Emergency Tweets with Follow Tweet Option

**Date:** May 19, 2020

**Author(s):** Charles Baiden, Larissa Hoshiko Suganuma, Carolina Cardozo, Abdiaziz Mohamoud, Aymen Dirar

Usually, there are two documents:

* Product Design Document
  + Product managers provide the feature specification and description
* Technical Design Document
  + Engineering provides the details on how they will implement the feature

We are trying to compress your project design documents into one document as a reference.

# Introduction

* Twitter represents “what’s happening”, therefore it’s essential for this platform to show it’s concerns about current events by raising awareness to emergency causes.
* Currently, there’s no visual difference between regular tweets and tweets that are “emergencies”.
* We want to introduce an “emergency tweet” option for emergency tweets. An emergency icon at the top right of the tweet would highlight it as a concerning topic. (This will help catch the users attention)
* Upon opening an emergency tweet, there will be an option to follow that tweet/thread (via a button) so that users can stay posted on the latest updates. Once following, users would receive notifications regarding that topic of the emergency tweet. (This will help the user engagement with the app)
* In relation to Twitter’s core values:
  + This feature will help earn people’s **#trust** in Twitter as a source for the latest minute-to-minute news regarding emergency situations
  + Some emergencies can be health-related (example below), so by raising awareness, the feature helps promote **#health**.
  + In terms of user-experience on Tweet Details page, the emergency indication via an image and the “Follow Emergency” button are easy enough to interpret and understand even to people who haven’t read the feature announcement, making this feature **#straightforward**.
  + The **#purpose** of this feature is to raise awareness as fast as possible, making it **#fast**. This feature is available at no added cost and is **#free**.

# Use Cases and Scenarios

* A good example to understand the purpose of our feature is the current COVID-19 situation: having emergency tweets will help, for example, fundraisers, professionals in need of equipment such as masks
* Missing person/child: many go to social media to seek help from their followers or authorities. In this case, the “Emergency Tweet” feature will allow people to raise more awareness about the missing person/child. It would attract more attention from the community who would then be able to help and share it with some organizations that may help them.
* Health and Security organizations would be able to use this feature more directly without the actual need to get through the algorithm. Some cases include:
  + Example: @RedCross, @WHO could share vital information about current concerns
  + Example: legal groups such as @ACLU could use this feature to solve cases involving signing petitions
  + Example: local hospitals can request blood donations through emergency tweets to increase awareness
  + Example: @ACLU or other organization which involves signing petitions can improve their collection of supporters.
* When concerned about an emergency situation and want to know how it develops, the “Follow Emergency Tweet” button allows you to get notifications updating you about the situation.
  + Example: in the case of the missing person(s), you’ll be able to get updates as to whether they’ve been found or if there is an active search still going on

# Experimentation

At Twitter, we run experiments to see if a feature will help or hurt our users. In this section try to describe an experiment you might conduct to test if the feature is doing well. Also, try to think about WHAT your feature might impact. At Twitter, we measure several different things. Think about if any of these measurements/metrics might be impacted. Here is a sample list:

* Success Metrics: AA/AB
  + AA -- verify metrics are balanced and random
  + AB -- two distinct versions
    - have groups with balanced/similar metrics: one group actually engages with the new Twitter version that has the emergency feature, and the other group would engage with the current version of Twitter
  + Compare the impact of our feature:

- Compare DAM (Daily Active Minutes)

-sessions/user: how many times the user opens the app in a day

- Compare number of new and active users (increase in DAU)

- Compare number of engagements on tweets that are marked as emergencies

# Technical Overview

## Server-Side Impact

* What are the new services/data that your feature will need?
  + A system would need to be put in place to check whether or not a tweet declared as an emergency is actually an emergency
    - The verification system could filter and text-match keywords in tweets such as “covid19”, “missing person”, “blood donation”, etc.
    - This review process is needed in order to not have users mark anything and everything as emergency. We want tweets that are marked as an emergency to actually be an emergency, rather than having someone mark it as one in order to get more views or for other personal gain.
  + Picture content in case of a picture through ML/AI image reading Algorithms or Optimal Character Recognition
* What are the existing services/data you’ll need access to? Examples:
  + Tweet content
  + User notifications
  + User location services (for local emergencies)

## Client-Side Impact

* + What web pages on the Twitter website will be impacted? We asked that you focus on the Tweet Details page (page that comes up when you click on a designated tweet)
    - * We acknowledge that this isn’t a part of the tweet details page, but here we are providing context as to how a tweet would be set as an emergency.
      * Both text tweets and image tweets should be able to be set as an emergency.
        + Example for image tweets: many missing person(s) tweets are images and we’d like for these tweets to be detected as emergencies.
    - Once designated as an Emergency tweet, an icon will appear on the upper right corner indicating relevance. When you hover over the icon, a text box will appear laying out information on what this new emergency feature is about.
    - In addition, after an Emergency Tweet is posted, users will be able to follow the thread. They will be able to receive future notifications of following information related to the tweet content and replies.
* What sections of the tweet details page will be changed?
  + Tweet Header (top right corner/opposite side of handle): a emergency icon will be shown
  + At the same bar as Retweets/Likes: a new follow button for users to click on if they want to follow the thread

When Composing a new Tweet

* + Button option for the user to mark a tweet as an emergency will be available while writing a new tweet.
  + In regards to the users profile page
* What sections **outside** of the tweet details page would change?
  + In the same way that users can follow lists and have a separate feed for those lists, users will be able to access a separate feed for emergencies for local emergencies/ emergency tweets that they follow.

# Testing & Quality

* How will you test this feature?
  + Check if the emergency algorithm is working properly
    - Rigorous test cases should be designed
    - Test with sample user groups
  + Check if appearance of the emergency tweet changed (emergency icon and follow button is added)
    - Icon: check if the hover function is working correctly and displaying information on emergency tweets.
    - Follow button: make sure the hover function is changing the background color of the button and changing “Follow Emergency” to “Following Emergency”
    - Follow Function: Make sure that users that follow the tweet/thread actually do get notifications
  + Check if users get notifications if they follow a tweet.
    - Have option for users to report not receiving notifications for whatever reason
  + Rollout should be conducted slowly to identify any early bugs

# Questions & Issues

* How to allow pictures to be declared as an emergency.
  + EX: If someone tweeted an image talking about an emergency would they be able to tag the tweet as an emergency tweet?
  + Some AI/ML algorithms can be implemented to detect images
  + Towards the beginning of the development, we will have to trust the users in regards to images/screenshots. Majority of missing persons posts include a picture of the missing person. As time goes on, the goal is for these pictures to be checked as well as the other content of the tweet to make sure it is indeed an emergency
* What would an algorithm with the correct parameters to only verify the correct emergencies and not tweets that aren’t emergencies look like?
* How to stop users from spamming emergencies
  + Should users be temporarily banned for abusing this system?
  + Protection against botting?
* What settings for emergency-related notifications should we offer users?