

Open Social Comparison and Analysis for  
Reddit

# The Team

---

Anna DeBrunner, [debrunac@mail.uc.edu](mailto:debrunac@mail.uc.edu)

Charles Greenman, [greenmcg@mail.uc.edu](mailto:greenmcg@mail.uc.edu)

Isaac Mathers, [matherit@mail.uc.edu](mailto:matherit@mail.uc.edu)

Advisor: Dr. Raj Bhatnagar

# Goals

---

- Explore users, communities, and topics on Reddit to discover flags of potentially harmful users.
- Present information, comparisons, and recommendations to users to make them more informed about the users and content they are interacting with online.
- Implement data analysis algorithms and visualizations to summarize available information about a user in an easily digestible way.
- Implement graph analysis algorithms and visualizations to summarize available information about user networks in an easily digestible way.

# Project Description

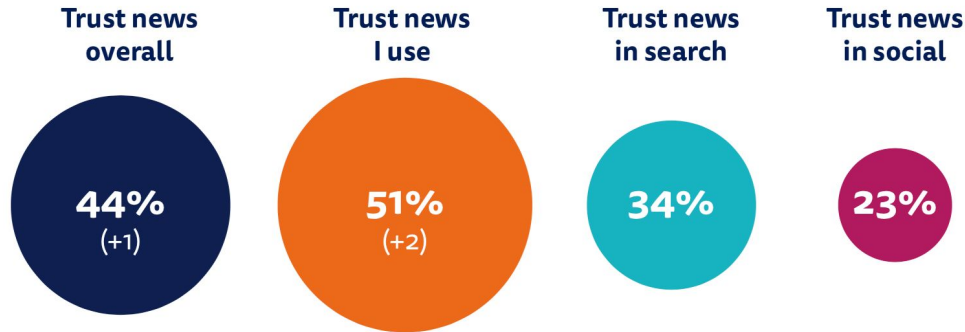
# Abstract/Background

---

The **exploitation of social media networks** came to light following the 2016 US Presidential election. Since then, the websites we use every day have become more aware of the issue, but disinformation has also become more pervasive. With another election cycle just around the corner, we must **arm ourselves with the correct tools and information** to ensure we do not fall prey to **false stories, targeted beliefs, and general discord**. With that in mind, the Open Social Comparison and Analysis for Reddit tool, or OSCAR, hopes to **provide a more informed experience** on Reddit regarding its users and communities.

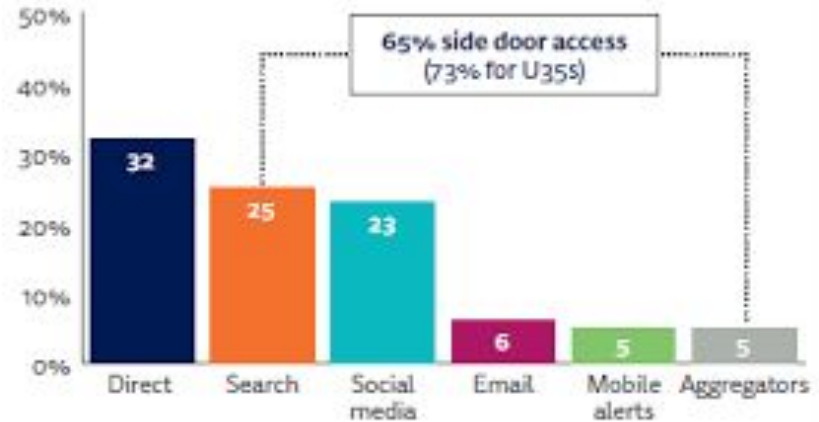
# Background

## PROPORTION THAT SAY THEY TRUST NEWS FROM EACH SOURCE – ALL MARKETS



From a study at the Reuters Institute:  
<https://reutersinstitute.politics.ox.ac.uk/risj-review/trust-misinformation-and-declining-use-social-media-news-digital-news-report-2018>

## PREFERRED GATEWAY TO NEWS CONTENT – ALL MARKETS



From a study at the Joint Research Centre:  
[https://ec.europa.eu/jrc/communities/sites/jrc/files/dewp\\_201802\\_digital\\_transformation\\_of\\_news\\_media\\_and\\_the\\_rise\\_of\\_fake\\_news\\_final\\_180418.pdf](https://ec.europa.eu/jrc/communities/sites/jrc/files/dewp_201802_digital_transformation_of_news_media_and_the_rise_of_fake_news_final_180418.pdf)

# Background

## About one in four report sharing fabricated news – whether aware at the time or not

*% of U.S. adults who say they ...*

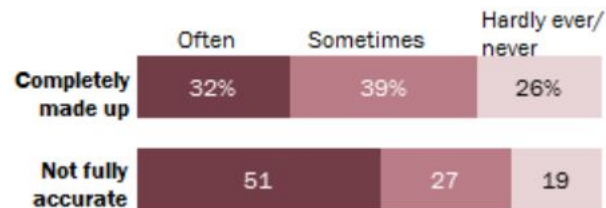


Source: Survey conducted Dec. 1-4, 2016.  
"Many Americans Believe Fake News Is Sowing Confusion"

PEW RESEARCH CENTER

## About a third say they often see made-up political news online; 51% say they see inaccurate news

*% of U.S. adults who often/sometimes/hardly ever or never come across political news online that is ...*



Source: Survey conducted Dec. 1-4, 2016.  
"Many Americans Believe Fake News Is Sowing Confusion"

PEW RESEARCH CENTER

From a study at the Pew Research Center:

<https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>

# Intellectual Merits

---

- The identification of misinformation via **analysis of social networks** rather than analysis of the content has been researched but has not been implemented in a way that can be easily used by consumers.
- We have taken research regarding this topic to create **our own implementation of a recommendation algorithm** that supplies a likelihood that a Reddit user could be a spreader of false information.
- This methodology could be improved and expanded to other social media websites to provide **a new technological solution to the problem of misinformation**.



# Broader Impacts

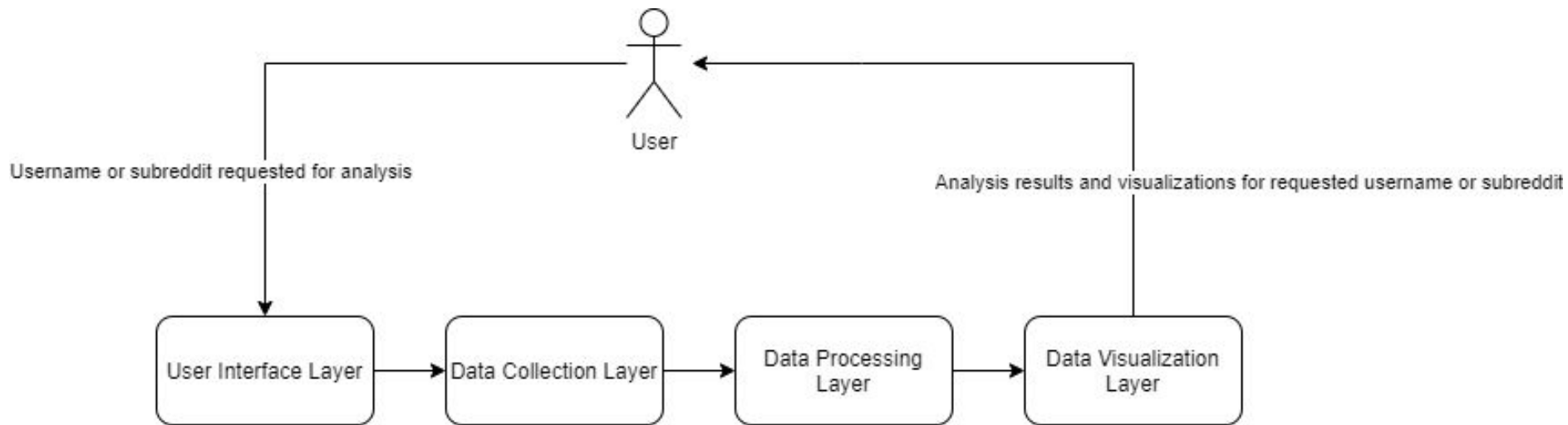
---

- Allows users to educate themselves about who they are interacting with online and determine if these users might have hidden intentions.
- Gives users an indication as to whether or not they should be particularly wary of information posted by a certain user.
- Provides an alternative, analytical way to view users and communities on Reddit.
- Helps fight the intentional and unintentional spread of misinformation, allowing us to live in a more fair and honest society.

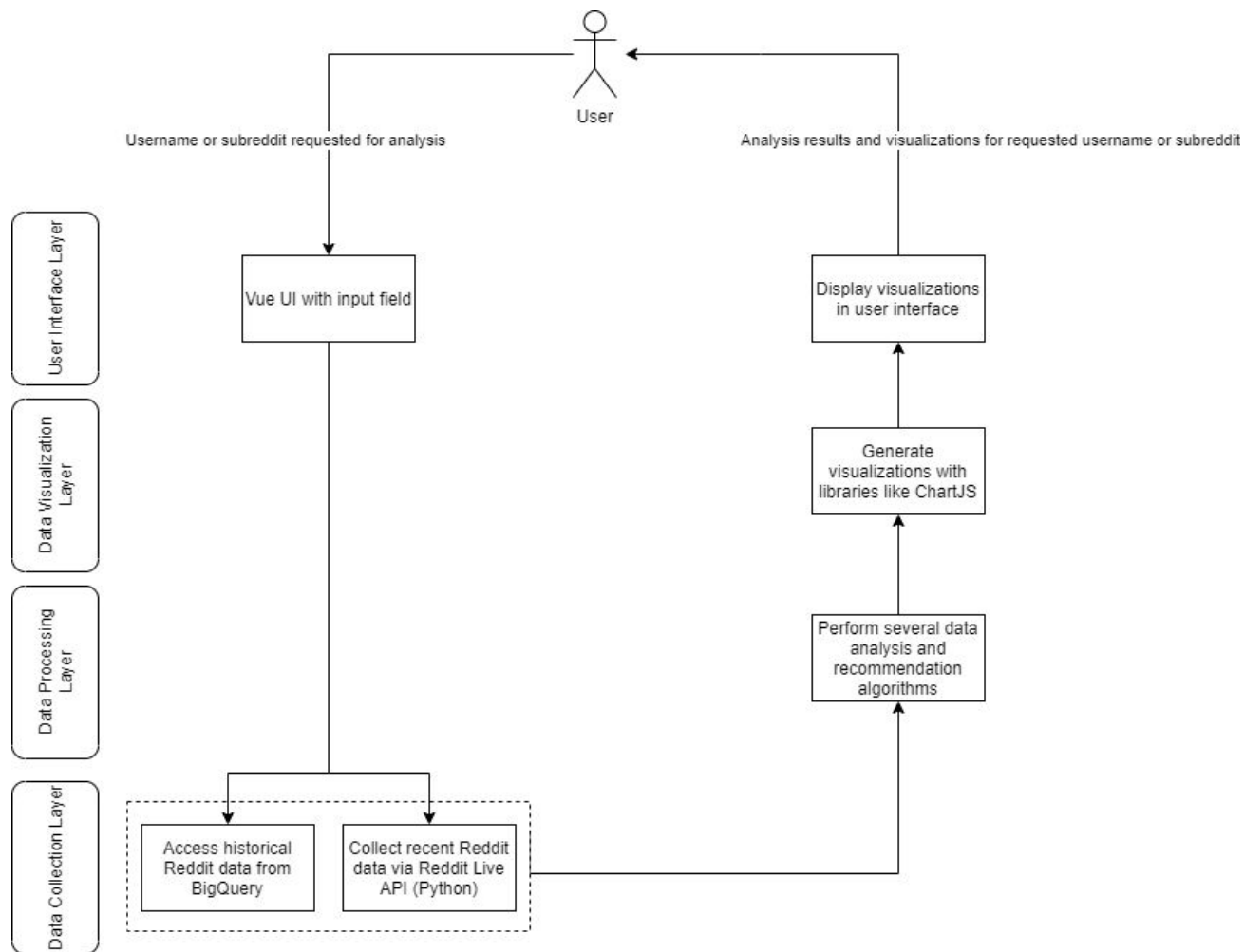
# Project Design

---

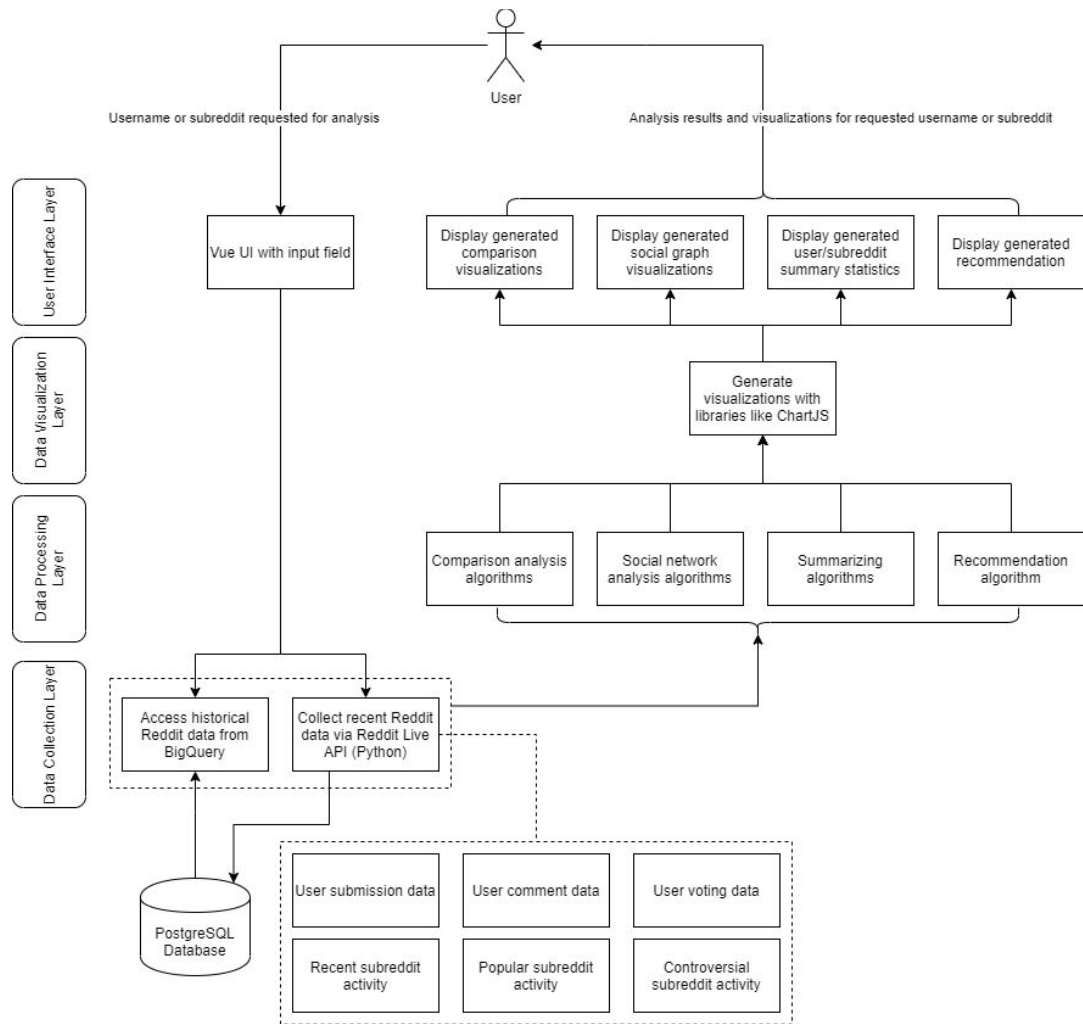
# High Level Diagram



# Mid Level Diagram



# Low Level Diagram



# Technologies

---

- **Reddit Live API** - We created a Python library that uses the Reddit PRAW library to obtain live data from Reddit. We clean that data so it appears in a format that can be used by our summarization and comparison algorithms that also live in this library.



# Technologies

---

- **Python Flask service** - We created a Python Flask service in order to pass information from the Reddit Live API back to the user interface. The user interface can query this service for baseline data or results from our algorithms.



# Flask

# Technologies

---

- **PostgreSQL Database** - We created a PostgreSQL database that contains an aggregated, historic sample of recent Reddit activity. This data is meant to be representative of “typical” Reddit activity around the time that the data was collected. We used research-driven techniques to ensure that we selected a sample that would be representative of Reddit activity as a whole.





# Technologies

---

- **VueJS User Interface** - We are creating a user interface to allow easy access to the insights we are calculating on the back end. The user will be able to submit a user or subreddit they want to analyze and then will be able to view and interact with the output of the analysis. During analysis, the user interface will query the Python Flask application for data and results.

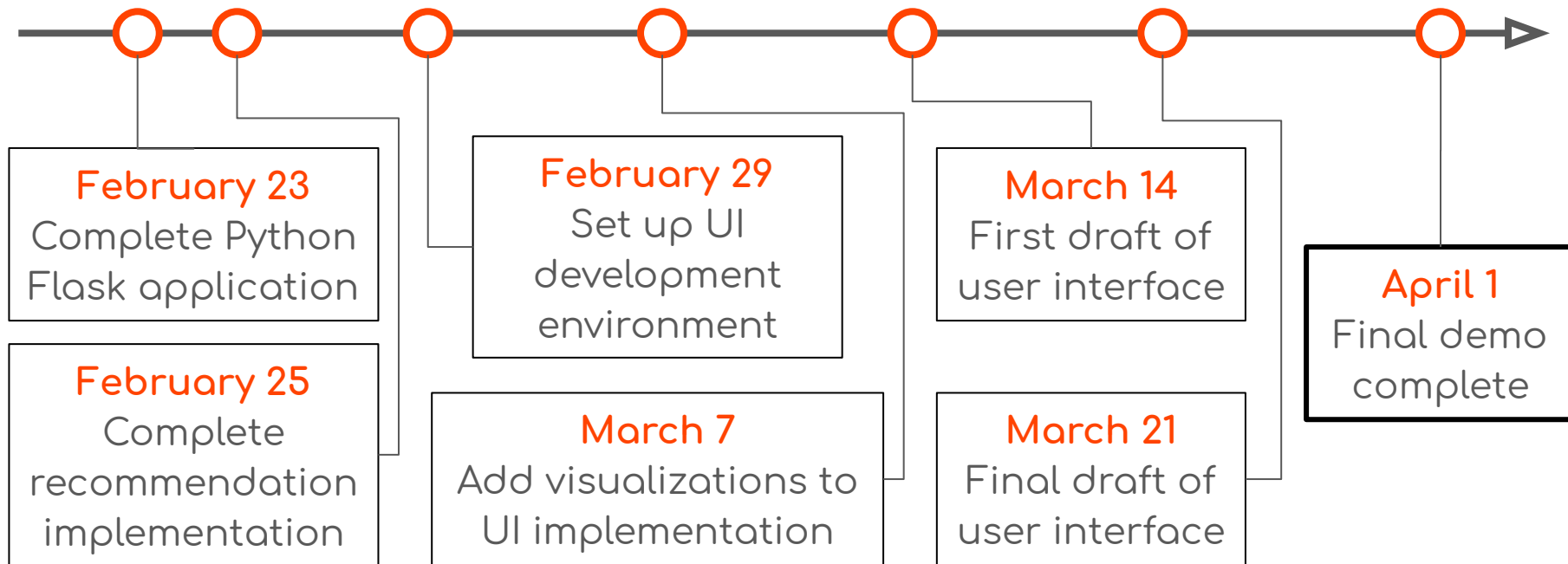


# Timeline and Milestones

---

# Timeline/Milestones

---



# Results

---

- Finalized tools, technologies, and development environments
- Completed Reddit Live API
- Set up database
- Completed baseline data analysis methodology
- Began implementation of recommendation methodology
- User interface mockups

# Challenges

---

- We have had to scale back and change details of our implementation due to limited budget and resources
  - We do not have unlimited memory or processing resources and are attempting to complete this project for free, causing some of our original ideas to become infeasible
- Reddit places limits on the amount of information you can request through their API for minute, so we have to ensure that we adhere to their standards while keeping the analysis time as low as possible.

# Challenges

---

- Since our implementation of the recommendation methodology is new, we have had to do a lot of research into options that would be viable for our product and combine several ideas to find a suitable solution.

Thank you!