

## Team-MI Project Progress Report: Identifying Misinformation and Disinformation in Search Results

*Disclaimer: Files listed below do not necessarily represent conclusive results as progress will be made and development will change based on additional research.*

- David Burrus - [dburrus3@illinois.edu](mailto:dburrus3@illinois.edu) – Team lead/Captain
- Jianci Zhai (Angela) - [jianci2@illinois.edu](mailto:jianci2@illinois.edu)
- Ginna Woo - [ginnaw2@illinois.edu](mailto:ginnaw2@illinois.edu)

### Main Tasks:

1. Web Scraper - scrape search results: In development
2. Manual Classification - building classification data to feed to model: In development
3. Model development - code to identify misinformation search results: Not Started
4. Chrome Extension Plugin - Package code into plugin: In development
5. Develop UI for misinformation tagging: Not Started
6. Build report and presentation: Not Started

### Current Progress:

- Teammates have tested various crawler packages available online for Google and Twitter.
- Teammates have explored possible plugin software options. A skeletal frame for the extension is currently being developed. This component will require packaging the entire model and will encompass the development of the UI projection on the search results page.
- Teammates have continued our class and learn more text clustering techniques.

### Remaining Tasks:

- Complete data gathering for the model development
- Provide feedback based on human judgment or online consensus panels (manual classification).
- Start building the analytics algorithm and improve results
- Package the algorithm engine with plugin softwares. This part will encompass the UI development of the code.
- Prepare the final presentation

### Any challenges/issues being faced:

- **Blocked Crawlers:** Our initial scraping attempts on Google and Twitter were unsuccessful because the websites blocked our crawlers
  - **Solution:** we will try more search engines (e.g.: Bing) with other packages (e.g.: selenium) . At the same time we will manually collect same data just in case the new crawlers still don't work
- **Results are Trustworthy:** Most research results from Google are pretty trustworthy, thus creating a biased dataset

- **Solution:** We will extend our searches out of google to search engine websites such as DuckDuckGo, Yahoo Search, and Bing. we will also try to collect data from 4chan and twitter to include more “misinformation”
- **Building Panel of Authorities:** It’s hard to collect “consensus” from the websites defined as “panel of authorities” to provide the final judgment because currently there is no API or other methods to integrate responses from those websites. There is a large number of fact-checking organizations, but with little commonality in rating methods or how in how ratings are made available online.
  - **Solution:** we will manually check the results on each websites and log in their “responses”
- **Learning Curve:** Building an extension from scratch and understanding how each of the file components operate.
- **New Language:** Team is not familiar with Javascript.