Public comment by the European Fact-Checking Standards Network on third-party fact-checking

1. How does third-party fact-checking impact freedom of expression and other human rights in discussions of public importance?

Clarifying the Nature and Function of the 3PFC Program

A central feature of Meta's third-party fact-checking (3PFC) program is that it labels, rather than removes, content assessed as false or misleading by independent fact-checkers. With limited exceptions (content directly linked to physical harm or election or census interference) the program does not result in content takedowns. To our knowledge, Meta's systems automatically send notifications to users who have previously shared or are about to share a post containing false information, along with additional context from third-party fact-checkers. This strengthens the factual basis of public debate on Meta's platforms. This retrospective, not proactive approach does not amount to censorship, but rather contributes to a richer information environment by adding context and verified information to public discourse.

A key mechanism within this framework is content demotion. When fact-checkers rate a post as false, Meta reduces its distribution, i.e. limiting its reach without removing it. This approach ensures that false content remains accessible, but is not amplified by recommendation systems. It prevents misinformation from crowding out more accurate content in users' feeds. In doing so, demotions strike a necessary balance between rights: they respect the speaker's ability to publish content even if it is demonstrably false while protecting the audience's right to accurate information. Demotions are subject to transparency and appeal requirements set forth by the DSA in the same way that removals are.

Alignment with International Human Rights Standards

Third-party fact-checking is aligned with international human rights instruments, namely the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR). These rights encompass both expression and access to reliable information.

Widespread mis- and disinformation undermine these rights by distorting public discourse, misleading citizens, and drowning out accurate information. By mitigating the reach and impact of false and misleading claims, fact-checking directly supports the right to freedom of information and the ability of citizens to form opinions freely.

It is also important to stress that freedom of expression is not absolute. As per Article 19(3) of the ICCPR, it may be subject to restrictions that are:

> *"provided by law and are necessary: (a) For respect of the rights or reputations of others; [or] (b) For the protection of national security or of public order (ordre public), or of public health or morals."*

The 3PFC program's scope and targeted interventions remain well within these boundaries. It does not restrict speech, but instead balances individual rights with the collective need for a truthful, safe, and democratic information environment.

Human Rights Protected Through Fact-Checking

Fact-checking and the 3PFC program do more than safeguard expression. By correcting false and misleading claims, thereby mitigating spread and impact, it also protects other human rights, including:

- **Right to Free and Fair Elections** (Article 25(b) of the ICCRP) Mis- and disinformation undermine electoral integrity by distorting political debate, suppressing voter turnout, or spreading falsehoods about voting procedures.
- **Right to Health** (Article 12 of the International Covenant on Economic, Social and Cultural Rights): Health misinformation can endanger lives by promoting false cures or anti-vaccine rhetoric. Fact-checking, including through the 3PFC program, corrects health misinformation and supports the public's right to make informed health choices.
- **Right to Reputation and Honor** (Article 17 of the ICCPR): False claims can cause direct harm to individuals' dignity and livelihoods. Fact-checking exposes and helps reduce defamatory or dehumanizing falsehoods.
- **Public Order and Safety:** Disinformation can fuel public disorder, panic, or violence as seen in events like the Southport riots or Valencia floods, where false online narratives escalated tensions. Fact-checking mitigates these risks and upholds the right to peace and security.
- **Protection from Fraud and Unlawful Conduct:** False claims and deception can be used to commit fraud such as scams or financial manipulation which pose tangible harm. Fact-checkers address these threats also through the 3PFC.

EFCSN and the Integrity of Third-Party Fact-Checking

Meta's 3PFC program rests on the credibility and independence of its participating fact-checkers. Third-party fact-checking ensures freedom to access accurate information thanks to a robust and comprehensive self-regulatory compliance structure, represented by intermediary bodies like the EFCSN and the IFCN. EFCSN membership is one prerequisite for European fact-checking organizations to join the Meta fact-checking program.

Fact-checking organizations and fact-checking units within larger media organizations are or can become EFCSN members by undergoing a thorough and comprehensive evaluation process. The EFCSN Code of Standards sets out in detail the standards member organizations have to comply with, e.g. in terms of fact-checking methodology, ethical standards and transparency obligations. Compliance with the Code is ensured through a rigorous assessment process, which sees a detailed questionnaire and the editorial output of an organisation reviewed by two independent assessors who are experts in the fields of fact-checking, journalism and disinformation. The entire assessment process is explained in more detail on our website. Approved members are listed on the EFCSN website together with their application and the corresponding assessments so that readers have full transparency on the assessors' comments and ratings.

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 assessment process as outlined above.

2. How do third-party fact-checkers prioritize content for review and address identical and near-identical content to posts that are already fact-checked?

Editorial Independence and Freedom of the Press

Fact-checking carried out under Meta's 3PFC program is professional journalism and as such protected by freedom of the press, including the essential right to exercise editorial discretion. This means that fact-checkers have the professional autonomy to determine what topics or claims to investigate, based on journalistic standards and public interest. While Meta may set guidelines or offer technical tools to inform editorial decision-making (e.g. report queue, dashboards, or insights into content trends through tools such as previously CrowdTangle or now the MCL), the choice of what to fact-check remains at the discretion of the participating organizations.

Fact-Checking Prioritization: Public Interest, Fairness, and Impartiality

EFCSN-certified fact-checking organizations are committed to editorial independence and follow the European Code of Standards for Independent Fact-Checking Organisations, which outlines how prioritization decisions should be guided by the principles of public interest, fairness, and impartiality. Relevant provisions include:

- 2.B: *"Assess the merits of the evidence found using the same standards applied to evidence on equivalent claims, regardless of who made the claim."*
- 2.C: *"Primarily focus on topics that are in the public interest, defined here as issues that concern the welfare of society or individuals, and be willing to explain, if requested by an assessor, how various investigations fit this definition."*
- 3.1.A: *"Be editorially free and politically independent."*
- 3.1.D: *"Not focus investigations unduly on one particular political party or side of the political spectrum."*

In practice, this means that content is selected for review based on a variety of factors:

- Its potential impact on the public, particularly during moments of high social relevance (e.g. elections, health crises, or natural disasters) which may include the fueling of prevalent disinformation narratives,
- The level of engagement or virality of the claim,
- The harm potential of the misinformation (e.g. in matters of health, safety, or democratic integrity),
- The availability of verifiable evidence to assess the truthfulness of the claim.

These editorial choices are made transparently and in alignment with each organization's internal editorial policies, which themselves must comply with EFCSN standards and are reviewed regularly through a rigorous assessment process. Selection decisions may also be made based on expertise of a given fact-checking organization. Some EFCSN members are, for example, focused on science-related claims, others on health misinformation or the verification of visual material.

Use of Claim Matching Technology for Duplicates and Near-Duplicates

The 3PFC program addresses identical and near-identical content by applying claim matching technology to scale the work of fact-checkers. When a claim has already been assessed and rated

by an independent fact-checker, this technology helps identify identical or near-identical versions of the same claim across Meta's platforms.

The claim matching that occurs within the 3PFC program is Meta's own, and works similarly to matching classifiers used in other harm areas Meta's platforms: when a match is detected, the system applies a fact-check label referencing the original assessment, and may demote the content to reduce visibility and prevent further spread. Fact-checkers can make small adjustments to the classifier's actions as they work, for example, to match only whole posts, rather than just a photo that has been checked. This allows the impact of a single verified fact-check to be extended to multiple instances of the same claim, without requiring fact-checkers to repeat the same investigation. According to Meta's latest Code of Conduct on Disinformation Transparency Report, claim matching enabled the labelling of over 27,000,000 pieces of content on Facebook between 01/07/2024 and 31/12/2024. When matching classifiers are at their best, they reduce the spread of false claims effectively while minimizing the risk of over-blocking or erroneous enforcement. Compared to other forms of automated content moderation, fact-checking-based claim matching is more precise, as it relies on human-reviewed assessments before labels are applied. Meta's own Digital Services Act (DSA) Transparency Reports underscore this point: The percentage of successful organic content demotion complaints for fact-checked misinformation is by far lower than for almost all other categories, see table below. Consequently, removing expert fact-checking from the system runs the risk that these classifiers become less accurate which would be a substantial risk to freedom of speech and fundamental human rights.

Number of organic content demotion complaints and resulting demotion lifted for Facebook and Instagram
Reporting period 1/10-31/12/2024

	Facebook			Instagram		
Organic content demotions	Total demoted complaints volume	Total demotion lifted after complaint	% of successful appeals	Total demoted complaints volume	Total demotion lifted after complaint	% of successful appeals
Adult Nudity and Sexual Activity	35059	25445	72.58%	293221	267920	91.37%
Adult Sexual Solicitation & Sexually Explicit Language	21744	16779	77.17%	87551	78576	89.75%
Bullying and Harassment	159	54	33.96%	1048	815	77.77%
Child Nudity	no data			45	34	75.56%
Fact-Checked Misinformation	76161	2477	3.25%	2279	243	10.66%
Hate Speech	9912	5326	53.73%	8294	6754	81.43%
(Restricted Goods and Services) Drugs	12963	7918	61.08%	8063	6779	84.08%
Suicide and Self-Injury	2852	2121	74.37%	11156	196	1.76%
Violence and Incitement	113	34	30.09%	4023	3295	81.90%
Violent and Graphic Content	65441	60075	91.80%	9887	8470	85.67%

sources: Meta DSA Transparency Reports accessed through https://transparency.meta.com/reports/regulatory-transparency-reports, column "% of successful appeals" own calculations

3. Research into the impacts of Meta's third-party fact-checking program, as well as alternative or complementary measures

3PFC Impact

The program has shown measurable improvements to information integrity on Meta's platforms:

- 47% (Facebook) and 46% (Instagram) of attempted reshares of fact-checked content were abandoned by users thereby reducing the spread of misinformation without limiting freedom of expression (see Meta's Code of Practice Transparency Report, March 2025, SLI 31.1.2).
- 95% of users do not click to view content labeled false (source), indicating that users do not want to be exposed to false claims.
- Less than 4% of appealed fact-checks result in reinstatement on Facebook, compared to over 60% for other moderated content, indicating high accuracy and reliability (see previous section).

Independent research supports the effectiveness of warning labels, showing significant reductions in belief and sharing of false content, even among users skeptical of fact-checkers (for an overview see Martel & Rand, 2023).

Community Notes

Community Notes, initially launched by X (formerly Twitter), is a crowdsourced annotation system allowing users to add context to misleading posts, with visibility dependent on achieving broad, cross-ideological consensus. Meta has begun rolling out a similar feature across its platforms. However, evidence shows that Community Notes is too slow and ineffective at mitigating the spread of misinformation at scale:

- On average fewer than 10% of submitted notes are displayed (see Community Notes Dashboard, developed by CheckFirst), likely a result of the concept that notes only get published if users with different previous opinions rate it as helpful; this often leads to demonstrably false or misleading tweets on polarizing or controversial issues not being treated with a note: A Science Feedback study showed nearly 70% of false tweets had no visible moderation. Valencia flood disinformation and 2024 US election falsehoods largely went uncorrected by Community Notes.
- Coordinated manipulation and limited contributor pools make the system vulnerable, particularly in small language markets.

Meta's version replicates key flaws:

- No downranking of posts tagged with Community Notes, allowing falsehoods to spread unchecked.
- Exclusion of ads from Community Notes, creating a loophole for disinformation to circulate as paid content, a common occurrence in financial scams.

In summary, while Community Notes may add value in select cases, it lacks the speed, reach, accountability, and the non-partisanship of professional fact-checking. Notably, many community notes rely on fact-checks as sources. If adopted as a substitute rather than a complement, it risks undermining information integrity on Meta's platforms.