

Jena, December 10th, 2025

Assessing Meta's Plans to Expand Community Notes

Comment by the Jena Institute for Democracy and Civil Society

The idea behind Community Notes is that users should fact-check content to the best of their ability, thereby contributing to more effective moderation and fact-checking and reducing bias. Developed by Twitter in 2021, the program was scaled up and opened to all users in 2023. Since then, over 2.2 million Community Notes have been published in various languages on X.

Insights about the usage of Community Notes on X can contribute to a better understanding of the risks and opportunities of this tool in terms of more collaborative platform moderation, including on Meta platforms. Several studies have examined Community Notes on X, but few have addressed national peculiarities. In a recently published study by the Institute for Democracy and Civil Society, we identified 48,149 German-language Community Notes for subsequent analysis.

In the German-language context, we found that:

1. Only a few community notes are published as helpful (only 2,228 out of 35,508 German community notes — 6%). The ones that are published are released late, on average (median) 19 hours after publication.
2. Content that is polarizing is unlikely to receive a published Community Note. Participation among German-speaking users is low, and bipartisan consensus requires voters with different political backgrounds. This explains why 91% of all German Community Notes do not receive enough votes. In this context, potentially helpful community notes also remain unnoticed by the public.
3. Of a sample of German Community Notes, a majority (55.5%) are not concerned with fact-checking, but rather with opinion-making, mischief, and discussion threads about the quality of other Community Notes.

We will address each point in more detail below and discuss the mechanisms behind them.

1. Publication rates and engagement decrease

Looking at German-speaking Community Note takers, we can see that there are few incentives to engage in the long term. Of the 7,353 people who have written Community Notes, 3,175 have only written one. In contrast, 50% of all community notes have been drafted by only 7% of community note takers. This reflects an imbalance, showing that relatively few people participate in the program and that engagement rates are unevenly distributed.

Furthermore, there is a significant disparity between the number of requested and written Community Notes. Only 10% of requested community notes are written online, and fewer than 1% are published. This discrepancy underlines that the Community Note system does not encourage enough voluntary participation. This problem is exacerbated in smaller language communities.

Since March 2025, we have seen a decrease in participation in the Community Note program in Germany and globally. This can be attributed to the rise of verifying AI chatbots, such as Grok, which are now used much more often than Community Notes. The Community Notes program, as it is currently implemented on X, does not reward users who voluntarily moderate the platform with sufficient recognition. There is little public recognition, which motivates other collaborative knowledge production projects.

Recommendation: Before implementing Community Notes globally, Meta must implement measures to encourage user participation in the program.

2. Polarizing content and bridging-based algorithms

The evaluation of a Community Note's ratings on X relies on a bridging-based algorithm that incorporates users' opposing "perspectives." Users are assigned to a particular "perspective" based on their previous rating decisions. This procedure ensures that only Community Notes that are widely agreed upon by large groups of users are displayed publicly.

While this mechanism primarily serves to safeguard the quality of Community Notes, it also makes it more difficult for a note to be displayed. In this sense, the algorithm operates as a filter for polarizing yet factually accurate content. Although the intention is the opposite, the publication mechanisms contribute to the acceptance of misleading information in polarizing posts. This can lead to further polarization.

Furthermore, community notes written in less widely used languages are at a disadvantage because they can only be rated by a comparatively small user base. Given the already low proportion of Community Notes rated as helpful, a more carefully calibrated adjustment of the algorithm's thresholds is necessary.

Recommendation: Meta should adjust the bridging-based algorithm according to political specifications in specific regions. Meta should furthermore prevent polarizing posts with misleading information from going unchallenged.

3. Community Notes and the efficacy of responses to misleading information

Community Notes, as rolled out by X (and prospectively by Meta), are said to provide important context for misleading information, raise information accuracy, and push back on disinformation. However, many Community Notes do not truly add context to questionable posts.

In our study of German Community Notes, we found that most Community Notes do not aim to correct false claims or provide necessary information. In fact, we found that opinions, off-topic statements, and housekeeping notes outnumber rectifications based on facts.

Many of the notes are not intended for publication. Instead, they aim to discuss which Community Notes are appropriate and which are not. As neither X nor Meta provides a forum for this, the discussion takes place in the community note system. Thus, the number of Community Notes that adhere to the initial idea of collective fact-checking must be considered much smaller in future calculations.

Recommendation: Future rollouts of community notes should provide forums for community note takers to reduce the inappropriate use of community notes. Best practices for community noting are recommended, especially for regions where engagement needs to be fostered.

4. Final remarks

For Community Notes to truly be a tool for less biased, bottom-up moderation, it needs a system of accountability for its participants. The success of this system depends on the long-term voluntary participation of many users, as well as features that recognize their contributions. If this cannot be guaranteed, the Community Notes system will quickly stagnate at an inadequate level. In any case, our study has shown that Community Notes cannot substitute for transparent content moderation.

Maik Fielitz
(Head of Unit on Digital Conflict Studies)

The Institute for Democracy and Civil Society (IDZ) is a research institute dedicated to strengthening democratic culture online and locally. The IDZ combines social science and computational methods to study extremism, polarization, and conflict in the digital world.

The study is available here: <https://www.doi.org/10.58668/matr/08.2>.