Harvard Kennedy School Misinformation Review[1]
July 2026, Volume 7, Issue 3, Special Issue on Fact-Checking

https://doi.org/10.37016/mr-2020-202
Website: https://misinforeview.hks.harvard.edu

Research Article

Fact-checking at a crossroads: Fact checkers' perspectives on Community Notes, AI integration, and design recommendations

Social media platforms are increasingly using community-based verification systems, such as Community Notes, and AI systems to flag and contextualize potentially misleading content at scale. While these approaches promise speed and broad coverage, concerns about accuracy, bias, and transparency persist. Drawing on interviews with 29 fact checkers, we find that practitioners see community-based verification and AI Note Writers as complementary tools that can support, but not replace, professional fact-checking. Our findings suggest that hybrid approaches, in which community contributors, AI tools, and expert fact checkers perform distinct but complementary roles, may offer a promising path toward scalable and trustworthy verification.

Authors: Basak Bozkurt (1), Mohsen Mosleh (1,2), Helen Margetts (1)
Affiliations: (1) Oxford Internet Institute, University of Oxford, UK, (2) Sloan School of Management, Massachusetts Institute of Technology, USA
How to cite: Bozkurt B., Mosleh, M., & Margetts, H. (2026). Fact-checking at a crossroads: Fact checkers' perspectives on Community Notes, AI integration, and design recommendations. *Harvard Kennedy School (HKS) Misinformation Review*, *7*(3).
Received: January 30th, 2026. Accepted: May 23rd, 2026. Published: July 7th, 2026.

Research questions

- RQ1: What opportunities and challenges do professional fact checkers identify in community-based verification systems?
- RQ2: How do professional fact checkers perceive AI Note Writers and their role in fact-checking practices?
- RQ3: What potential models could help address challenges in community-based verification systems?

Essay summary

- This study draws on semi-structured interviews with 29 fact checkers from member organizations of the International Fact-Checking Network (IFCN),[2] examining their perspectives on community-

[1] *A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.*
[2] The IFCN is a global network that promotes standards in fact-checking. Signatories are required to undergo an independent verification process to confirm their adherence to criteria including non-partisanship, transparency, and methodological rigor (IFCN, 2026a).

based fact-checking systems and the emerging artificial intelligence (AI) assisted features.

- Participants viewed community-based verification as promising but with significant implementation shortcomings. They emphasized its potential to expand coverage beyond the capacity of fact checkers, surface emerging claims, and provide rapid context. At the same time, they emphasized several limitations, including uneven quality of contributions, risks of bias, and slow publishing processes.
- Participants stressed that community-based verification systems are best understood as being complementary to professional fact-checking; community contributions are well-suited for low-stakes claims and for fast correction, while professional fact checkers remain essential for investigating high-stakes claims.
- Participants acknowledged AI Note Writers' potential to increase speed and scale, but expressed concern about factuality challenges, including hallucinations, bias, and limited contextual awareness.
- Building on prior scholarship proposing hybrid verification systems, this paper presents an empirical account of fact checkers' perspectives and their expressed support for participation in such systems.

Implications

Growing concern over the spread of misinformation online initially led many platforms to rely on professional fact checkers to identify and flag false content (Bettadapur, 2020; Instagram, 2019; Mosseri, 2016). Over time, platforms began to experiment with community-led fact-checking, most visibly through Community Notes (CN)[3] on X (see Figure A1 in Appendix A), formerly Birdwatch (Coleman, 2021). Other platforms have since adopted similar models. Meta has moved away from professional fact-checking partnerships toward CN in the United States (Kaplan, 2025). Whether this approach will be adopted in other regions remains unclear. TikTok introduced its own version of CN, branded as *Footnotes*, in the United States (Presser, 2025). These transitions are often framed as responses to the perceived limitations of fact-checking, including scalability challenges, alleged political bias, claims of over-enforcement, and declining public trust in fact checkers (Kaplan, 2025; Li et al., 2025).

CN enables scalable, user-facing detection and correction of misinformation beyond the limits of professional fact checkers (Augenstein et al., 2025; Slaughter et al., 2025; Solovev & Pröllochs, 2025). It relies on community contributors to add contextual notes to potentially misleading and false content, which become visible only when contributors with different rating histories agree that a note is helpful (Meta, n.d.; X, 2026). This design builds on the notion of the *wisdom of crowds*, the idea that ideologically diverse groups can deliver neutral, consensus-based judgments at scale (Allen et al., 2021).

Recent developments indicate that AI will play an increasing role in CN. In June 2025, X announced the integration of AI Note Writer into CN (see Figure 2A in Appendix A), through which large language models (LLMs) assist in generating or helping to draft notes, while users continue to rate helpfulness (Li et al., 2025). TikTok has similarly announced plans to introduce a combination of automated and human moderation through its Footnotes (Presser, 2025). Together, these developments signal growing interest in using AI to scale crowdsourced verification, with LLMs augmenting information retrieval and aggregation. At the same time, their use also raises well-documented challenges, including hallucination or lack of domain-specific expertise (Augenstein et al., 2024; Guan et al., 2024; Manakul et al., 2023; Wang et al., 2024).

[3] Throughout this paper, we use CN to refer to community-based verification systems.

As platforms transition toward CN and incorporate AI into these systems, the broader implications for the fact-checking ecosystem warrant careful examination. Fact checkers offer a particularly valuable perspective. First, their experience with large-scale misinformation leaves them well-positioned to evaluate the design and performance of CN. Second, CN are often presented as a remedy for perceived shortcomings in professional fact-checking, making it essential to examine how fact checkers themselves evaluate CN as an appropriate response. Third, the shift toward CN directly affects fact checkers' authority and future relevance. Fourth, prior work highlights the need for multilayered verification systems that combine CN with professional fact-checking (Barbera et al., 2022; Saeed et al., 2022; Vraga, 2025), yet it remains largely unclear how fact checkers themselves perceive this and whether they recognize or resist the complementary role that is ascribed to them. Finally, as AI becomes more embedded in CN, fact checkers' experience with AI tools puts them in a good position to assess the benefits and risks of AI-assisted verification.

Drawing on semi-structured, in-depth interviews with 29 professional fact checkers conducted as part of a larger project exploring the use of AI in fact-checking (Bozkurt et al., 2026), this study provides novel practitioner insights into CN and the integration of AI into community-based moderation. Participants positioned CN as a complementary component within a broader fact-checking ecosystem rather than a substitute for professional verification. They emphasized that fact checkers and community contributors can each bring distinct, but complementary, strengths: methodological rigor and subject-matter expertise on the one hand, and greater scale and reach on the other. Several also described CN as an early signal, helping fact checkers identify which claims are gaining traction and may warrant closer scrutiny. These approaches should therefore not be treated as binary choices but as mutually reinforcing components. This interpretation is also consistent with prior research showing that CN often address more straightforward forms of misinformation, such as doctored images or posts lacking context, while fact checkers tend to focus on more complex or highly contested claims (Matamoros-Fernández & Jude, 2025; Pilarski et al., 2024) and that CN frequently draw on professional fact-checks for complex or hard-to-verify claims (Borenstein et al., 2025). Together, these dynamics underscore the need for hybrid models in which crowd contributions and expert verification work together.

On the other hand, participants questioned whether CN could deliver the effectiveness and neutrality often attributed to community-based verification. They argued that CN often appear too slowly to meaningfully reduce misinformation spread, especially on polarized topics. This suggests that consensus-based publication may trade off neutrality against timeliness. This interpretation aligns with research indicating that CN did not significantly reduce engagement with misleading tweets, potentially because notes often appear too late to affect the early and most viral stages of diffusion (Chuai et al., 2024) and that consensus-based publication often results in low note visibility for the most contentious claims (Arjmandi-Lari et al., 2025; Yasseri & Menczer, 2023). Participants also underscored that the wisdom of crowds does not necessarily resolve partisan bias, especially when participation itself is politically motivated. Several described flagging practices as driven less by accuracy concerns and more by users' political allegiances. This is supported by research showing that politically motivated users are more likely to flag content (Martel et al., 2025) and that partisan asymmetries persist within CN. For example, posts by Republican users are flagged as misleading more frequently than those by Democrats (Renault et al., 2025). Together, these findings raise questions about CN's capacity to produce neutral and effective outcomes.

Further challenges relate to sustainability and representativeness. Participants questioned whether CN can function reliably when participation is concentrated among a relatively small subset of contributors and relies on unpaid volunteer labor. Some also noted uneven visibility across geographic and linguistic contexts, suggesting that not all users encounter CN in the same way. This indicates that CN's effectiveness depends on who participates and where content is produced and viewed. This aligns with research showing that most notes are produced by a small minority of contributors and that many

drop out over time, raising questions about both representativeness and reliance on unpaid volunteer labor (Arjmandi-Lari et al., 2025; Nudo et al., 2026). It also suggests significant variation in participation across regions and languages (Bassett, 2025; Stewart et al., 2025). Finally, participants emphasized that these challenges are further exacerbated by the absence of transparent and consistent methodological standards comparable to those guiding professional fact-checking.

With regard to AI Note Writers, participants identified both opportunities and risks. AI has long been used to support fact checkers in tasks, such as claim detection, check-worthiness estimation (prioritizing claims based on public interest, verifiability, urgency, and potential real-world harm), and the retrieval of previously fact-checked claims (Graves, 2018; Hasanain et al., 2024; Majer & Šnajder, 2024; Micallef et al., 2022; Nakov et al., 2021). Participants saw clear potential for AI to improve speed and scale in CN, particularly for tasks such as drafting, summarizing, and identifying existing fact-checks. They also highlighted AI's potential for synthesizing divergent contributions into a balanced overview, ensuring that users still receive contextual information rather than no note at all. Evidence from a small-scale experiment suggests that such a system was rated as more helpful than the highest-rated existing CN (De et al., 2025). On the other hand, participants also raised concerns about factuality, bias, misinterpretation, and errors. As a result, many emphasized that AI should play an assistive role and not publish notes without human oversight.

Taken together, these perspectives point toward an important implication: the need for a multi-tiered verification model integrating community input, AI assistance, and professional fact-checking. For platforms, an important direction is the development of models that combine community input with expert oversight, while also deploying AI. While platforms such as X have recently combined AI systems with CN (Li et al., 2025), our findings extend prior proposals for more systematic expert involvement (Augenstein et al., 2025; Barbera et al., 2022; Pilarski et al., 2024; Saeed et al., 2022; Vraga, 2025) by grounding these proposals in empirical evidence from professional fact checkers. Figure 1 synthesizes participants' perspectives and our suggestions into a proposed multi-tiered verification model. In this model, detection, contextualization, and in-depth investigation are distributed across community contributors, fact checkers, and AI systems. AI performs the initial classification, determines the review pathway, and supports drafting; community contributors provide fast contextual notes; and experts investigate complex or high-stakes claims.

Realizing such models also requires investment in participation infrastructures. CN relies on unpaid volunteer labor and offers limited incentives or community-building mechanisms compared to communities such as Wikipedia (Baytiyeh & Pfaffman, 2010) or Reddit (Moore & Chuang, 2017), which use reputation systems to facilitate long-term engagement. Platforms should invest in participation infrastructures, including community-building incentives. Fact checkers similarly require support, particularly financial support from platforms, to maintain their role in the ecosystem.

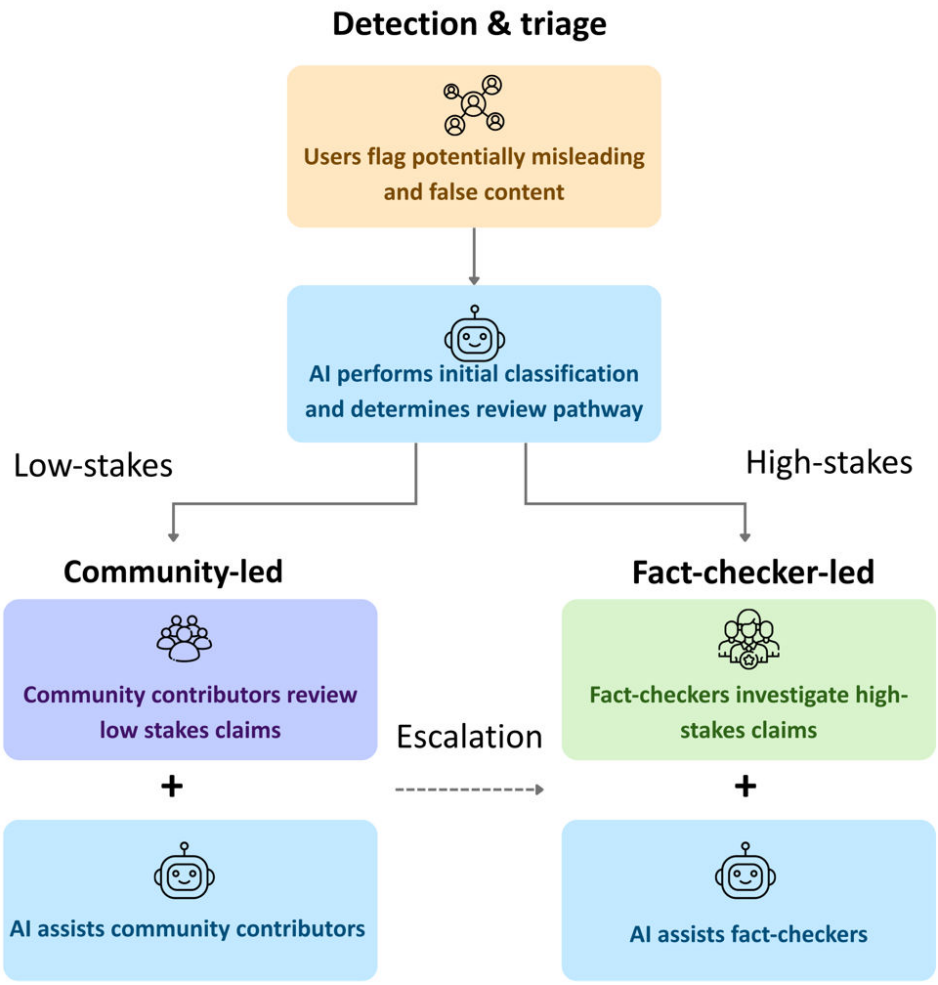

Figure 1. Proposed multi-tiered verification model. *The figure illustrates how AI assistance, community contributors, and professional fact checkers could interact across the verification pipeline. In the Detection & Triage stage, users flag posts, and AI performs initial classification and determines the review pathway. In the Community-led stage, contributors provide contextual notes, supported by AI-assisted drafting and editing. In the Fact checker-led stage, fact checkers investigate complex or high-stakes claims, assisted by AI tools. Claims may also be escalated from the Community-led stage to the Fact checker-led stage when contributors flag them as requiring expert investigation.*

For misinformation researchers, our findings highlight the need to design and evaluate hybrid verification systems, including task allocation across actors and their performance in accuracy, speed, and credibility. Existing check-worthiness systems have largely been developed around professional fact checkers' priorities (Hasanain et al., 2024), whereas community-based verification is shaped by user flagging. Two directions follow: extending check-worthiness systems to capture community perspectives and building on these systems to develop stakes-based routing that allocates posts across community, AI, and expert tiers.

These findings also have implications for educators and civil society organizations. As CN and AI Note Writers increasingly shape what citizens encounter when they face misinformation on social media, media literacy efforts should focus on helping audiences to interpret these systems, including their strengths, limitations, and appropriate use.

Evidence

Finding 1: CN is widely viewed as "a good idea that has been badly implemented."

In response to RQ1, participants identified both the opportunities and the shortcomings they saw in CN. Many viewed CN as promising in theory but falling short in practice. As one participant (P26)[4] said, "it's a good idea that has been badly implemented." They emphasized that CN could broaden participation in countering misinformation by extending coverage beyond the capacity of fact checkers, particularly for content circulating outside mainstream news, such as local issues or emerging claims. Some reported monitoring CN to track which topics are gaining traction, explaining that notes sometimes offer niche context, background, or early leads that help inform their investigations. They also noted that the use of accessible language and short explanations can make complex information easier for audiences to understand.

Participants identified several features that limit the effectiveness of CN. They emphasized a wide variation in the quality and visibility of contributions: while some notes align with professional standards by providing evidence and citing reliable sources, including fact checkers, the most frequently cited sources—X and Wikipedia—are not primary (Kangur et al., 2024). Participants also highlighted the opacity of the system, noting that it is unclear who writes notes, who rates them, or what standards apply. Unlike IFCN signatories, which operate under standards of non-partisanship, fairness, transparency of sources and methodology, and open correction policy (IFCN, 2026b), CN rely on anonymous or pseudonymous contributors whose expertise, methods, and motivations are not visible, raising concerns about bias, accountability, and transparency. Participants also observed that CN function unevenly across countries; for instance, one participant in Southeast Asia (P23) reported rarely encountering CN on X, reflecting lower participation and note production in some regions and languages (Bassett, 2025; Stewart et al., 2025).

Concerns were also raised about participation dynamics. Participants reported that low thresholds for participation, combined with a "loud minority" of users (P15), may allow unqualified or biased users to shape what is considered helpful. Participants highlighted that, while consensus requirements aim to reduce bias, they often delay publication or prevent notes from appearing at all, especially on polarizing topics. They also emphasized issues of contributor retention, due to a reliance on unpaid, unincentivized volunteer labor, which they believed leads to declining participation over time.

Taken together, these findings suggest that while CN has the potential to expand coverage and accessibility, current design choices around participation structures, its uneven operation, and the lack of clear standards and transparency limit its effectiveness and credibility in practice.

Finding 2: CN complements, not replaces, professional fact-checking.

Further addressing RQ1, drawing on both the strengths and limitations of CN, participants stressed that CN should be understood as a complement to, rather than a substitute for, fact checkers. As P19 explained, given that fact checkers already struggle with "the volume and pace of false information," "we [fact checkers] shouldn't discount the idea of community-led verification efforts." At the same time, participants emphasized that community systems cannot replace expert verification, particularly for complex or contentious claims. As P25 said:

[4] Participants are referred to throughout the paper using codes (P01–P29) to protect confidentiality.

> There is a fundamental mistake of thinking that it's a one-in, one-out system, and that community notes replace fact checkers. I think that is just terrible thinking. I would love to see a world where you have both. I think community notes are really good at a certain level of things, where they are not particularly contentious, or it's relatively small things. Community notes, from all the data we've seen, are pretty bad when it's a contentious issue, a hard issue. They're not fast enough for a start. They don't have the expertise needed to answer some of these questions.

Overall, participants emphasized the interdependence of these approaches, viewing CN as a supplementary layer rather than as a substitute for professional fact-checking.

Finding 3: AI Note Writers offer speed and scale but raise concerns about factuality.

Addressing RQ2, participants reflected on AI Note Writers and the role they might play in fact-checking practices. Many participants emphasized that they already use AI tools in their own work and do not see AI as "inherently a bad thing," but rather as "part of the solution to deal with misinformation at scale." (P26). In their own workflows, participants reported using AI across stages of fact-checking, including transcription, summarization, evidence gathering, and detection of previously fact-checked claims (Bozkurt et al., 2026). Drawing on this experience, participants saw potential in AI to increase speed and scale. They noted that AI could assist with drafting notes, synthesizing input from multiple contributors, and producing concise explanations more quickly than contributors, as well as identifying previously fact-checked claims.

Despite this openness, participants expressed strong reservations about the use of AI to generate notes. One concern expressed by participants was that AI systems developed in one part of the world are used across very different contexts on global platforms. For example, P18 explained:

> So, I think the problem with social media platforms' AI is that they only get programmed sort of in one part of the world. But the whole world uses the platform. I think there are a lot of questions that are misinterpreted or taken out of context because the AI doesn't understand what's being asked, and then it scrambles for information and scrambles to find an answer. And I think that's where it can become a real problem.

This points to a fundamental limitation: the difficulty in capturing the local political, cultural, social, and linguistic contexts in which claims circulate.

More broadly, participants worried that these systems could show some well-known factuality challenges that fact checkers have already encountered in their own work with AI tools, such as fabricated sources and citations, biases, and fluent but misleading outputs (Bozkurt et al., 2026). As P19 emphasized:

> There might be potential for AI to help with, for example, summarizing info or identifying patterns to help fact checkers (and) human reviewers to process the huge volume of information on social media. But this approach poses the same risks we've already identified with generative AI in general, i.e., perpetuating bias; hallucinating; sounding persuasive and confident even when inaccurate.

Participants therefore stressed that any use of AI to write or synthesize notes must remain transparent and operate under clear human supervision. Reflecting this view, P24 explained:

> AI alone should not decide what is true or false, as it is not fully capable of fact-checking all content. Human review is still necessary to ensure accuracy, fairness, and trust. I think AI can assist, but human judgment should guide the final decision.

This view was widely shared: AI is seen as a supportive tool for fact-checking, but its value depends on the human oversight that surrounds it.

Finding 4: Fact checkers proposed hybrid models that integrate community participation with professional oversight to enhance both the scale and credibility of verification efforts.

Turning to RQ3, participants put forward proposals for models that could help address the challenges identified in community-based verification. Many participants articulated a vision for hybrid models that integrates fact checkers and CN. As P05 stated, combating misinformation "is not one versus the other. It should be one with the other." This perspective illustrates participants' framing of fact checkers and CN as mutually reinforcing, rather than mutually exclusive. This naturally raises the question of what roles these different actors might take on. In these models, CN would contribute speed, reach, and contextual detail, while fact checkers supply expertise and methodological rigor. Participants described CN as an audience-facing fast context layer, offering concise clarifications and broad coverage for relatively low-stakes claims. These are claims that are straightforward to verify without expertise (e.g., outdated images presented as current or misattributed quotes) or that have already been fact-checked. Professional fact checkers would focus on in-depth investigations of complex and high-stakes claims: those which are highly contested, require expertise, or have the potential for harm.

Participants suggested that collaboration could strengthen the information ecosystem, with community contributors improving accessibility and scale, and fact checkers ensuring accuracy and accountability. Some participants pointed to Wikipedia as an illustrative example, where contributions are open to the public but moderated through a structured hierarchy of roles (Ren et al., 2023). These findings point toward hybrid verification models that combine the speed and reach of community contributions with the expertise of professional fact checkers.

Methods

The study draws on 28 semi-structured interviews with 29 professional fact checkers conducted between May and September 2025; one interview included two participants from the same organization. Participants came from IFCN signatory fact-checking organizations operating in 41 countries across South America, Africa (Western, Middle, Eastern, and Southern), Asia (Western, Central, Southern, Southeastern, and Eastern), Europe (Eastern, Northern, and Southern), and North America. Of the participants, 17 self-identified as women and 12 as men. Participants hold research, editorial, and technical roles and have a range of experience levels, with the majority at early- to mid-career stages.

We conducted all interviews remotely via Microsoft Teams using a semi-structured interview protocol as part of a project on the use of AI (Bozkurt et al., 2026). For this paper, we focused on participants' views on community-based verification systems and on the integration of generative AI into these systems. We analyzed the data using reflexive thematic analysis (Braun & Clarke, 2022). We first familiarized ourselves with the data by reviewing interview transcripts and cross-checking them against audio recordings. We then coded the dataset inductively using NVivo 14. Next, we examined the coded data for patterns and grouped related codes into initial themes. We iteratively developed and refined themes until we had a coherent interpretive account that adequately captured patterned meaning across the dataset. Finally,

we defined the final themes and wrote up the analysis (see Table B1 in Appendix B for an overview of themes and the research questions they address).

Further information about the participant demographics, recruitment, data collection, and limitations can be found in Appendix C.

Bibliography

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances*, *7*(36). https://doi.org/10.1126/sciadv.abf4393

Arjmandi-Lari, Z., Mantzarlis, A., & Stafford, T. (2025). *Threats to the sustainability of Community Notes on X*. arXiv. https://doi.org/10.48550/arXiv.2510.00650

Augenstein, I., Bakker, M., Chakraborty, T., Corney, D., Ferrara, E., Gurevych, I., Hale, S., Hovy, E., Ji, H., Larraz, I., Menczer, F., Nakov, P., Papotti, P., Sahnan, D., Warren, G., & Zagni, G. (2025). *Community moderation and the new epistemology of fact checking on social media*. arXiv. https://doi.org/10.48550/arXiv.2505.20067

Augenstein, I., Baldwin, T., Cha, M., Chakraborty, T., Ciampaglia, G. L., Corney, D., DiResta, R., Ferrara, E., Hale, S., Halevy, A., Hovy, E., Ji, H., Menczer, F., Miguez, R., Nakov, P., Scheufele, D., Sharma, S., & Zagni, G. (2024). Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence*, *6*(8), 852–863. https://doi.org/10.1038/s42256-024-00881-z

Barbera, D. L., Roitero, K., & Mizzaro, S. (2022). A hybrid human-in-the-loop framework for fact checking. In D. Nozza, L. C. Passaro, & M. Polignano (Eds.), *NL4AI 2022: Sixth Workshop on Natural Language for Artificial Intelligence* (pp. 1–11). CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3287/paper4.pdf

Bassett, K. (2025). *X's Community Notes and the South Asian misinformation crisis*. Center for the Study of Organized Hate. https://www.csohate.org/2025/06/30/x-community-notes-south-asia/

Baytiyeh, H., & Pfaffman, J. (2010). Volunteers in Wikipedia: Why the community matters. *Journal of Educational Technology & Society*, *13*(2), 128–140. https://www.jstor.org/stable/jeductechsoci.13.2.128

Bettadapur, A. N. (2020). *TikTok partners with fact-checking experts to combat misinformation*. TikTok Newsroom. https://newsroom.tiktok.com/tiktok-partners-with-fact-checking-experts-to-combat-misinformation?lang=en-AU

Borenstein, N., Warren, G., Elliott, D., & Augenstein, I. (2025). Can Community Notes replace professional fact-checkers? In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Vol. 2: Short Papers, pp. 535–552). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-short.42

Bozkurt, B., Mosleh, M., & Margetts, H. (2026). Fact-checkers navigating generative AI: Practices, boundaries, and design implications. In *FAccT '26: Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency* (pp. 5546–5569). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3805689.3806474

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Chuai, Y., Tian, H., Pröllochs, N., & Lenzini, G. (2024). Did the roll-out of Community Notes reduce engagement with misinformation on X/Twitter? *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW2), 1–52. https://doi.org/10.1145/3686967

Coleman, K. (2021). Introducing Birdwatch, a community-based approach to misinformation. *Twitter Blog*. https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation

De, S., Bakker, M. A., Baxter, J., & Saveski, M. (2025). Supernotes: Driving consensus in crowd-sourced fact-checking. In *WWW '25: Proceedings of the ACM Web Conference 2025* (pp. 3751–3761). Association for Computing Machinery. https://doi.org/10.1145/3696410.3714934

EFCSN. (2026). *European Fact-Checking Standards Network*. https://efcsn.com/

Full Fact. (2020). *The challenges of online fact checking*. https://fullfact.org/media/uploads/coof-2020.pdf

Graves, L. (2018). *Understanding the promise and limits of automated fact-checking*. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-NQNX-BG89

Graves, L., & Mantzarlis, A. (2020). Amid political spin and online misinformation, fact checking adapts. *The Political Quarterly*, *91*(3), 585–591. https://doi.org/10.1111/1467-923X.12896

Guan, J., Dodge, J., Wadden, D., Huang, M., & Peng, H. (2024). Language models hallucinate, but may excel at fact verification. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1: Long Papers, pp. 1090–1111). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.62

Hacks/Hackers. (2025). *Hacks/Hackers*. https://www.hackshackers.com/

Hasanain, M., Suwaileh, R., Weering, S., Li, C., Caselli, T., Zaghouani, W., Barrón-Cedeño, A., Nakov, P., & Alam, F. (2024). Overview of the CLEF-2024 CheckThat! Lab Task 1 on check-worthiness estimation of multigenre content. In G. Faggioli, N. Ferro, P. Galuščáková, & A. García Seco de Herrera (Eds.), *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024)* (pp. 276–286). CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3740/paper-24.pdf

IFCN. (2025a). *Global Fact Check Fund awards $2 million to 20 fact-checking groups across 15 countries*. Poynter. https://www.poynter.org/ifcn/2025/global-fact-check-fund-awards-2-million-to-20-fact-checking-groups-across-15-countries/

IFCN. (2025b). *GlobalFact 12 in Rio de Janeiro, Brazil, opens registration for June conference*. Poynter. https://www.poynter.org/news-release/2025/globalfact-12-rio-de-janeiro-2025-fact-checking-summit/

IFCN. (2026a). *About*. https://www.ifcncodeofprinciples.poynter.org/about

IFCN. (2026b). *The commitments of the code of principles*. https://ifcncodeofprinciples.poynter.org/the-commitments

Instagram. (2019). *Combatting misinformation on Instagram*. https://about.instagram.com/blog/announcements/combatting-misinformation-on-instagram

JournalismAI. (2024). *2024 innovation challenge*. https://www.journalismai.info/programmes/innovation/innovation-challenge-2024

Kangur, U., Chakraborty, R., & Sharma, R. (2024). *Who checks the checkers? Exploring source credibility in Twitter's Community Notes*. arXiv. http://arxiv.org/abs/2406.12444

Kaplan, J. (2025). *More speech and fewer mistakes*. Meta. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

Li, H., De, S., Revel, M., Haupt, A., Miller, B., Coleman, K., Baxter, J., Saveski, M., & Bakker, M. (2025). Scaling human judgment in Community Notes with LLMs. *Journal of Online Trust and Safety*, *3*(1). https://doi.org/10.54501/jots.v3i1.255

Majer, L., & Šnajder, J. (2024). Claim check-worthiness detection: How well do LLMs grasp annotation guidelines? In M. Schlichtkrull, Y. Chen, C. Whitehouse, Z. Deng, M. Akhtar, R. Aly, Z. Guo, C. Christodoulopoulos, O. Cocarascu, A. Mittal, J. Thorne, & A. Vlachos (Eds.), *Proceedings of the Seventh Fact Extraction and VERification Workshop (FEVER)* (pp. 245–263). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.fever-1.27

Manakul, P., Liusie, A., & Gales, M. (2023). SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 9004–9017). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.557

Martel, C., Allen, J. N. L., Pennycook, G., & Rand, D. G. (2025). *Political motives help rather than hinder crowdsourced fact-checking*. PsyArXiv. https://doi.org/10.31234/osf.io/8fhxz_v2

Matamoros-Fernández, A., & Jude, N. (2025). The importance of centering harm in data infrastructures for 'soft moderation': X's Community Notes as a case study. *New Media & Society*, *27*(4), 1986–2011. https://doi.org/10.1177/14614448251314399

Meta. (n.d.). *Community Notes: Keeping people better informed*. https://www.meta.com/technologies/community-notes/

Micallef, N., Armacost, V., Memon, N., & Patil, S. (2022). True or false: Studying the work practices of professional fact-checkers. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1–44. https://doi.org/10.1145/3512974

Moore, C., & Chuang, L. (2017). Redditors revealed: Motivational factors of the Reddit community. In T. X. Bui & R. Sprague, Jr. (Eds.), *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 2313–2322). Hawaii International Conference on System Sciences (HICSS). https://doi.org/10.24251/HICSS.2017.279

Mosseri, A. (2016). *Addressing hoaxes and fake news*. Meta. https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/

Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., & Da San Martino, G. (2021). Automated fact-checking for assisting human fact-checkers. In Z.-H. Zhou (Ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence* (pp. 4551–4558). International Joint Conferences on Artificial Intelligence. https://doi.org/10.24963/ijcai.2021/619

Nudo, J., Nemmi, E. N., Loru, E., Mei, A., Quattrociocchi, W., & Cinelli, M. (2026). *Hyperactive minority alter the stability of Community Notes*. arXiv. https://doi.org/10.48550/arXiv.2602.08970

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE.

Pilarski, M., Solovev, K. O., & Pröllochs, N. (2024). Community Notes vs. snoping: How the crowd selects fact-checking targets on social media. *Proceedings of the International AAAI Conference on Web and Social Media*, *18*, 1262–1275. https://doi.org/10.1609/icwsm.v18i1.31387

Poynter. (2026). *International Fact-Checking Network*. https://www.poynter.org/ifcn/

Presser, A. (2025). *Rolling out TikTok Footnotes in the US*. TikTok Newsroom. https://newsroom.tiktok.com/rolling-out-tiktok-footnotes-in-the-us?lang=en

Ren, Y., Zhang, H., & Kraut, R. E. (2023). How did they build the free encyclopedia? A literature review of collaboration and coordination among Wikipedia editors. *ACM Transactions on Computer-Human Interaction*, *31*(1), 1–48. https://doi.org/10.1145/3617369

Renault, T., Mosleh, M., & Rand, D. G. (2025). Republicans are flagged more often than Democrats for sharing misinformation on X's Community Notes. *Proceedings of the National Academy of Sciences*, *122*(25), Article e2502053122. https://doi.org/10.1073/pnas.2502053122

Saeed, M., Traub, N., Nicolas, M., Demartini, G., & Papotti, P. (2022). Crowdsourced fact-checking at Twitter: How does the crowd compare with experts? In *CIKM '22: Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (pp. 1736–1746). Association for Computing Machinery. https://doi.org/10.1145/3511808.3557279

Slaughter, I., Peytavin, A., Ugander, J., & Saveski, M. (2025). Community Notes reduce engagement with and diffusion of false information online. *Proceedings of the National Academy of Sciences*, *122*(38), Article e2503413122. https://doi.org/10.1073/pnas.2503413122

Solovev, K., & Pröllochs, N. (2025). References to unbiased sources increase the helpfulness of community fact-checks. *Scientific Reports*, *15*(1), Article 25749. https://doi.org/10.1038/s41598-025-09372-6

Stewart, E., Greenwold, S., & Marselo, T. (2025). *Community Notes: Crowd participation and dependence on professional fact-checking across languages*. arXiv. https://doi.org/10.48550/arXiv.2512.19947

Vraga, E. K. (2025). Understanding the strengths and limitations of community-based responses to misinformation. *Proceedings of the National Academy of Sciences*, *122*(48), Article e2524004122. https://doi.org/10.1073/pnas.2524004122

Wang, Y., Wang, M., Manzoor, M. A., Liu, F., Georgiev, G. N., Das, R. J., & Nakov, P. (2024). Factuality of large language models: A survey. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 19519–19529). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1088

X. (2026). *Diversity of perspectives*. https://communitynotes.x.com/guide/en/contributing/diversity-of-perspectives

Yasseri, T., & Menczer, F. (2023). Can crowdsourcing rescue the social marketplace of ideas? *Communications of the ACM*, *66*(9), 42–45. https://doi.org/10.1145/3578645

Acknowledgments
We thank the participants for their time and contributions to this study. We also thank the reviewers for their constructive feedback, Martha Stolze for helpful comments on an earlier draft, and the organizers of the workshop *The Future of Fact-checking in the Algorithmic Society*, held on 27–28 November 2025 at Exeter College, University of Oxford, for the opportunity to present and discuss our research.

Funding
Participant compensation for interviews was funded through the Oxford Internet Institute DPhil Student Research Fund at the University of Oxford.

Competing interests
The authors declare no competing interests.

Ethics
The study received institutional ethics approval from the Oxford Internet Institute Departmental Research Ethics Committee, University of Oxford (reference number: 706386). Informed oral consent was obtained from all participants. Participants self-reported their gender.

Data availability
Participants granted oral consent on the condition that interview transcripts would remain accessible only to the research team, as outlined in our ethics application. Therefore, we cannot share the data outside our research group.

Appendix A: Examples of Community Notes and AI Note Writer

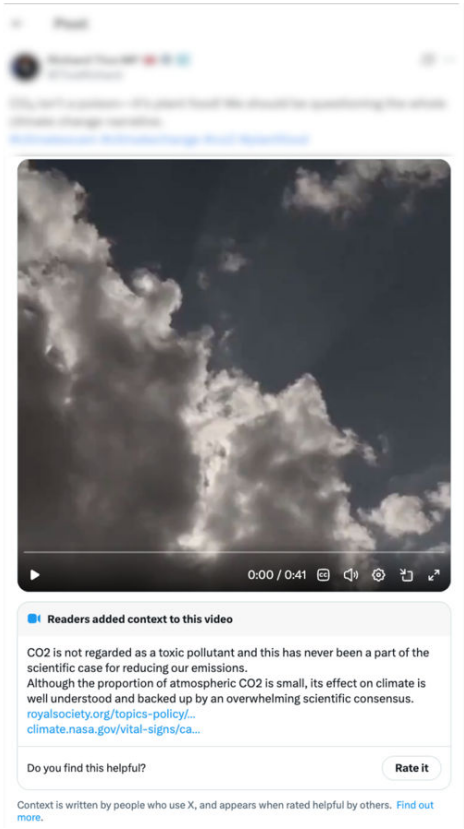

Figure A1. An example of a Community Note (tweet blurred out to protect privacy).

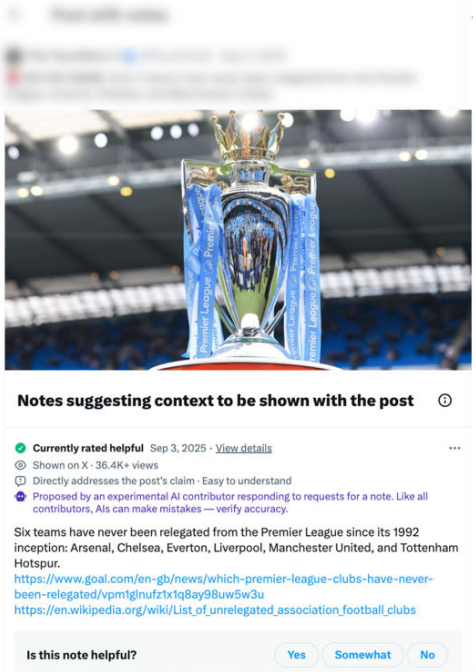

Figure A2. An example of a note proposed by an AI Note Writer (tweet blurred out to protect privacy).

Appendix B: Themes

Table B1. *Summary of themes.*

Theme	Description	Research Questions
CN as a promising but poorly implemented system	Participants saw Community Notes as a promising way to scale contextual information and broaden participation in verification, but noted that inconsistent quality, opaque governance, and uneven participation limit its effectiveness.	RQ1
CN as a complement to professional fact-checking	Participants framed CN as a complement to professional fact-checking, providing speed and scale for less complex claims, while experts address complex or high-stakes misinformation.	RQ1
AI Note Writers: Speed and scale with factuality risks	Participants saw AI Note Writers as enabling faster and more scalable notes but raised concerns about factual errors and the need for human review.	RQ2
Hybrid verification	Participants proposed hybrid verification systems that integrate community contributions with professional oversight.	RQ3

Appendix C: Participant demographics, data collection, and limitations

Participant demographics

***Table C1**. Participant demographics.*

ID	Gender	Age group	Highest education	Fact-checking experience	Role category	Region
P01	Woman	25–34	Master's degree	4–6 years	Editorial	Western Asia
P02	Man	>55	Master's degree	>10 years	Editorial	Western Europe
P03	Woman	25–34	Bachelor's degree	<1 year	Editorial	Eastern Europe
P04	Woman	25–34	Bachelor's degree	4–6 years	Editorial	Western Asia
P05	Man	35–44	Bachelor's degree	7–10 years	Editorial	Southern Asia
P06	Man	25–34	Bachelor's degree	7–10 years	Technical	Southern Europe
P07	Woman	45–54	Master's degree	4–6 years	Editorial	Northern Europe
P08	Woman	45–54	Bachelor's degree	<1 year	Technical	South America
P09	Man	18–24	Bachelor's degree	1–3 years	Technical	Western Africa; Middle Africa
P10	Woman	25–34	Bachelor's degree	1–3 years	Editorial	Western Asia
P11	Woman	25–34	Bachelor's degree	7–10 years	Editorial	Southeastern Asia
P12	Man	25–34	Master's degree	7–10 years	Editorial	South America
P13	Woman	35–44	Master's degree	7–10 years	Editorial	Eastern Europe
P14	Woman	35–44	Postgraduate diploma	7–10 years	Editorial	Southern Asia; Southeastern Asia
P15	Man	25–34	Master's degree	7–10 years	Editorial	Western Africa
P16	Woman	35–44	Master's degree	1–3 years	Editorial	South America
P17	Woman	45–54	Bachelor's degree	4–6 years	Editorial	North America
P18	Woman	25–34	Master's degree	4–6 years	Editorial	Western Africa; Eastern Africa; Southern Africa
P19	Woman	25–34	Master's degree	4–6 years	Editorial	Western Africa; Eastern Africa; Southern Africa
P20	Woman	35–44	Bachelor's degree	4–6 years	Editorial	Eastern Asia
P21	Woman	25–34	Master's degree	4–6 years	Editorial	Eastern Asia
P22	Man	25–34	Master's degree	1–3 years	Editorial	Central Asia; Eastern Asia
P23	Woman	35–44	Master's degree	1–3 years	Editorial	Southeastern Asia
P24	Man	25–34	Master's degree	4–6 years	Editorial	Southern Asia
P25	Man	45–54	Bachelor's degree	4–6 years	Technical	Northern Europe
P26	Man	35–44	Master's degree	4–6 years	Technical	South America
P27	Woman	35–44	Master's degree	1–3 years	Editorial	Southern Asia
P28	Man	35–44	Doctorate degree	1–3 years	Editorial	Southern Europe
P29	Man	>55	Master's degree	>10 years	Editorial	North America

Notes: Job titles were grouped into functional categories to reduce identification risk. We assigned role categories based on participants' self-reported job titles. Editorial roles include positions responsible for reporting, verification, writing, editing fact-checking content. Technical roles include positions focused on tool development, AI integration, and product design. At the same time, we recognize that, in practice, responsibilities are not always strictly divided, especially in smaller organizations where editorial staff also perform technical tasks. We did not observe different patterns between editorial and technical participants' perspectives. In many organizations, editorial staff also took on responsibilities related to technology. As a result, the distinction between editorial and technical roles was not always clear-cut in practice. Regions reflect where participants' organizations operate and are reported instead of countries to preserve anonymity.

Participant recruitment

We recruited participants by using purposive sampling, snowball sampling, and outreach through professional mailing lists (Patton, 2015). As part of a broader project on AI in fact-checking, we purposively identified organizations by reviewing grant funding records (e.g., IFCN, 2025a; JournalismAI, 2024) and conference programs such as Global Fact (IFCN, 2025b), as well as previous research and reports on AI in fact-checking (Full Fact, 2020). We contacted fact checkers by email and via LinkedIn. At the end of each interview, we asked participants to suggest additional potential participants, expanding the sample through snowball sampling. To broaden recruitment, we also used professional networks (EFCSN, 2026; Hacks/Hackers, 2025; Poynter, 2026) and distributed interview invites through their mailing lists.

Data collection

To be eligible for participation, individuals had to be at least 18 years old, employed by an IFCN signatory organization, and report using generative AI tools in their professional work. IFCN signatory status was used as a criterion because it indicates adherence to established norms of transparency, methodological rigor, and non-partisanship, and was more likely to represent well-established active organizations that conduct fact-checking as a core activity rather than an occasional project (Graves & Mantzarlis, 2020). Eligible respondents who completed the registration form were contacted to arrange an interview. Interviews followed a semi-structured format guided by a set of questions on fact-checking processes, the use of generative AI, AI-generated misinformation, and Community Notes. While the broader interview protocol covered these topics, the analysis presented in this paper focuses on participants' perspectives on two key questions: "What's your view on community-led models like Community Notes?" and "It looks like LLMs will soon be integrated into Community Notes on X. What do you think about using generative AI in this context?"

All interviews were conducted remotely using Microsoft Teams and generally lasted approximately one hour. With participants' oral consent, interviews were recorded and automatically transcribed using Microsoft Teams' transcription feature. Participants were offered a £35 digital gift card in recognition of their participation. Three participants donated the amount to the IFCN, and three donated it to a fact-checking organization of their choice.

Limitations

This study has several limitations that should be considered. First, we only conducted interviews in English, which may have discouraged participation from some fact-checking organizations and influenced the ways in which participants expressed their views. Second, as part of a broader project on generative AI in fact-checking, most participants had prior experience of using AI in their work. While this enabled informed reflections on both the opportunities and challenges associated with AI Note Writers, the findings may not fully reflect the perspectives of fact checkers without direct AI experience and may overemphasize concerns and advantages specific to more AI-experienced practitioners. Finally, focusing on IFCN signatory organizations may not fully capture practices across the broader fact-checking community.