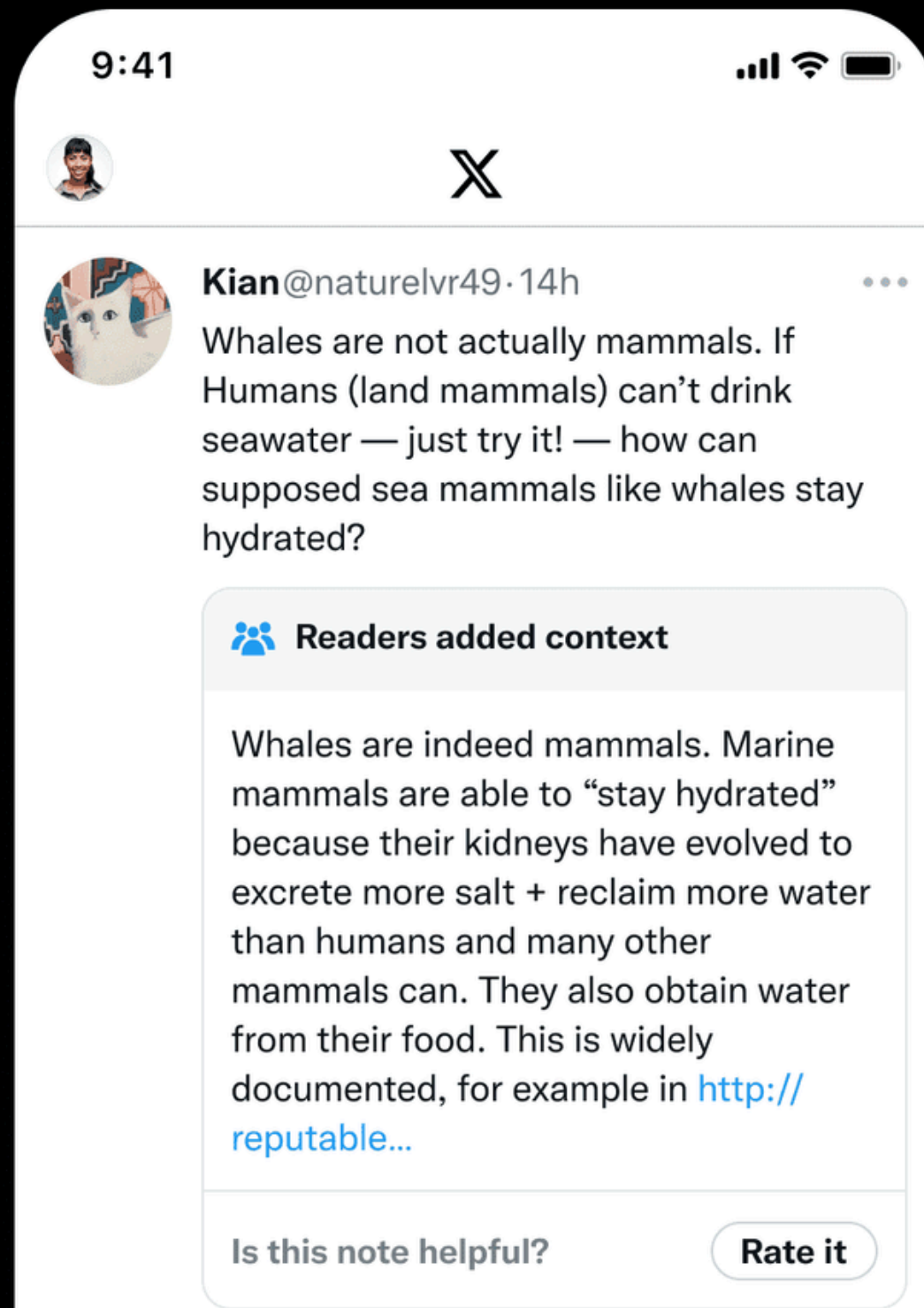COMMUNITY NOTES

What are community Notes

It is a feature which allows users to add additional context to a post, it allows verified users to add explanations, fact checks on misleading posts.

- A user adds community note to a specific post.
- The other users vote if they find the added context/note to be helpful.
- If more people vote the note to be helpful, it gets displayed beneath the post to every user on their feed.

Why Community Notes

By 2020, nearly 5 million tweets were generated or posted daily. With such a huge volume of tweets/posts, misinformation spread increases rapidly, with the introduction of community notes the spread is controlled as it provides more transparency and diverse perspectives to all users. Unlike traditional moderation it does not sensor posts rather it educates the user

Before the introduction, the misinformation usually spreads during the election or the information posted mostly has a political motive to it which is directly consumed by users who need to again fact check this either on the comments of the post or head to any of the search engines to find the validity of the same which was time-consuming and inefficient

User Interviews

1 Jos Smith - 34 M - Software Developer

- **How did you first learn about the Community Notes feature?**

I was quite new to X back then and I started to believe whatever tweets were posted to be true without noticing their auth across the Seattle Times post which had a community note to it that showcased a different dimension to the news posted.

- **How often do you participate in adding or rating Community Notes?**

Whenever I see tweets that contain misleading or fake information attached, I add a community note along with the correct trustworthy source link.

- **What motivates you to contribute to Community Notes?**

In the digital era, I feel all the information from social media should be validated for its authenticity as most of the users thin truth without digging deep.

- **Do you believe the note provided the necessary context? Why or why not?**

Since there aren't any such rules on what exactly must be specified on the note, I feel one can not only raise a flag for misle can even add a different perspective to an issue altogether

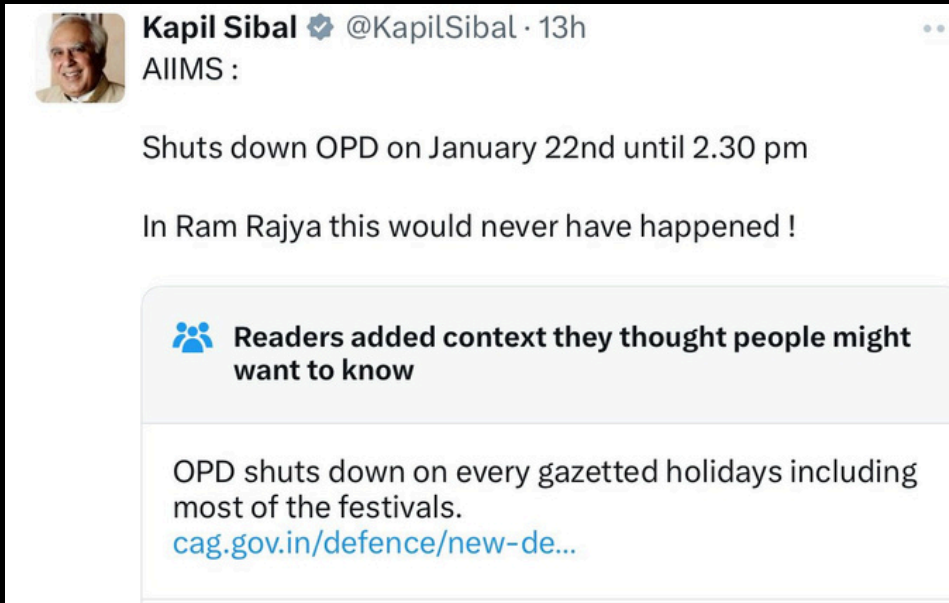- **Do you think Community Notes is an effective tool for countering misinformation?**

Yes, community notes is a great addition as it doesn't have a team to validate the post rather it leaves the ownership of the the bar only if they feel the note was helpful.

- **Would you encourage others to engage with Community Notes? Why or why not?**

Yes I would, knowing the posted information is false Icould move on but there would be many other users who may think the may act accordingly. so anyone coming across any tweets if they find it fake or misleading must go ahead nad add their no note which is already present

User Interviews

2 Simran - 27 F - Doctor

- **How did you first learn about the Community Notes feature?**

I am quite active in X, I used to skim information and follow Bollywood news mostly until I saw a post from an MP who specified a claim that had a note that negated the claim made

- **How often do you participate in adding or rating Community Notes?**

I haven't added a community note till now but have seen many posts which had community notes added in it.

- **Why haven't you used community note, didn't you find any misleading information?**

I don't feel like adding a note to any fake news or post, I have seen lots of misleading or fake news as I follow Bollywood news a lot.

- **Do you believe the note provided the necessary context? Why or why not?**

In most cases, it provides a correct context, but I mostly ignore the notes as I know which posts are true and which arent

- **Do you think Community Notes is an effective tool for countering misinformation?**

Might be, It may be a way but I don't see it eliminates the whole misleading posts on the platform

- **Would you encourage others to engage with Community Notes? Why or why not?**

If people feel they have time for this, I would indeed encourage others to either vote or come up with a note

Problem Framing

Verified/High reach User's post need validations to tackle personal agenda

- When users with large followers or high reach post misleading information, their followers are most likely to follow it.
- Fact-checking and adding a note may take time.
- This information is consumed by users who may not be aware of the personal motive of the user/group who are against an agenda
- This poses a significant difficulty to most of the users in preserving the neutrality and trust in the platform.
- The velocity of number of misleading information increases every day which directly impacts the huge chunk of users.

User Journey

- **Trigger**:

The user sees a tweet from a verified/high-reach account and suspects it may contain misleading information.

- **User Goals:**

The user wants to verify whether the information is accurate and get additional context.

- **How the user accesses the feature:**

The user notices a "Community Note" attached to the post or searches for related fact-checking notes.

- **How the feature fulfills the need:**

The Community Note summarizes verified facts, citing sources that confirm or debunk the original claim.

The user reads the note, evaluates its credibility, and either shares it, reports the tweet, or adjusts their perception based on the provided context.

The user feels informed and confident in understanding the situation, reducing the impact of misinformation.

Solution Analysis

Strengths of current implementation:

- Crowd sourced Fact checking.
- Transparent and publicly available.
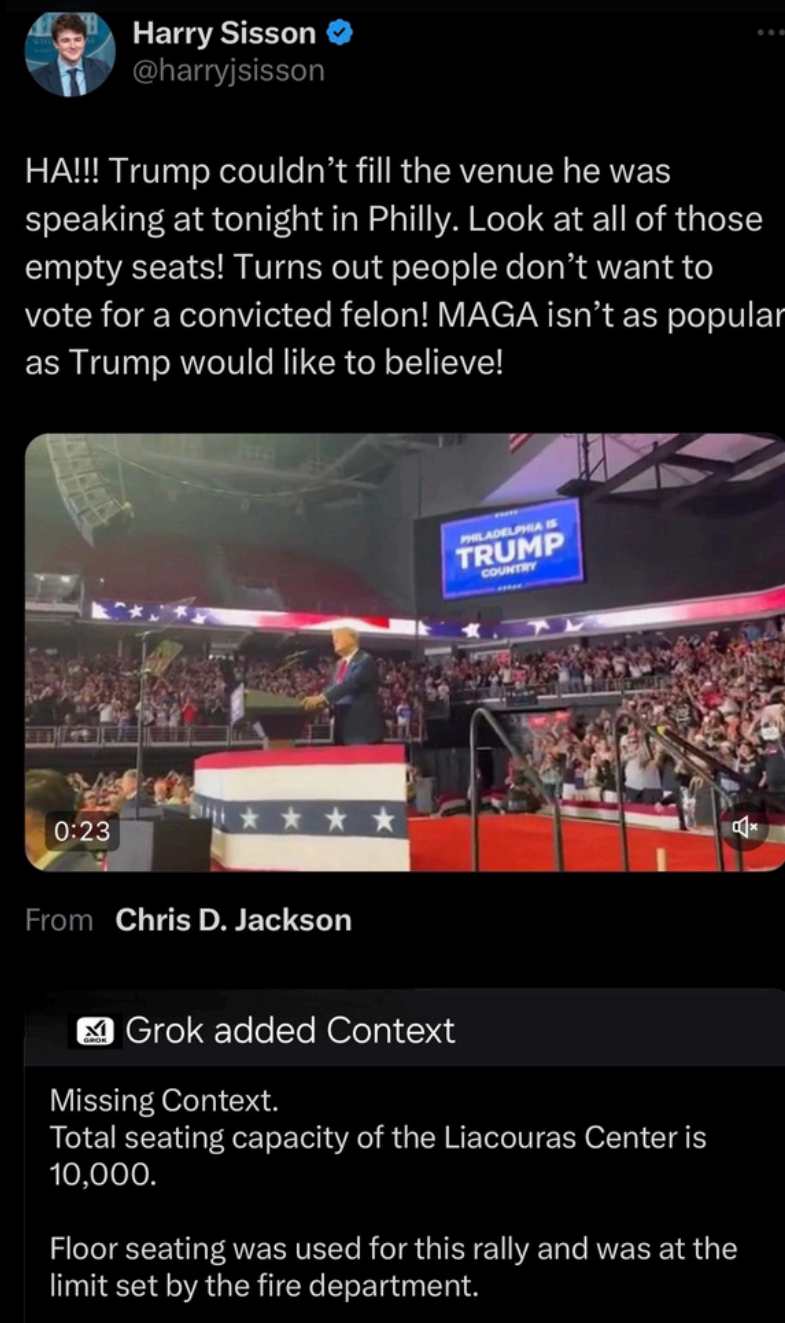- Source based justification.

Limitations of current implementation:

- Slow Response time.
- Dependence on active contributors.
- Does not prevent virality of misinformation.

Solution Analysis

Enhancement Proposal - Faster Fact check using Grok:

- Since the response time for adding a community note is more, it directly impacts the spread of misleading information.
- Grok can analyze trending posts and flag potential misleading information based on past fact-checking patterns, historical data, and known false claims.
- Grok can generate initial draft notes summarizing the claim, referencing fact-checked sources, and providing context.
- Instead of contributors manually searching for references, Grok can suggest reputable sources (news articles, research papers, government reports) that support or refute claims.
- Users who interacted with a flagged post (liked, shared, or commented) could receive AI-powered alerts once a note is added, preventing further spread of misinformation.

Business Considerations

- X the company, ensures people have a free and safe place to talk.
- The introduction of grok improving the response time for the addition of community notes makes the platform even more efficient, responsive, and reliable, ensuring that fact-checking and contextual information are provided in real time. This enhancement reduces misinformation spread, increases user trust, and encourages greater engagement with well-informed discussions, which is inline with the company goals.
- The improved trust makes brands come in for more ads, the users feel safe and post more content which increases the posts/per day which directly impacts the revenue.

Success Measurement

- Objective and key measurements (OKM)
- KPIs on tracing performance like the new response time taken for adding note
- North Star metric   - Feature Adaptation Rate