighting similarities/commonalities and differences with regular enforcement.

LinkedIn		
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention None for LinkedIn Indication of impact (at beginning of action: expected impact) including relevant metrics when available None for LinkedIn	
Bing Search		
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention See QRE 26.1.1 for general research support. Bing Search has not made datasets available pertaining specifically to this crisis. Indication of impact (at beginning of action: expected impact) including relevant metrics when available	
	n/a	
Microsoft Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention:         Microsoft has provided funding to Report for the World to support reporting positions with trustworthy newsrooms in Ukraine, including Kyiv Independent, ABO, Hromadaske and has also invested in the expansion of the Kyiv Independent's War Crimes Investigative Unit to allow for more in-depth reporting. This allows for leveraging authoritative sources of information in the region including by adding the Kyiv Independent our Start platform.         Indication of impact (at beginning of action: expected impact) including relevant metrics when available	
Empowering the Fact-Checking Community		
Outline approaches pertinent to this	chapter, highlighting similarities/commonalities and differences with regular enforcement.	

LinkedIn	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention LinkedIn works with globally-recognised factcheckers to receive their independent judgment regarding user-generated content. Content that violates LinkedIn's policy on false and misleading content is removed. Indication of impact (at beginning of action: expected impact) including relevant metrics when available LinkedIn does not separately track Ukraine related content sent to external fact checkers. Further detail regarding content sent to fact checkers is available at SLI 21.1.1, 30.1.1, and 31.1.1.
Bing Search	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing is in the process of engaging additional fact-checking support in EU Member States (including language coverage for countries in close proximity to Ukraine such as Poland, Romania, Slovakia, Hungary and Bulgaria). Bing also utilizes the ClaimReview open protocol to ingest fact checks into search results and articles from news and fact checking organizations, including Deutsche Welle, AP, Verify, AFP, Reuters, and many others may also appear as part of specialized Bing Answers. In addition, news and fact-check articles can appear in Bing News carousels, which are often presented at the top of search results pages, depending on the nature of user query.
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available

# Reporting on the services' response during a crisis

#### Israel-Hamas Conflict

Threats observed or anticipated at time of reporting:

**LinkedIn** is an online professional networking site with a real identity requirement, which means that content posted by our members is visible to that member's professional network, including colleagues, managers, and potential future employers. As a result of LinkedIn's professional context, our members come to LinkedIn for economic opportunity, and as such, do not tend to post misinformation, nor does misinformation content gain traction on LinkedIn. Nonetheless, LinkedIn may be subject to certain members posting misinformation during crisis situations.

**Bing Search** has not observed particular issues related to this crisis, but in general such crises pose a risk of data void manipulation where users might be unexpectedly exposed to low-authority content due to active manipulations by bad actors. This type of search algorithm manipulation could potentially be used as a tactic to spread disinformation.

#### Mitigations in place at time of reporting:

LinkedIn's Professional Community Policies expressly prohibit false and misleading content, including <u>misinformation and disinformation</u>, and its in-house Editorial team provides members with trustworthy content regarding global events, including the Israel-Hamas conflict. LinkedIn has an internal team of hundreds of content reviewers located all over the world (for 24/7 coverage), approximately 235 of which are located in the EMEA region, and includes specialists in languages supported on LinkedIn. These reviewers use policies and guidance developed by a dedicated content policy team and experienced lawyers, and work with external fact checkers as needed. When LinkedIn sees content or behaviour that violates its Professional Community Policies, it takes action, including the removal of content or the restriction of an account for repeated abusive behaviour. Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue, including the current Israel-Hamas conflict. LinkedIn also does not provide a mechanism for content creators to monetise the content they post on LinkedIn.

**LinkedIn** continues to mature its crisis response processes. In addition to the increase in resource allocation and process improvements, best practices include: 1) quickly coordinating with industry peers regarding the exchange of threat indicators; 2) engaging with external stakeholders regarding trends and TTPs; 3) continuously providing updated policy guidance to internal teams to assist with the removal of misinformation; and 4) continuing to proactively provide localised trustworthy information to our members.

LinkedIn has continued to mature its crisis response playbook by continually monitoring crisis situations globally, expanding internal teams that work on crisis response, and maturing our processes to respond more efficiently and effectively to crisis situations. LinkedIn will continue to follow its processes related to the removal of misinformation, and continually increase investments in resource allocation and process improvements where necessary to respond to the demands of the crisis.

**Bing Search**: To respond to the crisis, Bing Search has implemented the following measures: (1) defensive search interventions; (2) regular direction of users to high authority, high quality sources; (3) removed auto suggest and related search terms considered likely to lead users to low authority content; and (4) partnered with research and nonprofit

organizations to maintain threat intelligence and inform potential algorithmic interventions globally. These measures are also integrated into Bing generative AI experiences, along with the additional safeguards discussed at QRE 14.1.1 and QRE 14.1.2 and other measures discussed throughout this report.

[Note: Signatories are requested to provide information relevant to their particular response to the threats and challenges they observed on their service(s). They ensure that the information below provides an accurate and complete report of their relevant actions. As operational responses to crisis/election situations can vary from service to service, an absence of information should not be considered a priori a shortfall in the way a particular service has responded. Impact metrics are accurate to the best of signatories' abilities to measure them].

Policies and Terms and Conditions				
Outline any changes to your policies				
Policy	Changes (such as newly introduced policies, edits, adaptation in scope or implementation)	Rationale		
LinkedIn				
<u>False or misleading</u> <u>content</u>	LinkedIn continually updates its policies as appropriate during any crisis, including the Israel- Hamas conflict.	Misinformation, disinformation campaigns, coordinated manipulative behaviours, malicious use of advertising products, and the involvement of foreign state actors, are all harms that existed prior to the Israel-Hamas conflict, and therefore LinkedIn already had policies in place to address these harms.		
<u>Graphic content</u>	LinkedIn now provides sensitive content warnings as a way to notify members when a photo or video on LinkedIn contains sensitive or graphic content. This is particularly relevant to graphic content related to war.	This warning gives our members control over whether they view this content, which can serve a newsworthy or educational purpose while complying with our policies. More information about these warnings is available <u>here.</u>		
Microsoft Advertising				
Microsoft Advertising has not introduced specific policies related to this crisis.		Microsoft Advertising considers existing measures discussed throughout this report to sufficiently mitigate risks related to this crisis.		

Bing Search				
Bing Search has not introduced specific policies related to this crisis.		Bing Search considers existing measures discussed throughout this report to sufficiently mitigate risks related to this crisis.		
Scrutiny of Ads Placements				
Outline approaches pertinen	t to this chapter, highlighting similarities/commonalitie	s and differences with regular enforcement.		
LinkedIn				
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue, including the current Israel-Hamas conflict. LinkedIn also does not provide a mechanism for content creators to monetise the content they post on LinkedIn. LinkedIn does not pay its members for creating content, other than in rare instances where we individually select and manage a small group of content creators to create professionally relevant content. Indication of impact (at beginning of action: expected impact) including relevant metrics when available			
Microsoft Advertising				
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Microsoft Advertising is preventing serving advertising related to the Israel-Hamas conflict pursuant to its Sensitive Advertising Policies. Under this policy, Microsoft Advertising reserves the right to remove or limit advertising in response to a sensitive or high-profile news event to prevent the commercial exploitation of such events and to ensure user safety. Furthermore, our Misleading Content Policies prohibit advertising content that is misleading, deceptive, fraudulent, or that can be harmful to its users, including advertisements spreading disinformation. Microsoft Advertising requires its syndication partners (i.e., those partners that display our advertisements on their services) to abide by strict brand safety- oriented policies to avoid providing revenue streams to websites engaging in misleading, deceptive, harmful, or insensitive behaviours. Our publisher policies include a comprehensive list of prohibited content that ads cannot serve against, including, but not limited to, sensitive political content (e.g., extreme, aggressive, or misleading interpretations of news, events, or individuals), and unsavory content (such as content disparaging individuals or organisations). Partner properties that violate our policies are removed from our network until the partner remedies the issue.			

	Indication of impact (at beginning of action: expected impact) including relevant metrics when available During the relevant reporting period, Microsoft Advertising blocked any ads on 36 search queries and blocked 3 web domains across its network.
	Political Advertising
Outline approaches pertinen	t to this chapter, highlighting similarities/commonalities and differences with regular enforcement.
LinkedIn	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Not applicable. Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue, including the current Israel-Hamas conflict.
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a
Microsoft Advertising	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Not applicable. As described in Section 3 of the report, Microsoft Advertising does not support political advertising. Indication of impact (at beginning of action: expected impact) including relevant metrics when available
	n/a
Integrity of Services	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	

LinkedIn			
<b>Specific Action applied</b> (with reference to the	Description of intervention		
Code's relevant Commitment and Measure)	LinkedIn's Professional Community Policies prohibit misinformation, and misinformation is removed from the LinkedIn platform. Members that post misinformation are notified of LinkedIn's removal of their content. Members that repeatedly post misinformation are permanently restricted. State-sponsored attempts to post misinformation, if any, are removed.		
	LinkedIn also maintains a robust ad-review process to prevent malicious advertising wherein ads related to crises, including the Israel-Hamas conflict, are manually reviewed and approved by our internal team.		
	LinkedIn regularly meets with and exchanges information with industry peers to identify and share granular information related to manipulative behaviours, coordinated influence operations, and TTPs. All such content that violates LinkedIn's Professional Community Policies is removed.		
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available		
	Impact metrics for malicious advertising removals are available in SLI 2.1.1 and 2.3.1, whereas metrics related to manipulative behaviors and TTPs are available in SLI 12.1.1 and 14.2.1.		
Bing Search	Bing Search		
Specific Action applied	Description of intervention		
(with reference to the Code's relevant Commitment and Measure)	Bing Search has not introduced specific practices or policies related to this crisis, as it considers existing measures discussed throughout this report to sufficiently mitigate risks. However, Bing continues to monitor this crisis and may evolve its approach as needed.		
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available		
	n/a		
Empowering Users			
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.			
LinkedIn			

<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention LinkedIn has an internal team of global news editors that provides trustworthy and authoritative content to its member-base at all times. During important events in the Israel-Hamas conflict, this team provides manually curated and localised storylines. Indication of impact (at beginning of action: expected impact) including relevant metrics when available		
Bing Search	Bing Search		
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing Search has not introduced specific practices or policies related to this crisis, as it considers existing measures discussed throughout this report to sufficiently mitigate risks. However, Bing continues to monitor this crisis and may evolve its approach as needed. Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a		
	Empowering the Research Community		
Outline approaches pertinen	Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.		
LinkedIn			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention None for LinkedIn Indication of impact (at beginning of action: expected impact) including relevant metrics when available None for LinkedIn		
Bing Search	Bing Search		
<b>Specific Action applied</b> (with reference to the	Description of intervention See QRE 26.1.1 for general research support. Bing Search has not made datasets available pertaining specifically to this crisis.		

Code's relevant Commitment and Measure)	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
	Empowering the Fact-Checking Community
Outline approaches pertinen	t to this chapter, highlighting similarities/commonalities and differences with regular enforcement.
LinkedIn	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention LinkedIn works with globally-recognised factcheckers to receive their independent judgment regarding user- generated content. Content that violates LinkedIn's policy on <u>false and misleading content</u> is removed.
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available LinkedIn does not separately track Israel-Hamas conflict related content sent to external fact checkers. Further detail regarding content sent to fact checkers is available at SLI 21.1.1, 30.1.1, and 31.1.1.
Bing Search	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing is in the process of engaging additional fact-checking support in EU Member States but has not engaged fact checkers specifically to address this crisis. Bing also utilizes the ClaimReview open protocol to ingest fact checks into search results. Bing's generative AI features are subject to Copilot in Bing's robust safety classifiers (which are informed by data shared by information integrity partners and other authority signals) to prevent the creation of low authority materials or disinformation. As Copilot in Bing prohibits the creation of misleading content, and Bing does not allow users to publish or share content created by the system on the platform, in-product fact checks are not relevant to Bing's generative AI features. Indication of impact (at beginning of action: expected impact) including relevant metrics when available

Microsoft	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Microsoft continues to grow partnerships to strengthen the company's capacity and ability to combat information operations globally. Microsoft works with EFE to use their EFE Verifica to identify, track and expose emerging information operations spreading across Spanish-speaking communities around the world and Reporters Sans Frontières (RSF) to use their Journalism Trust Initiative (JTI) dataset to proactively promote trusted sources of news around the world that follow transparent and accepted journalistic practices. RSF's JTI data has expanded since our last report and currently includes Palestine, Bulgaria, France, Greece, Germany, Latvia, Poland, and Romania. Microsoft Democracy Forward and RSF are working to expand the dataset into more cities and countries throughout 2024. Indication of impact (at beginning of action: expected impact) including relevant metrics when available

## Reporting on the services' response during an election

## Reporting on the services' response during an election

## **European Elections 2024**

Threats observed or anticipated at time of reporting:

**LinkedIn** is an online professional networking site with a real identity requirement, which means that content posted by our members is visible to that member's professional network, including colleagues, managers, and potential future employers. As a result of LinkedIn's professional context, our members come to LinkedIn for economic opportunity, and as such, do not tend to post misinformation, nor does misinformation content gain traction on LinkedIn. Nonetheless, LinkedIn may be subject to certain members inadvertently posting misinformation during elections.

**Bing Search** expects that there could be attempts to utilize data void manipulation to show low-authority content to unsuspecting users related to elections, as well as more concerted efforts by bad actors attempting to spread mis/disinformation to users across online ecosystems. Bing deploys an automated monitoring on election-related content and potential abuses. Bing also relies on fact-check supports and ingests fact-check and trustworthiness signals into the automated pipelines. Outside of the insights from fact-checking and partner organizations whose signals we continuously ingest, we haven't observed any other threads or manipulation trends

Mitigations in place - or planned - at time of reporting:

LinkedIn's Professional Community Policies expressly prohibit false and misleading content, including <u>misinformation and disinformation</u>, and its in-house Editorial team provides members with trustworthy content regarding global events, including European Elections. LinkedIn has an internal team of hundreds of content reviewers located all over the world (for 24/7 coverage), approximately 235 of which are located in the EMEA region, and includes specialists in a number of languages including English, German, French, Spanish, Portuguese, Russian, and Ukrainian. These reviewers use policies and guidance developed by a dedicated content policy team and experienced lawyers, and work with external fact checkers as needed. When LinkedIn sees content or behaviour that violates its Professional Community Policies, it takes action, including the removal of content or the restriction of an account for repeated abusive behaviour. Further, as noted in Section VIII, LinkedIn is also working on increasing the language coverage of its fact-checking arrangement.

Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue, including European Elections. LinkedIn also does not provide a mechanism for content creators to monetise the content they post on LinkedIn.

**LinkedIn** continues to mature its crisis response processes. In addition to the increase in resource allocation and process improvements, best practices include: 1) quickly coordinating with industry peers regarding the exchange of threat indicators; 2) engaging with external stakeholders regarding trends and TTPs; 3) continuously providing updated policy guidance to internal teams to assist with the removal of misinformation; and 4) continuing to proactively provide localised trustworthy information to our members.

LinkedIn has continued to mature its crisis response playbook by continually monitoring crisis situations globally, expanding internal teams that work on crisis response, and maturing our processes to respond more efficiently and effectively to crisis situations. LinkedIn will continue to follow its processes related to the removal of misinformation, and continually increase investments in resource allocation and process improvements where necessary to respond to the demands of the crisis.

**Bing Search** takes a multifaceted approach to protecting election integrity and regularly updates its processes, policies, and practices to adapt to evolving risks, trends, and technological innovations. This approach includes: (1) defensive search interventions; (2) regular direction of users to high authority, high quality sources; (3) removal of auto suggest and related search terms considered likely to lead users to low authority content; (4) partnerships with research and nonprofit organisations to maintain threat intelligence and inform potential algorithmic interventions and contribute to broader research community; (5) special information panels and answers to direct users to high authority sources concerning elections and voting; (6) internal working groups dedicated to addressing company-wide election initiatives; (7) establishing special election-focused product feature teams; (8) conducting internal research on content provenance and elections; (9) evaluating and undertaking red-team testing for generative AI features with respect to elections and political content; (9) ensuring Responsible AI reviews for all AI features; (10) seeking additional fact checking support in EU member states; (11) undertaking comprehensive risk assessments related to elections and electoral processes; (12) developing and continuing to improve targeted monitoring both for web search and Copilot in Bing; and (13) regularly evaluating whether additional measures, metrics, or mitigations should be implemented. These measures are integrated into Bing Search and Copilot in Bing generative AI experiences, along with the additional safeguards discussed at QRE 14.1.1 and QRE 14.1.2 and other measures discussed throughout this report.

**Microsoft**'s Democracy Forward team continues to expand its collaborations with organizations that provide information on authoritative sources, ensuring that queries about global events will surface reputable sites. Microsoft works with Reporters Without Borders (RSF) and their Journalism Trust Initiative (JTI) data to proactively promote trusted sources of news around the world.

While not announced during the current reporting period, it is worth mentioning that in February 2024, Microsoft and LinkedIn came together with the tech sector at the Munich Security Conference to take a vital step forward against AI deepfakes, which will make it more difficult for bad actors to use legitimate tools to create deepfakes. This focuses on the work of companies that create content generation tools and calls on them to strengthen the safety architecture in AI services by assessing risks and strengthening controls to help prevent abuse. This includes aspects such as ongoing red team analysis, preemptive classifiers, the blocking of abusive prompts, automated testing, and rapid bans of users who abuse the system. The accord brings the tech sector together to detect and respond to deepfakes in elections and will help advance transparency and build societal resilience to deepfakes in elections.

We combined this work with the launch of an expanded Digital Safety Unit. This will extend the work of our existing digital safety team, which has long addressed abusive online content and conduct that impacts children or that promotes extremist violence, among other categories. This team has special ability in responding on a 24/7 basis to weaponized content from mass shootings that we act immediately to remove from our services. The accord's commitments oblige Microsoft and the tech sector to continue to engage with a diverse set of global civil society organizations, academics, and other subject matter experts. These groups and individuals play an indispensable role in the promotion and protection of the world's democracies.

See more at Microsoft announces new steps to help protect elections - Microsoft On the Issues and Meeting the moment: combating AI deepfakes in elections through today's new tech accord - Microsoft On the Issues.

[Note: Signatories are requested to provide information relevant to their particular response to the threats and challenges they observed on their service(s). They ensure that the information below provides an accurate and complete report of their relevant actions. As operational responses to crisis/election situations can vary from service to service, an absence of information should not be considered a priori a shortfall in the way a particular service has responded. Impact metrics are accurate to the best of signatories' abilities to measure them].

## **Policies and Terms and Conditions**

Outline any changes to your policies

Policy	Changes (such as newly introduced policies, edits, adaptation in scope or implementation)	Rationale	
Bing Search	3ing Search		
Authoritative Information sources	Bing Search will empower voters with authoritative election information on Bing. We do this in partnership with organizations that provide information on authoritative sources, ensuring that queries about election administration will surface reputable sites. Bing joined forces with Reporters Without Borders (RSF) and their Journalism Trust Initiative (JTI) data to proactively promote trusted sources of news around the world, including Bulgaria, France, Greece, Germany, Latvia, Poland, and Romania. Microsoft Democracy Forward and RSF are working to expand the dataset into more countries throughout 2024.	See more at <u>RSF to Implement the Journalism Trust Initiative in collaboration</u> with Microsoft   RSF	
Authoritative Information sources	In addition, Bing is working to provide user-facing information panels on election information for the European Elections and directing users to official government sources for these elections, most notably the European Parliament.	See more at <u>Microsoft announces new steps to help protect elections -</u> <u>Microsoft On the Issues</u>	

Microsoft		
Deepfakes and Elections	Microsoft recently launched a new web page – <u>Microsoft-2024</u> <u>Elections</u> – where political candidates can report alleged deepfakes of themselves to Microsoft.	See more at <u>Meeting the moment: combating AI deepfakes in elections</u> through new tech accord
Content Credentials as a Service	Microsoft will help candidates and campaigns maintain greater control over their content and likeness by launching Content Credentials as a Service. This new tool enables users to digitally sign and authenticate media using the <u>Coalition for Content</u> <u>Provenance and Authenticity's (C2PA)</u> digital watermarking credentials, a set of metadata that encode details about the content's provenance using cryptography. Users can attach Content Credentials to their images or videos to show how, when, and by whom the content was created or edited, including if it was generated by Al. Content Credentials as a Service will be initially launched in private preview for political campaigns and election purposes in the United States and will be the first-of-a-kind SaaS application for adding content credentials to images and video.	These watermarking credentials empower an individual or organization to assert that an image or video came from them while protecting against tampering by showing if content was altered after its credentials were created. See more at <u>Microsoft announces new steps to help protect elections - Microsoft On the</u> <u>Issues</u> See more at <u>Meeting the moment: combating AI deepfakes in elections through</u> <u>today's new tech accord</u>
Campaign Success Teams	Microsoft will help political campaigns navigate cybersecurity challenges and the new world of AI by deploying a newly formed "Campaign Success Team" within Microsoft Philanthropies' Tech for Social Impact organization. This team will advise and support campaigns as they navigate the world of AI, combat the spread of cyber influence campaigns, and protect the authenticity of their own content and images.	See more at <u>Microsoft announces new steps to help protect elections - Microsoft</u> <u>On the Issues</u>
AI Elections Trainings	Microsoft's Democracy Forward and Tech Social Impact teams will deliver a series of trainings on election security and AI, including cybersecurity and information security, dedicated to political groups in the European Parliament.	

Communications Hub	Microsoft will create and provide access to a new "Election Communications Hub" to support democratic governments around the world as they build secure and resilient election processes.	See more at <u>Microsoft announces new steps to help protect elections -</u> <u>Microsoft On the Issues</u>
Situation Room	Ahead of the European elections, Microsoft will establish a Situation Room, bringing together resources across the company to monitor, support, and protect the European elections.	
Foreign Malign Influence Updates	Microsoft will also publish regular reports on foreign malign influence researched and reported by the company's MTAC team. The <u>first report</u> was released in November 2023.	See more at <u>Microsoft announces new steps to help protect elections -</u> <u>Microsoft On the Issues</u> See more at <u>Meeting the moment: combating AI deepfakes in elections</u> <u>through new tech accord</u>
	Scrutiny of Ads Plac	cements
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.		
LinkedIn		

Political Advertising			
Outline approaches pertine	Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.		
LinkedIn			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Political ads are banned on LinkedIn, which includes prohibitions on ads that exploit a sensitive political issue, including European Elections. Indication of impact (at beginning of action: expected impact) including relevant metrics when available		
Microsoft Advertising			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Not applicable. As described in Section 3 of the report, Microsoft Advertising does not support political advertising. Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a		
	Integrity of Services		
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.			
LinkedIn			
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention LinkedIn maintains a robust ad-review process to prevent malicious advertising wherein ads related to the European Commission elections are manually reviewed and approved by our internal team. LinkedIn regularly meets with and exchanges information with industry peers to identify and share granular information related to manipulative behaviours, coordinated influence operations, and TTPs. All such content that violates LinkedIn's Professional Community Policies is removed.		

	Indication of impact (at beginning of action: expected impact) including relevant metrics when available
	LinkedIn does not track misinformation content removal by subject matter, and a comprehensive listing of its misinformation content takedowns is provided in this disclosure and in its Transparency Report. See also SLI 18.2.1.
	Impact metrics for malicious advertising removals are available in SLI 2.1.1 and 2.3.1, whereas metrics related to manipulative behaviors and TTPs are available in SLI 12.1.1 and 14.2.1.
Bing Search	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing Search takes a multifaceted approach to protecting election integrity and regularly updates its processes, policies, and practices to adapt to evolving risks, trends, and technological innovations. This approach includes: (1) Defensive search interventions; (2) regular direction of users to high authority, high quality sources; (3) removal of auto suggest and related search terms considered likely to lead users to low authority content; (4) partnerships with research and nonprofit organizations to maintain threat intelligence and inform potential algorithmic interventions; (5) special information panels and answers to direct users to high authority sources concerning elections and voting; (6) internal working groups dedicated to addressing company-wide election initiatives; (7) establishing special election-focused product feature teams; (8) conducting internal research on content provenance and elections; (9) evaluating and undertaking red-team testing for generative AI features with respect to elections and political content; (10) ensuring Responsible AI reviews for all features; (11) seeking additional fact checking support in EU member states; (12) undertaking comprehensive risk assessments related to elections and electoral processes; (13) developing and are continuing to improve targeted monitoring both for web search and Copilot in Bing; (14) conducts additional monitoring of fact check signals and third-party intelligence sources for election related information to inform potential defensive interventions and maintain awareness and preparations for potential disinformation narratives and tactics; and (15) regularly evaluating whether additional measures, along with the additional safeguards discussed at QRE 14.1.1 and QRE 14.1.2 and other measures discussed throughout this report. Indication of impact (at beginning of action: expected impact) including relevant metrics when available Bing does not currently have applicable metrics. Bing hopes to be able to share metrics

Empowering Users	
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.	
LinkedIn	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention LinkedIn has an internal team of global news editors that provides trustworthy and authoritative content to its member-base at all times. During the European Elections, this team will provide manually curated and localised storylines. Indication of impact (at beginning of action: expected impact) including relevant metrics when available
Bing Search	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Partnerships with Election Authorities and Trusted Partners: Bing Search empowers voters with authoritative election information on Bing. We do this in partnership with organizations that provide information on authoritative sources, ensuring that queries about election administration will surface reputable sites. Bing joined forces with Reporters Without Borders (RSF) and their Journalism Trust Initiative (JTI) data to proactively promote trusted sources of news around the world, including Bulgaria, France, Greece, Germany, Latvia, Poland, and Romania. Microsoft Democracy Forward and RSF are working to expand the dataset into more countries throughout 2024. Microsoft continues to build on its partnership with EFE focused on identifying Spanish language influence campaigns. In addition, Bing is working to provide user-facing information panels on election information for European elections.
<b>Specific Action applied</b> (with reference to the Code's relevant	Indication of impact (at beginning of action: expected impact) including relevant metrics when available Description of intervention Bing Search is developing and improving special informational and Answer segments for the European elections to show high-authority information related to the election, such as election dates and information about how to vote (examples of which are provided below).

Commitment and	European elections • 6-9 June 2024	
Measure)	How to vote in Italand	
	How to vote in Ireland	
	The European elections will take place in Ireland in June 2024. Irish and other EU citizens resident in Ireland and aged 18 or over can vote provided they register 15 days before polling day.	
	Learn more Feedback & Q	
	Elecciones europeas • 6-9 de junio de 2024	
	Cómo votar en España	
	elections.europa.eu Las elecciones europeas se celebrarán en España el 9 de junio de 2024. La edad mínima para	
	votar es de 18 años. Pueden votar en ellas los ciudadanos españoles residentes en España o en el extranjero, así como los ciudadanos de otros países de la UE residentes en España.	
	Learn more Feedback 🖉 🖓	
	Bing Search also directs users to official sources related to the European elections, such as https://elections.europa.eu/, and other high authority	
	sources.	
	Indication of impact (at beginning of action: expected impact) including relevant metrics when available	
	Bing does not currently have applicable metrics. Bing hopes to be able to share metrics concerning impressions of recently launched and	
	forthcoming informational features and Answers pertaining to European elections in its next report.	
Empowering the Research Community		
Outline approaches pertinent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.		
Bing Search		
<b>Specific Action applied</b> (with reference to the	Description of intervention	

Code's relevant Commitment and Measure)	Microsoft is undertaking additional research on how users interact with content provenance tools and the use of content provenance tools for Al and elections and other research concerning Al and elections. In addition, Bing engaged third party experts to support internal evaluations of misinformation risks related to image generation features.         Indication of impact (at beginning of action: expected impact) including relevant metrics when available         Such research helps to inform Bing and Copilot safety mitigations.
	Empowering the Fact-Checking Community
Outline approaches pertine	ent to this chapter, highlighting similarities/commonalities and differences with regular enforcement.
LinkedIn	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention         LinkedIn works with globally-recognised factcheckers to receive their independent judgment regarding user-generated content. Content that violates LinkedIn's policy on false and misleading content is removed.         Indication of impact (at beginning of action: expected impact) including relevant metrics when available         LinkedIn does not separately track election related content sent to external fact checkers. Further detail regarding content sent to fact checkers is available at SLI 21.1.1, 30.1.1, and 31.1.1.
Bing Search	
<b>Specific Action applied</b> (with reference to the Code's relevant Commitment and Measure)	Description of intervention Bing Search is formalizing an additional fact-check agreement covering EU member states and languages, which includes localized coverage of the European elections. Bing continues to explore additional fact-checking partnerships and opportunities. Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a

Microsoft	
Specific Action applied (with reference to the Code's relevant Commitment and Measure)	Description of intervention In December 2023, the Institute for Nonprofit News teamed up with DeleteMe, funded by Microsoft's Democracy Forward Initiative, to help decrease the amount of personal information about journalists that can be discovered online. The collaboration protects local journalists from threats during a critical election year and ensures they can provide high-quality information to local communities. The <u>Microsoft Journalism Hub</u> is an ongoing resource center built to connect the journalism community more seamlessly with tools, technology, services, and partner programs designed to support their unique business needs. These resources are distributed across Microsoft's three journalism initiative commitments: rebuilding capacity in local news ecosystems; restoring trust in news; and reducing legal and cyber risks for journalists. Microsoft has partnered with Nota to allow journalists to access helpful AI tools for tasks such as improving SEO, creating videos of published stories, and writing headlines. Nota's assistive AI tools build time-saving production infrastructure for digital publishing—creating an accessible engagement economy, offering instant content conversion and content amplification, multiplying video content revenue and doing away with repetitive tasks that have long required journalists to act as generalists in multiple fields. Indication of impact (at beginning of action: expected impact) including relevant metrics when available n/a