Recurrent and prominent systemic risks in the EU and measures for their mitigation

Working paper for the preparation of the systemic risks report by the European Board for Digital Services and the European Commission.

March 2026

Submitted by

European Fact-Checking Standards Network

4 rue Belgrand
75020 Paris
France
efcsn.com
EU Transparency Register
Number: 847550548807-30

Executive summary

This working paper, submitted by the European Fact-Checking Standards Network (EFCSN), details some of the recurrent and prominent systemic risks posed by misinformation and disinformation on Very Large Online Platforms (VLOPs) and Search Engines (VLOSEs), risk factors influencing these risks and an evaluation of potential mitigation measures.

The report highlights how deceptive content as well as false or misleading information undermines democratic processes, public health, and fundamental rights across Europe. Research and investigations by EFCSN members identify five prominent risk areas:

- **Electoral Integrity:** Coordinated inauthentic behavior and AI-generated content have targeted elections in Denmark, Hungary, Romania, and Moldova.
- **Scams:** Platforms frequently profit from scam advertisements; for example, one member reported over 4,000 unique scam links to Meta in a nine-month period.
- **Climate Disinformation:** Repetitive narratives designed to erode trust in climate science; disinformation can outperform accurate information in engagement metrics.
- **Health Misinformation:** AI-generated "experts" are increasingly used to bypass regulations and promote unproven treatments.
- **Non-Consensual Image Abuse:** The proliferation of deepfakes targeting both public figures and private individuals undermines privacy and dignity.

Risk Factors

There is a significant "misinformation premium" where low-credibility accounts receive substantially higher engagement than high-credibility sources. On YouTube, low-credibility accounts receive 8× more interactions; on Facebook, the ratio is 7×.

Other risk factors are: AI chatbots and AI overviews that have been documented reinforcing and amplifying misleading narratives, and monetization schemes that may incentivize the production and dissemination of disinformation.

Evaluation of Mitigation Measures

The EFCSN expresses concern over the reduction of effective mitigation tools:

- **Cuts to Fact-Checking:** Meta has reduced remuneration for third-party fact-checkers by approximately 25% for 2026, despite the proven efficacy of the program in reducing the spread of false content.
- **Google Search Changes:** The removal of the "fact-checking snippet" in June 2025 has reduced the visibility of verified data and, in some cases, led to snippets that suggest the opposite of the actual fact-check conclusion.
- **Limitations of Community Notes:** Current crowd-based systems are often too slow, vulnerable to manipulation by small groups, and ill-suited for highly polarized topics.

Importantly, this paper is not comprehensive. There are likely to be more prominent and recurring systemic risks posed by VLOPSEs, there certainly are more risk factors (e.g. failures in enforcement of content moderation policies, intransparent recommender systems, lack of age verification, to name a few), and risk mitigation measures should be evaluated more in–depth and comprehensively than we are able to in this report. We focus instead on aspects that our community is involved with and has a deep understanding of. Time and resource constraints did not allow a more in-depth work. The EFCSN has recently summarised its view of European information integrity in this report.

Content

Q1: Evidence of potentially prominent or recurrent systemic risks on VLOPs and VLOSEs

Research, investigations and fact-checking carried out by EFCSN member organizations has extensively documented the prominent and recurrent systemic risks misinformation and disinformation pose to European societies and individuals. The study "Measuring the State of Online Disinformation in Europe on Very Large Online Platforms", led by EFCSN member Science Feedback, demonstrated the concerning share of misinformation on VLOPs: TikTok shows the highest prevalence of misinformation aggregated across four languages (~20% of exposure-weighted posts on public-interest topics), followed by Facebook (~13%) and X/Twitter (~11%). Instagram and YouTube are lower (~8%), and LinkedIn is lowest (~2%). A second iteration of the same study[1] which will be published after this paper has been submitted confirms the results. It even finds an increase of misinformation prevalence on some platforms.

Misinformation and disinformation pose significant and persistent threats to Europe, undermining democratic processes, social cohesion, public health, individual well-being and fundamental rights. These phenomena are not isolated risks but manifest across all four systemic risk categories outlined in Article 34(1) of the Digital Services Act (DSA):

> (a) the dissemination of illegal content, as false or misleading narratives often incite hatred, violence, and extremism, or might contain illegal defamation. Illegal content such as non-consensual image abuse material may also infringe on privacy and image rights. Lastly, scams and frauds fall in the illegal content category and can also be classified as disinformation (i.e. false or misleading information produced, published and distributed with the intent to cause harm).
> (b) the protection of fundamental rights, as disinformation campaigns erode trust in institutions, manipulate public discourse, and contribute to discrimination and marginalization. By flooding communication spaces with falsehoods or inauthentic content, the freedom of expression of EU citizens is unfairly limited as their speech might be drowned out by inauthentic speech (e.g. from bots or sockpuppets).
> (c) electoral processes and civic discourse, as misinformation and disinformation distort political debates, mislead voters, and weaken democratic participation. It is important to note that this systemic risk should not be limited to election periods and campaigns. Disinformation is a continuous challenge for European societies, weaknesses that are exploited during election periods, such as low trust in institutions or susceptibility to conspiracies, have often been caused by a constant drip of disinformation over the years; and
> (d) public security, as the spread of deceptive content undermines crisis response, from health emergencies and disaster response to geopolitical conflicts.

[1] Vincent EM & Crisan D (2026) Second Measurement of the State of Online Disinformation in Europe on Very Large Online Platforms. Second report of the SIMODS project (Structural Indicators to Monitor Online Disinformation Scientifically). Science Feedback. Results will be accessible via https://science.feedback.org/second-measurement-mis-disinformation-major-platforms-europe/ from March 19 onwards.

Given the scale and persistence of these risks, robust enforcement of the DSA is critical to mitigating the societal harm caused by misinformation and disinformation across the European digital ecosystem.

In the following sections, we summarize and link to research, investigations and fact-checking done by our member organizations that proves the existence of the systemic risk of misinformation and disinformation. To provide additional context, we also reference other investigations and studies.

Nonetheless, our reporting very likely only covers the tip of the iceberg. Various factors prevent our community as well as other stakeholders from gaining a complete picture of the prevalence and impact of mis- and disinformation spread through VLOPSEs, among them a lack of comprehensive monitoring (due to lack of resources and insufficient data access) as well as the covert nature of disinformation campaigns. The fact-checking community is well-positioned to mitigate this problem as we are working to pool data on fact-checks and debunked claims in a centralized European repository. Such a repository will likely become operational in 2026.

Risk #1: Electoral Integrity

Monitoring the information environment before elections is essential to enable early detection of misinformation and foreign interference that could distort voter perception. Identifying disinformation or information manipulation can help mitigate their spread, ensuring voters make informed decisions based on facts rather than manipulation. Proactive monitoring also strengthens transparency, holds actors accountable, and protects the fairness and legitimacy of electoral outcomes. EFCSN member organisations contribute to this monitoring extensively and have documented several instances in which electoral integrity was being targeted.

Denmark, 2026
Misleading claims targeting election integrity sometimes exploit religious and cultural narratives to sow division and influence voter behavior. This is what EFCSN member TjekDet revealed ahead of the upcoming Parliamentary election in Denmark. TjekDet fact-checked claims that have surfaced, asserting that Muslims are prohibited from voting, a narrative linked to fringe groups like Hizb ut-Tahrir. These campaigns employ tactics such as selective interpretation of religious texts, decontextualization, and the strategic use of social media to amplify divisive messages.

Hungary, 2026
In Hungary EFCSN member Lakmusz.hu is registering several manipulation attempts to sway the electorate ahead of the parliamentary vote in April. For example, AI-generated videos spread by obscure Facebook pages trying to instill fear of war, implying that the EU's support of Ukraine may lead to Hungarians suffering.

Lack of enforcement of Meta's self-imposed political ad ban is also an issue: Lakmusz.hu reported that 14 of the 106 Fidesz candidates placed 181 ads in January 2026. Of these, Meta classified only 19 as political, while the other 162 passed through filters without any problems.

> *"This is not the first time that pro-government actors have circumvented Meta's rules: the National Resistance Movement (NEM) constantly places political ads on the platform, with their videos generally aimed at discrediting Péter Magyar, the main opposition leader. One recent video, which ran as a political ad during Christmas season, received 17 million views." From reporting by Lakmusz.hu*

Romania, 2025

During the 2025 Romanian presidential elections, EFCSN member Funky Citizens exposed a systematic disinformation campaign characterized by coordinated inauthentic behavior (CIB) and fabricated narratives.

False claims that the second round was unjustly annulled, supported by fabricated legal rulings and fake foreign endorsements, dominated online discourse, portraying Romanian institutions and international powers as conspiring to suppress democracy and fueling public anxiety about war and economic collapse.

Candidates were targeted with personalized smear campaigns, such as Nicușor Dan being falsely linked to the former Securitate, while conspiracy theories about the WHO pandemic treaty, EU dictatorship, and foreign interference proliferated.

Platform manipulation was rampant, with Funky Citizens identifying hundreds of accounts exhibiting suspicious, bot-like activity—posting identical, repetitive messages in multiple languages and from locations unrelated to Romania, exploiting the lack of transparency on platforms like TikTok and Metaedmo.eu. The information landscape was further poisoned by antisemitic tropes and false claims of voter fraud, illegal voting booth footage, and premature victory declarations. Funky Citizens noted that "the presence of commenters from countries with no ties to Romania and the use of rare languages such as Amharic or Khmer indicates a deliberate strategy of inauthentic engagement," and that "the information landscape has become a battlefield dominated by disinformation".

Moldova, September 2025

EFCSN member Demagog closely monitored the elections in Moldova and identified Russian interference through coordinated online campaigns aimed at influencing the electoral environment. According to the investigation, Russia sought to shape the information space through propaganda, disinformation, and coordinated activities supporting pro-Russian interests. The operation included the use of Facebook advertisements targeting Moldovan diaspora communities, recruiting individuals to act as "independent election observers" in exchange for financial compensation. The collected reports were allegedly centralised on infrastructure connected to entities operating in the separatist region of Transnistria.

Risk #2: Scams

Scams and schemes to defraud unwitting victims that are facilitated via platform services pose a significant recurring and prominent systemic risk to the safety of European users. The scale of the problem is enormous and oftentimes scams are promoted using VLOPs own advertising systems, i.e. platforms even profit from this systemic risk.

According to a report by Reuters Meta's total global profit off of scam ads is in the billions. Our members' investigations find endless examples of scam posts; from March 24 to December 31, 2025, EFCSN member Greece Fact Check, a Trusted Flagger in Greece, reported 4,177 unique links to Meta alone. The majority of these links were sponsored content and the content was almost exclusively financial and medical scams. The organisation stated towards us, they could have reported more if they had time and human resources to do so.

There are many other examples of scam detection and debunking from our network, e.g. a debunk an offer of a free espresso machine in Romania (Funky Citizens) or warnings about handmade products supposedly made by the elderly using AI-generated content in France, Slovakia and Greece that are actually cheap goods from China (AFP). In fact, when we consulted our members, they told us that they produce dozens of fact-checks warning their audiences about scams, with many having a dedicated section of their website (Fact Check Cyrus, Hibrid.info, Infoveritas, among others). In several cases, the likenesses of prominent figures were instrumentalized for scam ads, for example, on Facebook in Denmark.

Risk #3: Climate change-related disinformation

Climate change-related disinformation undermines public support for policy responses to climate change, may discredit innovative technologies and distort public discourse about the issue.

The EFCSN and 21 of its member organisations worked alongside TU Dortmund, Newtral, Maldita.es, Science Feedback, and Ontotext on a long term project, concluded in 2025, which tracked and analyzed climate-related misinformation across Europe. Over the course of the project, 30 EFCSN member organisations collectively catalogued over 1,100 climate-related fact-checks, structured more than 80,000 data points, and identified 20 cross-border narratives and 52 sub-narratives. The research revealed that climate disinformation rarely takes the form of a single viral post — instead, it spreads through many minor, repeating storylines that collectively shape public perception and erode trust in climate science.

With regards to misinformation about electric vehicles, fact-checkers found that the same falsehoods repeatedly appeared across countries and platforms, a clear sign that climate disinformation ignores borders. The project found that disinformation in these areas thrives through strategic doubt, distortion, and emotional appeal rather than outright denial, exploiting country-specific anxieties about personal freedom, economic costs, and national identity.

The role of the platforms in amplifying these narratives is a growing concern. A study by EFCSN member Fundación Maldita.es and AI Forensics, analysing disinformation in the aftermath of the 2024 Valencia floods, found that videos featuring mis- and disinformation garnered over 21 million views on YouTube and TikTok combined — with disinformation content significantly outperforming accurate information in engagement metrics on both platforms. Fewer than one in four videos on YouTube containing disinformation carried any warning label, while on TikTok none of the climate disinformation identified had a warning attached — this despite TikTok's stated policy banning content that denies the existence of climate change. These findings point to a systemic failure by major platforms to enforce their Terms of Service and they indicate that recommender systems and platform architectures might favour low quality and disinformation content.

Risk #4: Health misinformation

Health misinformation poses a threat not only to individuals' health and wellbeing but also to public health more broadly, and it carries steep economic costs. The most recurring topics exposed to misinformation are cancer treatments, vaccines, their effectiveness, origin and contents and so called wellness trends promoted on social media according to qualitative feedback from our members.

The consequences are tangible and severe: false claims have fuelled hesitancy around life-saving vaccines, led patients to delay or reject evidence-based treatments in favour of unproven alternatives, and generated significant economic costs — including an estimated $2 billion in additional healthcare expenditure linked to COVID-19 vaccine hesitancy alone.

Generative AI is fuelling this systemic risk as well. EFCSN member Science Feedback recently uncovered a sophisticated operation using synthetic "experts" to bypass regulations on platforms such as TikTok, where AI avatars dispense questionable medical advice for engagement (a similar manipulative scheme was uncovered in Germany), and YouTube, where channels directly impersonate real doctors by using their names and likenesses to mislead vulnerable audiences. Similarly, the Pravda Association uncovered a network of at least 30 social media profiles across Europe using AI-generated "experts" to sell unproven dietary supplements. Full Fact has also highlighted how these deceptive tactics are used to market weight-loss products. Their investigation reveals a surge in accounts on platforms such as Facebook that use AI-generated doctors to sell patches by promising impossible results.

Risk #5: Non-consensual image abuse

The abuse of the likenesses of real people in AI-generated content and the reach this kind of content gains on VLOPs poses a prominent and recurring risk. This phenomenon has been documented extensively as it affects public figures as well as ordinary people. High-profile figures such as Marine Le Pen, French President Emmanuel Macron or Irish presidential candidate Catherine Connolly have confronted a proliferation of AI manipulations on platforms, which depict them saying fictitious things to sow confusion or discredit their positions. Deepfakes of politicians have also surfaced in Italy (here and

here), and Hungary. Such political deepfakes are persuasive and are difficult to correct (Barari et al., 2025). Generally, AI-generated content is becoming ever more prevalent. In December 2025, survey results from the European Digital Media Observatory (EDMO) showed that 16% of all fact-checked articles in Europe focused on AI-related content, a new record. GenAI drastically lowers the cost and time required to produce persuasive, emotionally charged, and realistic-looking content tied to current events.

Additionally, non-consensual AI-generated content, specifically through 'nudifier' applications, continues to undermine privacy rights, dignity and personal safety. It likely classifies as illegal content in most instances.

Q2: Risk factors

A study led by EFCSN member Science Feedback analysed data across six VLOPs (Facebook, Instagram, LinkedIn, TikTok, X/Twitter, YouTube) and four EU Member States (France, Poland, Slovakia, Spain). They found a substantial 'misinformation premium', where low-credibility accounts consistently outperform high-credibility sources in user engagement. Across most platforms, low-credibility accounts enjoy a clear interaction advantage. To ensure a fair comparison, researchers measured interactions (reactions, comments, and reshares) per post per 1,000 followers:

- **YouTube:** Low-credibility accounts receive, on average, **8×** more interactions than high-credibility accounts.
- **Facebook:** The ratio is approximately **7×**.
- **Instagram & X:** The ratio is approximately **5×**.
- **TikTok:** The ratio is approximately **2×**.
- **LinkedIn:** The only platform studied where the difference was not statistically significant.

Datasets studied were either platform-provided (LinkedIn) or the result of large-scale keyword searches on high-salience topics (Ukraine/Russia, climate, health, migration, national politics). The corpus covered ~2.6 million posts totalling ~24 billion views.

These findings indicate that the design and mechanisms of platforms can not only facilitate but amplify the spread of disinformation, as other research has also demonstrated.

Monetization schemes

The monetization of polarizing and misleading content on social media is driven by engagement-based revenue models that reward virality rather than accuracy. These incentive structures favour emotionally charged and divisive narratives, which are systematically amplified through algorithmic recommendation systems.

An investigation by EFCSN member Maldita.es identified 550 TikTok accounts that posted over 5,800 AI-generated videos of fake protests, designed to exploit emotional and political themes for monetization. These accounts used generative AI tools to create realistic but false scenarios, such as "Venezuelans celebrating Maduro's capture" or "Iranians protesting for freedom," to drive engagement and qualify for TikTok's creator monetization program. The content was not only polarizing but also directly violated platform policies, yet enforcement was inconsistent or absent.

Another Maldita.es investigation exposed a network of 15 YouTube channels that used deepfake technology to impersonate political analyst John Mearsheimer, creating over 200 videos. While not always visibly monetized, these channels operated in a "value accumulation phase," building audiences for future monetization or sale, and avoiding detection by staying under the platform's radar. The goal was to create reusable assets for long-term profit, regardless of the misinformation spread.

Chatbots & AI overviews

In the first week of the war in Iran, EFCSN member Correctiv, debunked misleading claims circulating online after footage of an attack on a school in Minab, Iran. The case illustrates several common dynamics in crisis-related disinformation. Authentic visual material was falsely attributed to a different event, a form of miscontextualised content frequently used to cast doubt on verified reporting, while the claim spread rapidly through social media posts. The narrative was further amplified when the AI chatbot Grok repeated the allegation and presented incorrect supporting evidence.

The episode highlights how AI systems embedded in online platforms can unintentionally reproduce and reinforce misleading narratives circulating on the platform itself, while their authoritative tone may increase the perceived credibility of false claims. A similar case has been documented for Google Lens AI overviews by EFCSN member organisation Full Fact. Even when such systems later correct their responses, the initial misinformation may already have circulated widely, contributing to the rapid spread of misleading information during fast-moving crises.

Efficacy and risks of community notes

Considering available academic literature, investigations into and reports about community notes, we must conclude that the approach is ill-suited for countering misinformation and disinformation when it is not complemented by other measures, for the following reasons:

1. **Meta Community Notes have no impact:** in its first six months in the US, by Meta's own account, only around 900 Community Notes became visible. In a similar period, again according to Meta's own public numbers, Facebook alone labeled around 35 million Facebook posts in the European Union alone, using 150,000 individual articles available to the platform through its partnerships with European fact-checking organisations.
2. **Only a small fraction of submitted notes are ever shown publicly:** it is not that users do not care, it is that the notes they submit almost never become visible

because the system is built to ensure little impact. On X, that first adopted the bridging algorithm Meta relies on to decide which notes become visible, less than 12% of all notes proposed ever become visible (Bouchaud & Ramaciotti, 2025).

3. **Community Notes are more and more often AI-generated because users get tired of proposing notes that never show:** According to a recent study the number of monthly active authors on X's Community Notes (defined as users who contributed at least one note a month) has dropped significantly in 2025 (Arjmandi-Lari et al., 2025). X is trying to make up for this loss by having AI propose and write the notes. Thus, community notes might turn into "machine notes" with minimal human involvement, opening the system up for LLM-specific vulnerabilities, such as partisanship (Renault et al., 2026).
4. **Current Community Notes programs are irrelevant for polarizing or politicized topics.** The "bridge-based" design creates a significant bottleneck for polarizing issues, as notes regarding divisive topics or prominent figures rarely achieve the necessary agreement to reach "Helpful" status. This effect has been demonstrated across different countries and contexts (Bouchaud & Ramaciotti, 2025).
5. **Community Notes can be manipulated** Razuvayevskaya et al. (2025) found that the top 10% of contributors produced 58% of all notes. Based on a simulation, Truong et al. (2025) concluded that "a small minority (5–20%) of bad raters can strategically suppress targeted helpful notes". Small language areas in which there is only a limited number of contributors, might therefore be at an even greater risk.
6. **Community Notes are too slow** to curb the spread of misinformation. Razuvayevskaya et al. (2025) found that "notes, on average, are published 65.7 hours after the original post, with longer delays significantly reducing the likelihood of consensus".
7. Meta has previously announced that **users will not be able to apply Community Notes to advertisements**. This exclusion creates a **powerful incentive for disinformation actors** to simply promote their content as ads—ensuring broader reach and immunity from public correction—while simultaneously generating revenue for Meta. Additionally, it shields scammers and fraudsters from this intervention. Given the enormous scale at which ads promoting scams and frauds are being proliferated, this policy choice should be reversed (Horwitz, 2025).
8. Meta has also stated that posts tagged with Community Notes will not be downranked in feeds. This creates a serious loophole: bad actors can repeatedly spread falsehoods, and even if their posts are eventually annotated by users, they **may continue to receive algorithmic amplification and reach**. To our knowledge, being labelled with community notes also does not impede on an account's or page's eligibility for Meta's monetisation programs. This weakens any deterrent against coordinated or habitual disinformation.

End of fact-checking snippet in Google Search results

In June 2025 Google announced in a developer blog that it would no longer display the dedicated "fact-checking snippet" in search results. According to the announcement, Google's internal data indicated that the snippet was "not commonly used in Search" and

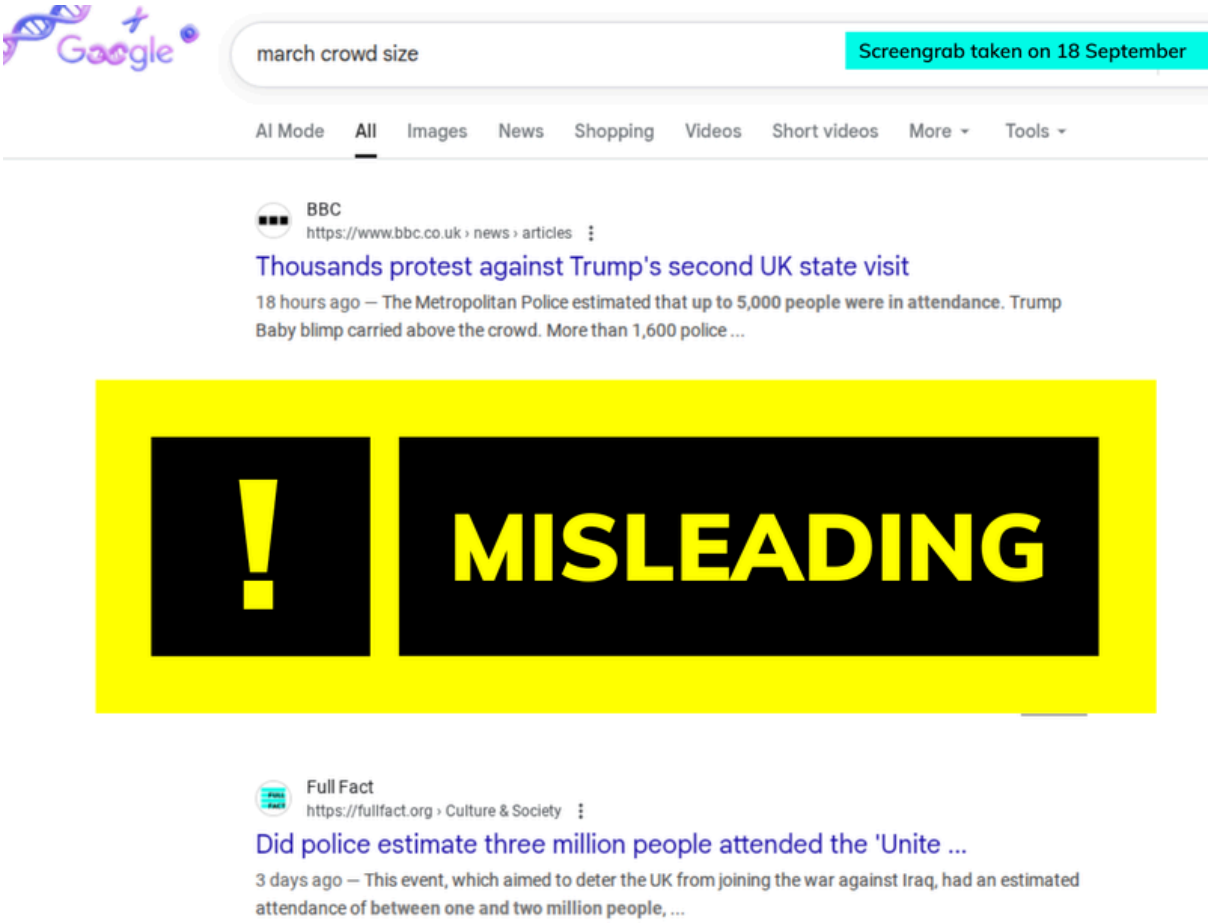
no longer provides “significant additional value for users.” The decision was made without prior consultation with the fact-checking community.

In a transparency report under the then Code of Practice on Disinformation in 2024, Google reported the snippet had been used to serve fact-check entries over 120 million times within the first six months in the EU alone. Furthermore, the Digital News Report by the Reuters Institute found that globally about 25 % of online news consumers say they look for a fact-check to verify information; in Europe the average was around 23 %.

In our view, this change reduces the visibility of structured fact-check content in a key discovery channel. The removal of the snippet means that fact-check articles may now appear as ordinary links rather than being presented in a clear claim-verdict schema.

Additionally, a few months after the announcement an analysis by EFCSN member Full Fact found that ordinary search result snippets sometimes distorts fact-check content, essentially exposing users to misinformation in some cases.

Full Fact reported that the crowd size of the “Unite the Kingdom” rally in London was widely exaggerated, with organisers claiming three million attendees while police estimates placed the number between 110,000 and 150,000. Despite Full Fact’s verified correction of the data, Google’s search snippet extracted a fragment from the article out of context, misleadingly suggesting attendance of “over a million” and incorrectly linking the figure to the rally.



Full Fact's further review found that in approximately 10–15 % of searches on recent fact-checks from their site, Google offered snippets that explicitly suggested the opposite conclusion of the fact-check itself. Google's stated response was that it aims "to surface relevant, high-quality information in all our Search features" and that "when issues arise … we use those examples to improve and take appropriate action under our policies."

Overall, these observations point to a growing concern: when search engine presentation mechanisms (snippets, rich-results) mis-represent or fail to accurately encode the output of independent fact-checkers, the public dissemination of correct information is undermined at the "first glance" stage. The user may draw a wrong inference before even clicking through, simply because the snippet incorrectly frames the claim-verdict structure. The removal of structured metadata support for fact-checks further raises the barrier to ensuring clarity and consistency for users. This issue deserves closer scrutiny and additional research.

Q3: Risk mitigation measures

Meta's Third-Party Fact-Checking Program

Meta's Third-Party Fact-Checking Program is a collaborative initiative launched to combat misinformation across its platforms. By partnering with independent fact-checking organizations certified through the EFCSN or the International Fact-Checking Network (IFCN), Meta aimed to enhance the accuracy of information shared online.

Program Mechanics and Impact

Independent fact-checkers assess the accuracy of content by conducting original reporting, consulting public data, and analyzing media, including photos and videos. When content is rated as false, Meta significantly reduces its distribution, labels it accordingly, and notifies users attempting to share it. Notably, fact-checkers do not remove content, accounts, or pages; removal occurs only when content violates Meta's Community Standards, separate from the fact-checking program.

The positive impact of the program was acknowledged by Meta who previously reported that more than 53-62% of reshare attempts were not completed on fact-checked content (see Meta Code of Practice Transparency Reporting, H1 2025), and when a fact-check label was placed on a post, 95% of users did not click through to view it (source). These statistics underscore the program's effectiveness in curbing the spread of misinformation. Nevertheless, in his January 2025 speech, Mark Zuckerberg stated that Meta would end the fact-checking program in the US and roll out community notes.

Furthermore, Meta's September 2024 Digital Services Act (DSA) Transparency Reports[2] for Facebook and Instagram reveal that only about 3-4% of fact-checked misinformation that was challenged by users resulted in content being reinstated. In contrast, for other categories, over 60% of appealed content was restored. This disparity highlights the robustness of the fact-checking process and the accuracy of the interventions made by fact-checkers.

Scientific Validation

Research supports the efficacy of fact-checking labels. A 2023 research review by MIT researchers Cameron Martel and David G. Rand concluded that warning labels are reliably effective at reducing belief in, and spread of, false content online. Another study, published in Nature Human Behavior, found that warning labels on average led to a 27.6% reduction in belief and a 24.7% reduction in sharing of false headlines. Importantly, these effects persisted (to a slightly lesser extent) even among individuals with low trust in fact-checkers, indicating the broad applicability of such interventions.

Cuts to the program

[2] To our knowledge these data points are not available in more recent Transparency Reports.

However, the program has been quietly reduced. According to reports, the maximum number of fact-checks a third-party fact-checker can get remunerated for was cut by around 25% on average in Europe for 2026 contracts. This will likely reduce the effectiveness of the program as less articles written and less fact-checking data ingested into Meta's system will equal less pieces of false or misleading content labelled and downranked. The likely result will be a greater prevalence of misinformation on Meta's platforms

In summary, Meta's Third-Party Fact-Checking Program serves as a significant measure in mitigating various systemic risks associated with misinformation and disinformation, aligning with the objectives of the Digital Services Act. However, recent cuts to the program and the roll-out of largely ineffective community notes are undermining the program.

Complementing professional fact-checking with community notes

Considering the prevalence of disinformation online and its detrimental effects, it would be appropriate for digital platforms to run a fact-checking program and an additional community notes program. Such programs could either be independent of one another or they could be integrated. The EFCSN has proposed such an integrated approach publicly already.

Crowd-based notes systems can support fact-checking in various ways, keeping in mind that time and accuracy are both of the essence: by suggesting verified, high quality notes themselves; by checking the crowds' notes to accelerate their visibility to other users; by offering innovative, technical solutions to quickly match claims and verified notes.

Building on prior work by EFCSN member organisation Lead Stories presented at Global Fact 2025, we propose seven recommendations to successfully integrate professional fact-checking with a community notes approach:

- **A "fast lane" for fact-checkers**: A consensus voting system may suppress valid notes if they benefit one side of a partisan debate too much. A fast lane for fact-checkers could mitigate this effect, e.g. by auto-approving notes by fact checkers, weighing their votes more heavily, or having fact-checkers crosscheck each other's notes.
- **Fact-checking crowd-based notes**: Sometimes user-based notes make false claims. The consensus voting system used by platforms may not stop the publication of such notes, because accuracy is not central to their voting system. Fact-checkers could review such cases and submit a reliable rating.
- **Quick access for better results**: User notes can be an early warning to a content problem, and some of them provide useful clues, but professional fact-checkers are better equipped, after years of experience, at collating, researching, evaluating and summarizing the available evidence into a clear conclusion. Thus, giving fact-checkers access to all the proposed or requested notes in real time can improve situational awareness and lower response times.

- **Impact at scale with more information**: The neutral and informative way in which community notes are displayed, as well as the efficiency of AI-powered algorithms, could help to match verified information to claims more effectively and at scale. Adding more information is always preferable over moderation tools like takedowns.
- **Transparency on partnerships**: Notes could be a great opportunity to make the partnerships between platforms and independent fact-checkers more transparent to the users, and linking back to, whenever possible, in-app fact-checking videos, articles or other content.
- **Avoid bias**: Work with professional fact-checkers who are part of networks such as the EFCSN adhere to standards on non-partisanship and independence. They must prove that they are compliant with these principles when applying to be certified. They also advocate for all types of content to be tackled by any sort of labelling (including political speech, when false, and ads).
- **Independence is key**: Making sure that fact-checkers are fairly remunerated for the work provided while remaining editorially independent is a way to ensure sustainable, reliable quality of the online service and information access.

About the EFCSN

The EFCSN is the voice of European fact-checkers who uphold and promote the highest standards of fact-checking and media literacy in their effort to combat misinformation for the public benefit. The EFCSN and its verified members are committed to upholding the principles of freedom of expression. They work to promote the public's access to fact-checked trustworthy data and information and to educate the public in how to assess the veracity of information in the public sphere. Our network currently represents 60 organizations from 36 European countries. A visual representation of our network can be found below.

Fact-checking organizations and fact-checking units within larger media organizations can become EFCSN members by undergoing a rigorous evaluation process. The EFCSN Code of Standards sets out in detail the standards member organizations have to comply with, e.g. in terms of rigorous application of the fact-checking methodology, ethical standards and transparency obligations. Compliance with the Code is ensured through a rigorous certification process. Applications are reviewed by two independent assessors, usually academics, scholars or experienced media practitioners with proven expertise in the fields of fact-checking, journalism and disinformation. Based on the reviews of the independent assessors, the Governance Body decides whether an applicant organization is approved as a member or not. Each approved member may display the EFCSN badge on its website.

To ensure that high standards are upheld over time, membership expires after two years. To renew their membership, fact-checking organizations have to reapply and must undergo the same rigorous assessment process as outlined above.