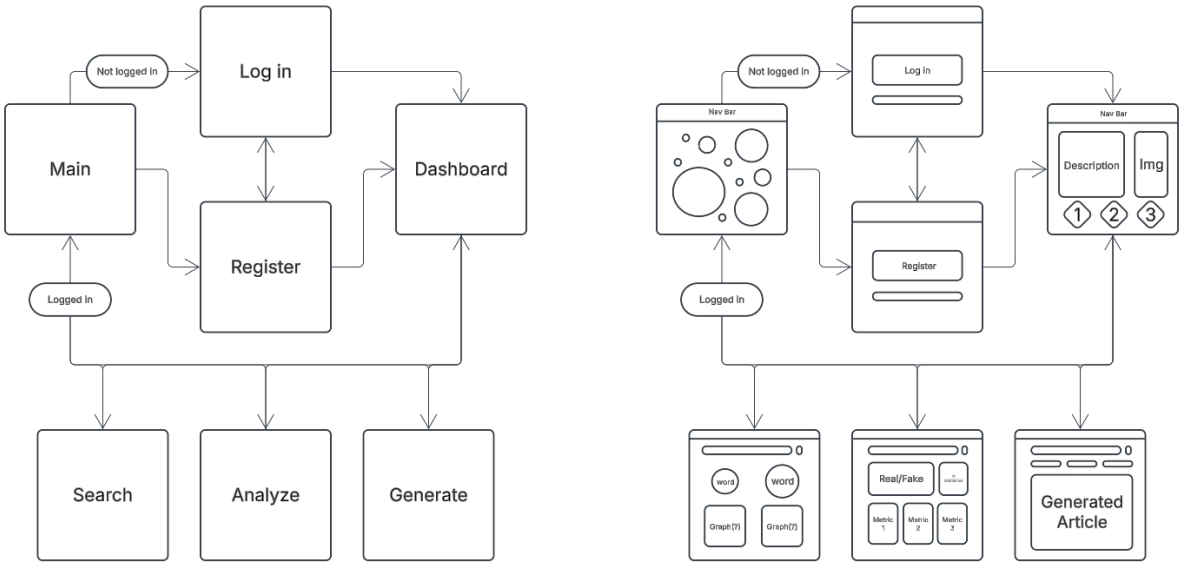
SoftDev P04
TPNG: Salt&Pepper
Roster: Tracy Ye, Princeden Hom, Sascha Gordon-Zolov, Ankita Saha
2025-03-27
Time spent: 2 hours
Target Ship Date: 2025-04-10

Fake News Analyzer

Our website combs through a dataset of over 38,000 fake and true news articles to determine commonalities that are shared between them that may be used to predict other article's trueness or falseness. The first page users will see is a page of randomized "bubbles" which each contain a word. Each bubble's size will be proportional to the prevalence of that word throughout the dataset it was trained on while the color of the bubble will show whether the size is referring to the word's prevalence in fake-news articles or true-news articles. Once the user is logged in, they will have access to more functions, including a search page where they can search a specific word and compare its appearances in fake vs. true articles, an analysis page where the user can input a link to an article on the web and a rating of real or fake will be returned with an analysis of metrics used to reach the conclusion, and lastly a page that allows the user to generate an article that contains either fake or real news related to their inputted word (super stretch goal).

Added ability for users to write their own blocks of text to be rated "true" or "false." Rather than being assigned a boolean true or false value, each article is rather given a *score* (some real number from 0 to 1) determining its truth value. This will turn our model from discrete to continuous, further allowing one to better understand the outputted data and allowing one to draw larger conclusions.

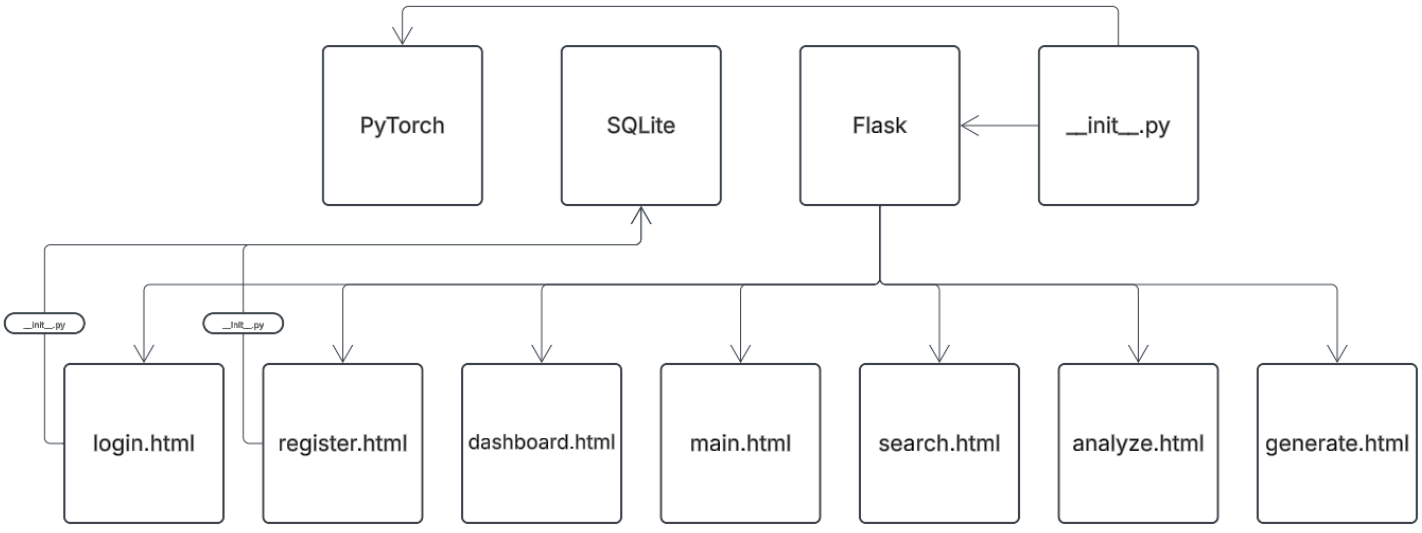
Site Map:



Components:

- __init__.py
  - Main python program that will run the flask application
  - Additionally will contain all data processing functions for the app and interact with the database and apis
  - Utilize sqlite to analyze database and calculate word counts
  - Utilize pytorch to create a fake news classifier
- Register.html
  - Allows user to register a new account
- Login.html
  - Allows user to log into a preexisting account
- Main.html
  - Has redirection to login and register
  - Display bubbles with most popular fake and real news words with size being proportional to prevalence
  - Also contains information about what the website is
- Dashboard.html
  - Contains the bubbles in main as well as other metrics and information we gleam from the dataset
- Generate.html (STRETCH)
  - If we have time we plan to create an algorithm that will be able to generate text that replicates the tone and structure of fake news
- Analyze.html
  - Allows users to upload txt files and will tell the user whether the text is more similar to fake or real news
  - Allows users to write text in their browser, with a dynamic returned truth value (Stretch?)
- Search.html
  - Allows user to input words in a search bar to see the prevalence of that word in the fake and real news articles

Component Map:



Database Organization: SQLite3

user

- Username (Key)
- Password (str)

search

- Username (foreign key reference username in user table)
- Searched Words (str)
- Graph Visuals (output of our algorithm as a str)

articles:

- Title (str)
- Content (str)
- Topic (str)
- Real (boolean)

Datasets

- Fake-News-dataset // Fake News detection dataset

APIs

- Sentiment analysis API
- +others…

Data Visualization Library

We will utilize d3 due to its ability to display data in a unique variety of ways beyond just charts, especially since we plan on using custom shapes like bubbles to display data

Front End Framework: Bootstrap:

We decided to use this because it will fit best for our project components, especially those that require specific stylistic changes (color, border, layout, etc.), and is easily understandable.

- Change color of bubble text or words to differentiate fake news vs real news
- Create layout for data visualizations
- Create formatting for articles (grid structure)
- Create a dropdown for different webpages

Task Breakdown:

Tracy - (PM)

- Research our stretch goal of a fake news generator
- Create a python file that searches for an draws conclusions from the dataset
- Create Main
- APIs

Princeden

- Utilize pytorch to create a fake vs real news classifier
- Research our stretch goal of a fake news generator
- Work on flask app routing for user login and logout

Sascha

- Flask app routing
- Create search function for fake vs real news terms
- Write a function to assign a truth value to a given portion of text
- Research our potential stretch goal of a fake news generator

Ankita

- Comb through necessary sets of data and remove null, duplicate, etc. values and neaten to work smoothly for our project
- Use D3 to visualize data for fake news vs. real news and show growing trends, statistics, probabilities
- Use Bootstrap on finished components to make the site aesthetically pleasing