Title: Findings appendix for "Professional and community-based fact-checking show different strengths, but neither performs strongly across trust, scalability, and impact"
Authors: Petter Bae Brandtzaeg (1), Silje Susanne Alvestad (2,3), Jonas R. Kunst (4,5), Asbjørn Følstad (6)
Date: July 30th, 2026
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix A: Findings

Tables A1, A2, and A3 synthesize the findings from 21 peer-reviewed studies examining fact-checking interventions on social media, focusing on three dimensions: trust, scalability, and impact. The tables also report platform context, method, region, and, where available, sample size or corpus size. Our analysis shows how methodological choices shape the assessment of these dimensions. Studies of professional fact-checking primarily examine individual-level trust and attitudinal outcomes using experiments, surveys, and interviews, whereas studies of community-based fact-checking more often evaluate scalability and systemic impact through platform-level, data-intensive analyses. The studies are geographically diverse, covering North America, Latin America, Europe, and Asia, but the evidence base remains concentrated in U.S.-oriented X/Twitter research. Platforms examined include X/Twitter, Facebook, WhatsApp, online newspapers, community radio, and professional fact-checking sites, with X/Twitter hosting the most prominent community-based initiative, Community Notes (formerly Birdwatch). Table A1 below sums up the results from the reviewed studies on professional fact-checking. Table A2 shows reviewed studies on community-based fact-checking, while Table A3 shows the reviewed comparative studies.

Table A1. *Studies on professional fact-checking (n = 7).*

Authors	Trust	Scalability	Impact	Platform and method
Aruguete et al. (2024)	N/A	N/A	Moderate; users more likely to like and engage with confirmatory fact-checks aligned with own beliefs.	Facebook. Argentina, Brazil, Chile, and Colombia. Questionnaire-based experiment (N = 9,512). Participants exposed to Facebook-style posts about COVID-19 vaccines and other topics, framed as confirmational (e.g., "True:") or refutational (e.g., "False:") fact-checks.
Aruguete et al. (2025)	N/A	N/A	Moderate; users more likely to share confirmatory factchecks that align with own beliefs.	Twitter. Chequeado (leading Argentine fact-checking organization). Argentina. Experimental manipulation of tweet labels; analysis of user engagement behavior (N = 2,041), combined with observation data on Twitter.

Authors	Trust	Scalability	Impact	Platform and method
Ejaz et al. (2025)	Moderate; the effectiveness and the engagement with both the fact-checks and mis-information where moderate by trust.	Moderate; liking or sharing correction on social media higher among those concerned with mis-information.	Moderate; fact-checking most effective among participants who trust the news the most, and least effective among those trusting social media.	Facebook, Twitter, online newspapers. Pakistan. Questionnaire-based experiment (N participants = 302, N observations = 1,208) investigating the effectiveness of fact-checking, and the determinants of engagement with fact-checks and misinformation.
Freiling et al. (2023)	N/A	N/A	Moderate; users' motivational and affective states limit the impact of fact-checking.	Facebook. U.S. Experimental design using real-world misinformation/fact-check content in a politically charged public health context (COVID-19) (N = 719).
Lavilles et al. (2023)	Low: Only 25–27% expressed explicit trust in Rappler and VeraFiles (Those two are the focus in the study, however, two other fact-checking initiatives with higher trust score are mentioned).	Low; only 22% had used Rappler or VeraFiles for verification. Low capacity to respond to all mis-information.	Low; fact-checks with limited behavioral impact. Some users reported fact-checking to improve own verification behavior, others prefer doing own checks.	Facebook, Rappler, VeraFiles. Philippines. Mixed methods: FGDs, focus group interviews (n = 121), survey (n = 685). Filipino social media users in 2020.
Park (2023)	Moderate; perceived severity of fake news positively associated with trust in fact-checking.	N/A	Moderate; perceived severity of fake news and, to some extent dissatisfaction with news media, positively associated with use of fact checkers.	Social media platforms. U.S. Questionnaire study (N = 2,350).

Authors	Trust	Scalability	Impact	Platform and method
Seelam et al. (2024)	Low-moderate; credible by design, but awareness and engagement are limited. Quote: "Fact-checks are for the top 0.1%... educated population[s] at least have the know-how and agency…"	Low; major bottlenecks including financial constraints, limited manpower, and technological gaps in hyperlocal outreach.	Low; fact-checks with some impact via outreach efforts (e.g., stringers, local language materials), but limited by low media literacy, lack of awareness, and infrastructural gaps.	WhatsApp, Facebook, community radio, traditional media. India. Interviews with 12 representatives from fact-checking organizations (e.g., editors, fact checkers, founders).

References (Table A1)

Aruguete, N., Batista, F., Calvo, E., Guizzo-Altube, M., Scartaschini, C., & Ventura, T. (2024). Framing fact-checks as a "confirmation" increases engagement with corrections of misinformation: A four-country study. *Scientific Reports, 14,* Article 3201. https://doi.org/10.1038/s41598-024-53337-0

Aruguete, N., Bachmann, I., Calvo, E., Valenzuela, S., & Ventura, T. (2025). Truth be told: How "true" and "false" labels influence user engagement with fact-checks. *New Media & Society, 27*(3), 1443–1464. https://doi.org/10.1177/14614448231193709

Ejaz, W., Altay, S., & Ittefaq, M. (2025). How effective are fact-checks in Pakistan and who engages with them? *Information, Communication & Society, 28*(14), 2439–2461. https://doi.org/10.1080/1369118X.2024.2445636

Freiling, I., Krause, N. M., Scheufele, D. A., & Brossard, D. (2023). Believing and sharing misinformation, fact-checks, and accurate information on social media: The role of anxiety during COVID-19. *New Media & Society, 25*(1), 141–162. https://doi.org/10.1177/14614448211011451

Lavilles, R. Q., Naga, J. F., Tinam-isan, M. A. C., Perez, J. E., & Palad, E. B. B. (2023). Information verification practices and perception of social media users on fact-checking services. *Journal of Information Science Theory and Practice*, *11*(1), 1–13. https://doi.org/10.1633/JISTaP.2023.11.1.1

Park, C. S. (2023). Why people rely on fact-checkers? Testing theses of "perceived severity of fake news" and "disappointment in news media." *Journalism Studies, 25*(1), 1–18. https://doi.org/10.1080/1461670X.2023.2289878

Seelam, A., Paul Choudhury, A., Liu, C., Goay, M., Bali, K., & Vashistha, A. (2024). "Fact-checks are for the Top 0.1%": Examining reach, awareness, and relevance of fact-checking in rural India. Proceedings of the *ACM on Human-Computer Interaction, 8*(CSCW1), Article 56. https://doi.org/10.1145/3637333

Table A2. *Studies on community-based fact-checking (n = 9).*

Authors	Trust	Scalability	Impact	Platform and method
Allen et al. (2022)	N/A	N/A	Moderate; partisan bias may weaken corrective effects, limits visibility and diminishes real-world influence of Birdwatch notes.	X-platform. U.S.-oriented. Analysis of Twitter/Birdwatch data on tweets, notes (>4,9K), and ratings (>28K). Concerning how political partisanship influences users' behavior within Twitter's experimental fact-checking Birdwatch.
Allen et al. (2024)	N/A	High; 97% of notes found to be scientifically accurate. Posts with the studied notes reached over 200 million views. Number of COVID-specific notes was still low relative to the problem's size.	N/A	X-platform. U.S.-oriented. Expert assessment of 205 Community Notes on COVID-19 vaccination. Notes gathered from open data.
Chuai et al. (2024)	N/A	Moderate; substantial participation in Community Notes (>64K contributors in study sample).	Low; no significant effect of Community Notes, assumed due to timing delays. (This in contrast to Twitter/X's own studies during pilot, where an effect of notes was found—likely due to timing issues).	X-platform. U.S. and global. Analysis of Notes (>68K), ratings (2,3M), and enrolled contributors (>64K) during U.S. and global implementation of the Twitter/X Community Note feature (2021–2023).

Authors	Trust	Scalability	Impact	Platform and method
Drolsbach & Pröllochs (2022)	N/A	High; Birdwatch found to leverage large numbers of ordinary users to identify, annotate, and contextualize misleading content in real time.	Moderate; Birdwatch notes associated with misleading posts receiving fewer retweets than non-misleading posts, indicating reduced diffusion.	X-platform. U.S.-oriented. Behavioral and observational data analysis of Birdwatch.
Drolsbach et al. (2024)	Moderate; Community Notes were rated as more trustworthy than simple mis-information flags. Skepticism among politically misaligned users.	N/A	Moderate; Community Notes increase participants' ability to correctly identify misleading posts compared to no fact-checking and simple flags. Variation across partisan lines.	X-platform. U.S. questionnaire-based experiment (N = 1,810) involving a sample of 36 misleading and non-misleading social media posts. Assessment of users' trust in different types of fact-checking interventions
Kuuse et al. (2025)	N/A	Moderate; Community Notes demonstrate potential for wide coverage and volume processing, yet qualitative consistency varies across ideological and contextual dimensions.	N/A	X-platform. English language. Large-scale computational content analysis of 323,382 Community Notes, collected between Jan. 2021 and Nov. 2023, using bias scoring and sentiment analysis.

Authors	Trust	Scalability	Impact	Platform and method
La Barbera et al. (2024)	N/A	Moderate; scales well via crowd-sourcing, but high consensus threshold limits broad visibility.	N/A	X-platform. U.S.-oriented. NLP-based lexical, sentiment, and temporal analysis; source bias via media monitors. Challenge with ecological validity (what is tested, is whether anyone can fact-check in view of a selection of parameters).
Pröllochs (2022)	Moderate; struggles with consensus. Skepticism about bias but recognizes transparency mechanisms.	N/A	Moderate; Community Notes reduce tweet visibility and engagement but may be unevenly distributed due to polarization.	X-platform. (Birdwatch on then Twitter). U.S.-oriented. Behavioral data from 11,802 notes analyzed.
Zhou et al. (2025)	N/A	N/A	Moderate; Community Notes can decrease audience engagement with labeled authors long-term but boost short-term engagement.	X-platform. Region not specified. Analysis of Community Notes, investigating audience engagement with authors of misinformation.

References (Table A2)

Allen, J., Martel, C., & Rand, D. G. (2022). Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *CHI '22: Proceedings of the conference on human factors in computing systems.* Association for Computing Machinery. https://doi.org/10.1145/3491102.3502040

Allen, M. R., Desai, N., Namazi, A., et al. (2024). Characteristics of X (formerly Twitter) Community Notes addressing COVID-19 vaccine misinformation. *JAMA, 331*(19), 1670–1672. https://doi.org/10.1001/jama.2024.4800

Chuai, Y., Tian, H., Pröllochs, N., & Lenzini, G. (2024a). Did the roll-out of Community Notes reduce engagement with misinformation on X/Twitter? *Proceedings of the ACM on Human-Computer Interaction, 8(*CSCW2), Article 428. https://doi.org/10.1145/3686967

Drolsbach, C., & Pröllochs, N. (2022). Diffusion of community fact-checked misinformation on Twitter. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), Article 428. https://doi.org/10.1145/3610058

Drolsbach, C. P., Solovev, K., & Pröllochs, N. (2024). Community Notes increase trust in fact-checking on social media. *PNAS Nexus, 3*(7), pgae217. https://doi.org/10.1093/pnasnexus/pgae217

Kuuse, S. F., Kangur, U., Chakraborty, R., & Sharma, R. (2025). Crowdsourced fact-checking or biased commentary? Analyzing political bias in Twitter's Community Notes. In WWW '25: *Companion Proceedings of the ACM on Web Conference 2025* (pp. 2661–2669). Association for Computing Machinery. https://doi.org/10.1145/3701716.3717533

La Barbera, D., Maddalena, E., Soprano, M., Roitero, K., Demartini, G., Ceolin, D., Spina, D., & Mizzaro, S. (2024). Crowdsourced fact-checking: Does it actually work? *Information Processing & Management, 61(*5), Article 103792. https://doi.org/10.1016/j.ipm.2024.103792

Pröllochs, N. (2022). Community-based fact-checking on Twitter's Birdwatch platform. *Proceedings of the International AAAI Conference on Web and Social Media, 16*(1), 794–805. https://doi.org/10.1609/icwsm.v16i1.19335

Zhou, Y., Hou, J., Gao, Y., & Chen, P.-Y. (2025). How does crowdsourced fact-checking approach tackling misinformation affect audience engagement? Evidence from Twitter's Community Notes program. In T. X. Bui (Ed.), *Proceedings of the 58th Hawaii International Conference on System Sciences* (pp. 6554–6561). University of Hawaii. https://doi.org/10.24251/HICSS.2025.783

Table A3. *Comparative studies of professional and community-based fact-checking (n = 5).*

Authors	Trust	Scalability	Impact	Platform and method
Liu et al. (2025)	Moderate; professional fact checkers perceived as significantly more credible than news outlets and social media, but at same level as community notes and AI-based fact-checking.	N/A	Moderate: reduction in misinformation for all fact-checking, but no significant difference between professional fact checkers and community notes.	Social media, news media, and AI-based fact-checking. U.S. online experiment (N = 859) comparing professional fact checker (FactCheck.org), community notes (Birdwatch), and others, including news media, social media, and AI-based fact-checking.

Authors	Trust	Scalability	Impact	Platform and method
Pilarski et al. (2024)	N/A	Professional Moderate; broader coverage and more diverse target base, less visibility in high-traffic threads may limit reach.	Professional N/A	X-platform. U.S. Comparative analysis of Community Notes (>25K) and replies with links to professional fact checkers (52K), during Community Notes' pilot phase (Jan. 2021–Oct. 2022).
		Community Moderate; target highly viral and visible posts, but slow response time.	Community N/A	
Primig (2022)	Professional Moderate; (dpa, Correctiv) were perceived as more trustworthy but also perceived by some as state-aligned or propaganda.	Professional Low; real-world reach is limited: many in the sample had not encountered the fact-checks organically and only reacted when shown to them in the study.	Professional Moderate; reduced misinformation among participants who trust media and institutions.	Facebook. Germany. Investigates how German Facebook users assess the credibility of fact-checking sources (dpa, Corrective vs a fictitious "Facebook Community") and messages related to COVID-19 disinformation (N = 123).
	Community Low; "Facebook community" source was rated as least credible.	Community N/A	Community Low; no evidence of change in perception of community-checked content.	

Authors	Trust	Scalability	Impact	Platform and method
Saeed et al. 2022)	Professional High; experts rely on relatively small set of high-quality sources to verify claims.	Professional Low; scale globally but are slower and less responsive to novel mis-information.	Professional Moderate; have longer-term and institutional influence.	Birdwatch (Twitter/X): U.S. Comparing crowdsourced fact-checking (Birdwatch) with expert fact-checking and using computational methods. ClaimReview checks contain 76,769 links covering 73 domains of fact-checking groups and journalists.
	Community Moderate; cites diverse sources, most evaluated as credible and useful, but are more vulnerable to partisan bias, credibility challenge identified as malign users may game the algorithm.	Community High; rapid crowd response (~10× faster than experts).	Community Moderate; high potential impact due to speed and breadth, but - vulnerable to partisanship.	

Authors	Trust	Scalability	Impact	Platform and method
Zhao & Naaman (2023)	Professional High; crowdsourced fact checkers cite professionals heavily (454/897 cases). Dominates expertise-dependent topics (medicine and health, international affairs).	Professional Low; focuses on complex, low-volume topics. Slower response (>24hr median for new claims).	Professional Moderate; critical impact in specialized domains (COVID-19). Enables crowdsourced scale (50% citations).	A variety of platforms. Taiwan. Compares a fact-checking community (Cofacts) with two professional sites (MyGoPen and Taiwan FactCheck Center), on "variety" (breadth of coverage/topic coverage), "velocity" (speed in answering new requests), veracity (accuracy, reliability, and objectivity), and viability (persuasiveness and effectiveness of fact-checks).
	Community N/A	Community High; higher volume (e.g., 11,747 vs. 418 health articles). 57% faster response for new claims (faster in 754/879 cases).	Community N/A	

References (Table A3)

Liu, X., Qi, L., Wang, L., & Metzger, M. J. (2025). Checking the fact-checkers: The role of source type, perceived credibility, and individual differences in fact-checking effectiveness. *Communication Research, 52*(6), 719–746. https://doi.org/10.1177/00936502231206419

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*(1), 1262–1275. https://doi.org/10.1609/icwsm.v18i1.31387

Primig, F. (2022). The influence of media trust and normative role expectations on the credibility of fact checkers. *Journalism Practice, 18*(5), 1137–1157. https://doi.org/10.1080/17512786.2022.2080102

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

Zhao, A., & Naaman, M. (2023). Insights from a comparative study on the variety, velocity, veracity, and viability of crowdsourced and professional fact-checking services. *Journal of Online Trust and Safety, 2*(1). https://doi.org/10.54501/jots.v2i1.118