

---

# Proposal for a Reddit FAQ Responder Bot Application

---

Greene, Joseph  
jgreenemi@gmail.com

August 23, 2017

## 1 Problem Statement

This proposal is for creating an automated system of responding to Reddit user's Frequently Asked Questions. At present, large subreddits that see several thousand users a day will invariably encounter a large number of postings that are readily explained through either previous posts or through the subreddit's Wiki page resources. In an effort to educate the users who make these posts, often times a moderator or regular subreddit user will need to go out of their way to provide the answer to their question as a comment on such posts, thereby reducing the effectiveness of having an FAQ Wiki in the first place. Although having such a Wiki can simplify the process of answering such questions, the process still involves human effort to complete.

The aim of this project is to automate this process with a Reddit bot application. Where a human normally identifies a post with a common question by reading through the post, then typing a comment in response to deliver a suitable answer, a bot can detect the nature of the post and respond with an informative comment to answer the question being asked. The bot can be trained on a subreddit's post history using a convolutional neural network (abbreviated as "CNN") on the normalized text from each post to assign a prediction score, telling of the likelihood that the poster is asking a common question. The bot will leverage the Reddit API for reading text

posts from a specific subreddit, as well as making comments to posts as appropriate.

## 2 Measures of Success

- The number of posts that the bot correctly identifies as FAQs versus those posts that the bot incorrectly marks as such. The goal is a 70% accuracy in identifying posts correctly.
- The number of responses that the bot submits for the FAQ posts that are considered a valuable response, measured by the upvote score on the comment. The goal is an upvote score of at least 1 for 70% of the bot's responses.

## 3 Risks and Failure Cases

There is a risk that the Reddit bot logic could be improperly implemented where posts that get responses may get picked up again in later checks by the bot, and receive multiple responses when this is not necessary. This will be mitigated by having the bot check for posts according to one of two approaches: 1) check only posts that were submitted in the last 10 minutes, or 2) maintain a database of posts that have received responses, such that any post that is already present in the database will not receive further responses. If neither check is implemented, and the bot only checks for whether or not it already has a comment on it, this has the potential to introduce a looping behaviour.<sup>1</sup>

There is also the risk that an improperly trained implementation of the Reddit bot's algorithm could produce unnecessary responses on posts that did not need them, or that posts asking FAQs would be missed entirely. To address this risk, the bot will only be tested against subreddits that have sufficiently large post histories of FAQs to train and evaluate with. Those subreddits with too few samples to provide of both kinds of posts will not be able to take advantage of the bot as the algorithm will be unable to generalize well enough to accurately predict the nature of the posts. The

---

<sup>1</sup>If a subreddit moderator chose to delete the bot's response for some reason, the bot would not know about this and would comment once more.

exact threshold of how many samples will be needed of each kind of post remains to be found as development is not yet underway.

## 4 In Scope

- Create an automated Reddit account that watched a specific subreddit for new self-text posts.
- Extract, normalize and process the post content to determine if the question is likely or unlikely to be a Frequently Asked Question.
- Create logic to respond to each post predicted to be an FAQ, with content and/or links to the subreddit Wiki to answer the poster's question. The content to be posted can include an overview of the information contained in the Wiki, such that it would apply in any question's case (that is, it would not post just the information for just a single specific question).

## 5 Out of Scope

- Answering questions directly. While we could do this, the answers to FAQs should be provided in the form of Wiki links/Wiki content. This will simplify the initial build of the application as we need only do binary classification of whether or not a post is an FAQ or not, rather than which of the FAQs it is asking. This can be done as a stretch goal, as post-launch feature development.
- Flagging or otherwise marking posts as being FAQs after answering them. This would require that every subreddit have post flair set up prior to launching the bot, and would require that the bot have moderator permissions on its Reddit account. This is not necessary for the base build of the application.

## 6 Assumptions

It is assumed that the subreddits for which the bot will be used will have a sufficient number of variations for each FAQ for training against. Sub-

reddits that have too few samples for FAQ posts versus regular ones will likely produce an unreliable algorithm, and will ultimately take the wrong action on posts, such as responding to a post that doesn't pose a common question.

It is also assumed that every subreddit to use the bot has sufficient Wiki content to answer the poster's question. Otherwise the bot will not provide value to the users of the subreddit.

###