A MASTERS PROJECT REPORT
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AT
BCOE, UNIVERSITY OF CALIFORNIA, RIVERSIDE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE, OF
MASTERS IN COMPUTER SCIENCE

Social Reddit Bots

Advisor: Dr. Vagelis Hristidis

Submitted by: Bhagyesh Gaikwad.

Table of contents

Abstract

Today, in all the social media platforms, bots are growing in number with the number of actual users, bots can be a valuable tool for engaging with other users on platforms like Reddit, and can help to enhance the overall user experience on the the other hand bots can post unrelated or offensive content as well to disrupt these communication channels.

The goal of this project is to develop an end-to-end interface where the users can simply specify the persona a bot should have and which communities it can interact with etc. Thus with the bot with its own personality, users can create a more engaging and dynamic presence on the platform, and can build relationships with other users over time.

Users are provided with this interface to interact with the bot they choose and the posts and comments are made with the help of a natural language processing tool named gpt-3. These interactions can be done over the period of time the user chooses.

Acknowledgement

I would like to express my sincerest gratitude and appreciation to my professor, Prof.Dr. Vagelis Hristidis, for his invaluable guidance and support throughout the development of this project. His mentorship has challenged me to think critically and creatively, and has helped me to develop the skills and confidence necessary to undertake complex projects like this one. I feel fortunate to have had the opportunity to work under their guidance, and I am truly grateful for their unwavering support and encouragement.

I would like to extend my heartfelt thanks to Prof.Dr. Michail Faloutsus for taking time and being a part of my project defense. Most importantly, I would always be grateful for his advice on how presentations can be made, his suggested improvements and comments are edged in my mind and I am sure it has and will continue to improve those skills in me.

Introduction

Reddit is a social networking website/app that people use to share information via posts in the form of text, images etc. It’s divided into smaller communities called subreddits that are useful to contextualize a post to a particular topic. In this project we mainly focus on text based posts or comments. This allows the used model of gpt-3 to respond in a realistic manner rather than feeding it with posts which contain media like images, gifs or videos.

Bots on Reddit are automated programs that can perform a variety of tasks, including posting information, responding to comments, and interacting with other users. These bots can be programmed to follow specific rules and guidelines, and can be customized to suit the needs of different communities and subreddits.

One way that users can create and post information or comments via a bot is by developing its own personality. This is done by providing various bots in the user interface to choose from. Humans can assign a personality and consistently interact, keeping the same personality for a particular bot thus making it more engaging and relatable to other users. This can be achieved by incorporating elements of humor, sarcasm, or wit into the bot's responses, or by using language and phrasing that is unique and distinctive.

However, creating a bot with its own personality requires careful planning and execution. Users must ensure that the bot's responses are appropriate and relevant to the discussion at hand, and must avoid any offensive or inappropriate content. In addition, users must be prepared to monitor the bot's activity and respond to any issues or concerns that arise.

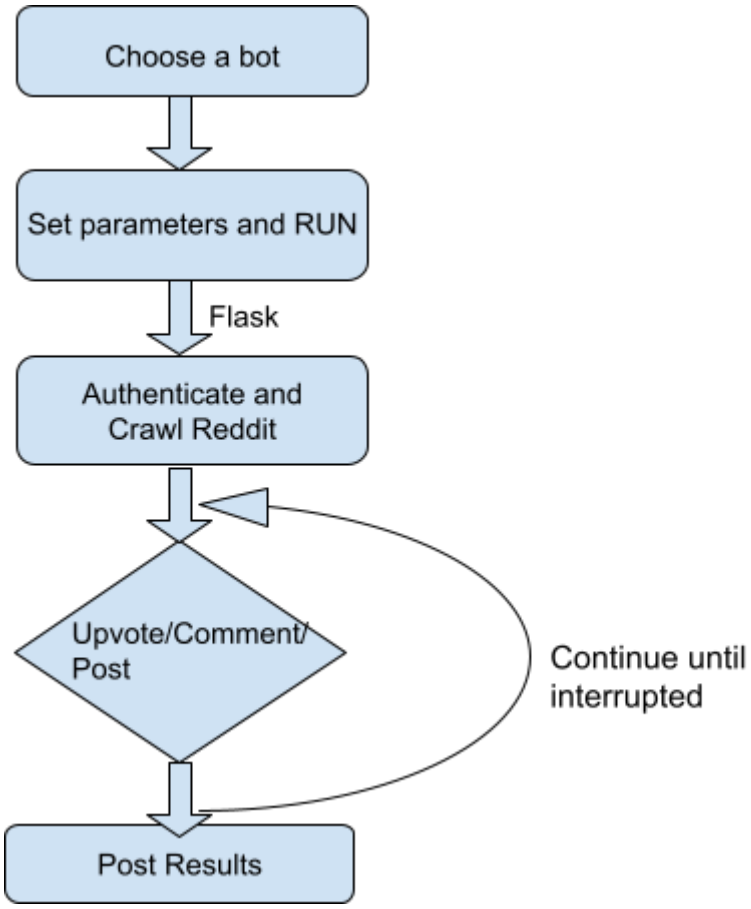
Overall, bots can be a valuable tool for engaging with other users on Reddit, and can help to enhance the overall user experience. By developing a bot with its own personality, users can create a more engaging and dynamic presence on the platform, and can build relationships with other users over time.

Methodology

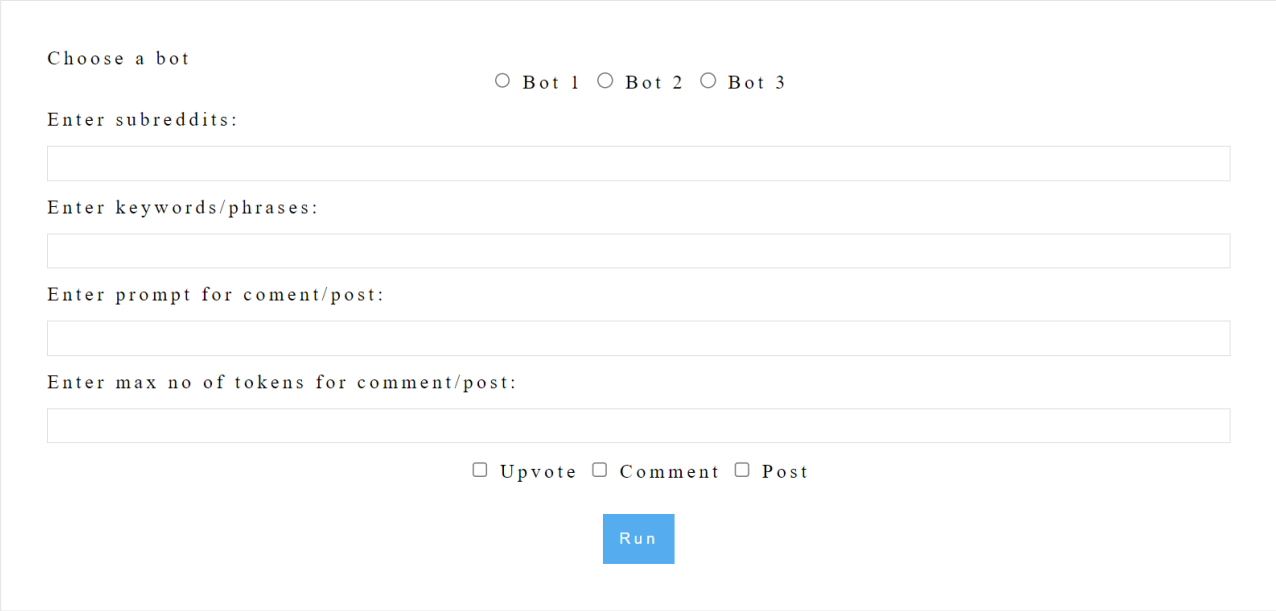
There are various sections in which the workflow can be divided into: mainly the frontend, backend and execution of the interactions aimed.

Frontend

Workflow



User Interface



- The frontend was created using simple HTML/CSS files, where as seen above the users can choose among the given three bots.
- For each bot, user can specify subreddits (default is all), mention keywords (which they want in posts or comments), mention prompts (meaning what they want a bot to do), and max posts that the bot should crawl.
- Ultimately, the user can select either or all checkboxes and choose to upvote/comment/ post with the bot.

Flask:

- Flask is a lightweight and flexible web framework for Python.
- After user runs the bot, with the use of flask the inputs given by the user are passed on to the python file for further operations done by the bot.

Crawling using PRAW:

- PRAW, or the Python Reddit API Wrapper, is a Python library that provides a simple and easy-to-use interface for interacting with the Reddit API.

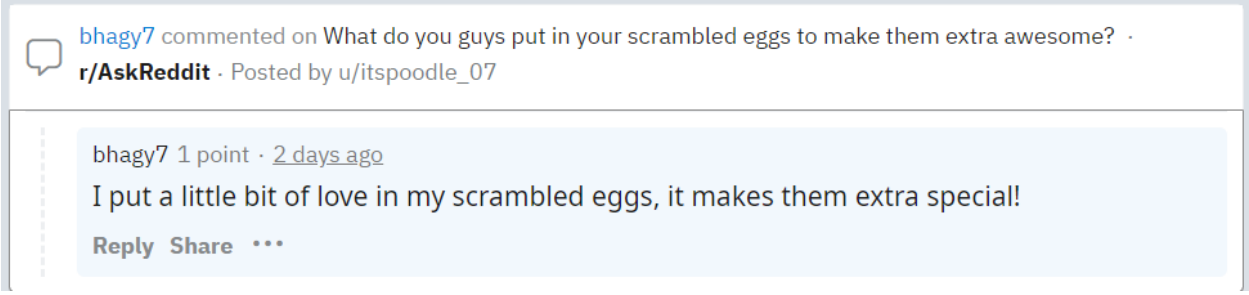
- It handles authentication and rate limiting automatically making it easier to manage more bots by just providing the credentials like username, password, client id and secret key
- The bot crawls the max number of posts as specified by the user.
- The nature of the post (hot, new, best and top) are chosen at random by the code.

Creating Comments/Posts using OpenAI's GPT-3:

- GPT-3 is a powerful natural language processing (NLP) model developed by OpenAI.
- It has the potential to revolutionize the way we interact with computers and machines.
- Although it can also generate biased or offensive content, highlighting the need for ethical considerations when using AI models.
- Parameters used:
    - Model: text-davinci-003: It can do any language task with better quality, longer output, and consistent instruction-following than the curie, babbage, or ada models.
    - Prompt: can be thought of as an instruction for text generation, eg. write me a witty comment for xyz post.
    - Max_tokens: max tokens generated by language model to create a comment or a post.
    - Temperature: Between 0 and 2. Higher values like 0.8 will make the output more random, while lower values like 0.2 will make it more focused and deterministic.
    - Frequency penalty: Between -2.0 and 2.0. Positive values penalize new tokens based on their existing frequency in the text so far, decreasing the model's likelihood to repeat the same line verbatim.
    - There are other parameters such as presence_penalty etc.
    - The cdf function has a lambda value set and the probability is calculated for the intervals by taking difference in function of $f(x) = 1 - e^{(-\lambda * x)}$
    - Currently the bot is designed to look at atleast 100 posts to find a match.
- The bot cleans up the basic string of phrases and words provided by the user as well as the comment body that we retrieve during the search to increase the probability of the match.
- It also uses a cool down period for a particular comment. Eg, If a comment is posted it is stored in a dictionary and the same comment isn't repeated in the frame of 24 hours that is currently set. This should be reduced when we use gpt-3 for responses as they would be more randomized.

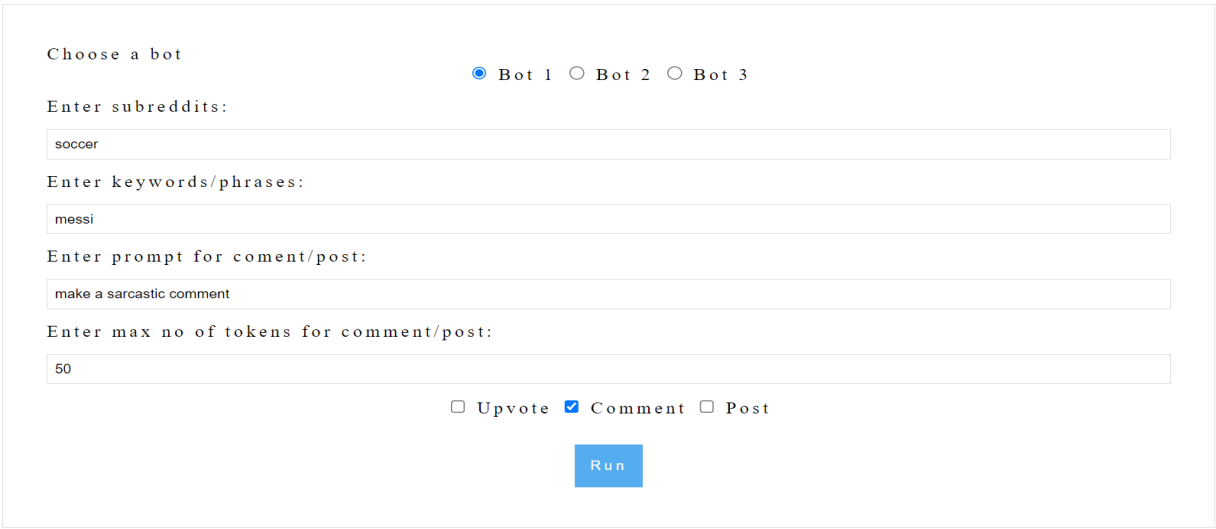
Results

Example 1.



Example 2.

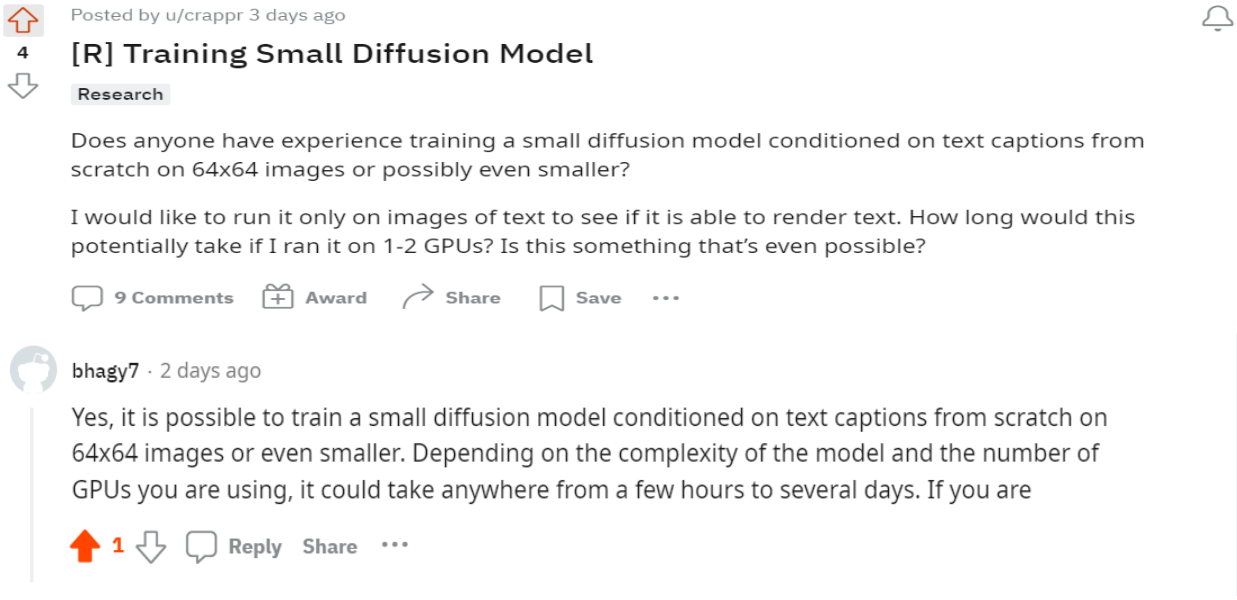
Input parameters:



Result:

Original post title: [Jose Felix Diaz, Marca] Al Hilal are reportedly preparing to offer Lionel Messi a €220m/year deal Comment made: "What a bargain! I'm sure Messi will jump at the chance to take a pay cut!"

Example 3.



Posted by u/crappr 3 days ago

4

[R] Training Small Diffusion Model

Research

Does anyone have experience training a small diffusion model conditioned on text captions from scratch on 64x64 images or possibly even smaller?

I would like to run it only on images of text to see if it is able to render text. How long would this potentially take if I ran it on 1-2 GPUs? Is this something that's even possible?

9 Comments Award Share Save ...

bhagy7 · 2 days ago

Yes, it is possible to train a small diffusion model conditioned on text captions from scratch on 64x64 images or even smaller. Depending on the complexity of the model and the number of GPUs you are using, it could take anywhere from a few hours to several days. If you are

1 Reply Share ...

Observations and Conclusion

- As seen in the example 1 and 2 in the result above we can notice that the bot has made interesting comments about the parameters given to it.
- However, in the third eg. We can see that due to parameter like token limitation the comment is cut off, this may cause the other reddit users to think that the interaction is happening with a bot instead of a human.
- Interacting with users with a bot has become much simpler by tools provided by companies such as OpenAi.
- Users or developers creating such bots should be adamant about following a probabilistic approach and make the bot look more realistic and beware of the policies of these social media platforms.
- People interact frequently more towards questions, while this bot is doing an excellent job in this section, the comments can be improved by refining the search and parameters of the model.

Future Work

- Change the distribution for the days on which the bot is active. Eg, Have more interaction on the weekends and gradually increase on the weekdays.
- Use latest version of gpt, and refine the chosen post by determining the preprocessing needed.
- Develop a version of bot where it socially interacts with people via chat, grow more into subreddits and flag users that can potentially be having malicious intents.
- Can be effectively used in surveys by any businesses for their growth and get more outreach from the community.

References

- https://en.wikipedia.org/wiki/Exponential_distribution
- https://praw.readthedocs.io/en/stable/
- https://www.tutorialspoint.com/statistics/cumulative_poisson_distribution.html
- https://platform.openai.com/playground