Team-MI Project Proposal:
Identifying Misinformation and Disinformation in Search Results

What are the names and NetIDs of all your team members? Who is the captain?

- David Burrus - dburrus3@illinois.edu – Team lead/Captain
- Jianci Zhai (Angela) - jianciz2@illinois.edu
- Ginna Woo - ginnaw2@illinois.edu

What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

Our topic is a "search result markup" browser extension that will flag links known or suspected to contain disinformation and misinformation, so that users can make a more informed decision of whether to click such links.

Dissemination of information that is deliberately misleading (disinformation) or simply inaccurate (misinformation) has been recognized as a threat to society. This is not a new problem: Sensationalistic "yellow journalism" during the 1890s is widely credited with inflaming public opinion and triggering the Spanish-American war in Cuba and the Philippines. More recently, the national cost of COVID-19 non-vaccination due to disinformation and misinformation has been estimated at between $50 and $300 million per day[1].

With the advent of online search engines and social media, disinformation and misinformation can and do spread quickly[2] and irreversibly. This is due partly to the influence of confirmation bias[3], and partly to algorithms deployed by Google and Meta (Facebook)[4].

Provoking content that elicit emotional response from users tend to receive more click-throughs, and is therefore favored by search engines. The search engine will feed users additional recommendations based on the user's behaviors, which can cause authoritative or neutral sources to struggle to gain audience attention. Misinformation can cause users to develop and receive biased opinions, which in turn makes them more likely to search for and read biased content. Such interactions between algorithms and self-reinforcement amplify the impact of misinformation and ultimately leave the audience trapped in an information bubble filled with untruths and deceptions. A user's perception of the information that they receive can be distorted as a result.

Suppose a student is researching the long-term effects of COVID-19 and is presented with a list of search results in a browser. Interested in learning about COVID-19's effects, he selects a paper titled "COVID-19 Long-Term Effects: Humans Turn into Vampires" over other sources that seemed uninteresting to him. By gathering and analyzing this user's interests, the algorithms will then provide the user with more content about COVID and vampires. Such feedback

---

[1] Richard Bruns et al., "COVID-19 Vaccine Misinformation and Disinformation Costs", The Johns Hopkins Center for Health Security, October 20, 2021.

[2] Center for Countering Digital Hate, "The Disinformation Dozen", at https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf.

[3] Liz Hamel et al., "KFF COVID-19 Vaccine Monitor: Media and Misinformation", Kaiser Family Foundation, November 8, 2021, at https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-media-and-misinformation/

[4] Max Fisher, *The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World*, Little, Brown & Company, 2022.

mechanisms can obtain insights about users, but it cannot assist users in identifying misinformation. It has become more difficult to determine which content is based on facts because of the unlimited sources of information available today. The element of measuring the level of trustworthiness should be integrated in an algorithm as users want reliable sources; however, many algorithms fail to deliver. Without acknowledging human beings' own limitations in information screening and judgment, users need technology to help make informative decisions.

By providing a "disinformation score" for each search result, our "search result markup" browser extension will influence user click-through[5] by navigating the users towards more reliable sources. This should offer three benefits: (a) providing users with a dimension other than ranking to enable users to decide whether they should click a link, (b) reducing disinformation spread by reducing its initial consumption, and (c) changing the feedbacks used by Google and Meta to customize their search and recommendation algorithms by encouraging users browse less misinformation.

Briefly describe any datasets, algorithms or techniques you plan to use

It is possible to explore many aspects of this topic beyond what is covered in the course. The scope of this project will be narrowed to a specific topic of interest – COVID-19 disinformation and misinformation – to build a proof-of-concept.

Our scope also is influenced by the knowledge and time required to create a true browser extension of any kind. The members of our team have neither previous experience writing browser extensions nor any particular expertise with JavaScript. We therefore propose a two-phase approach:

Phase 1:
Screen scrape search results from an arbitrary (but COVID-related) user query and extract relevant text data. This might be accomplished using the BeautifulSoup (https://pypi.org/project/beautifulsoup4/) or scrapy (https://scrapy.org/) libraries.

Instead of relying on a group of real users to judge whether a search retrieval result is trustworthy or not, we find a panel of websites dedicated to providing unbiased information related to COVID, and fact-checking widespread misinformation, so we decide to regard the overall consensus given by this "panel of authorities" as the dependent variables. Considering the fast and wide spreading nature of misinformation, it poses great challenges to rely on ordinary users to make a judgment, because their opinions may already be influenced by fake information. By comparison, the "authorities" are usually operated by experts with good domain knowledge or investigating the truth thoroughly, and are believed to be more reliable. On the contrary, troll farms online create and spread fake news seeking to interfere with politics and public opinions. Once identified, we use news/posts from troll farms to provide the additional cues for misinformation. Our ideal is in the future, an API can integrate information from those websites to help search engines improve the algorithm. As for this project, we will collect and augment the information manually. Two reference lists containing sources of "truth" and "falsehood" are described below. We may input these variables as binary values or on a scale by counting the resources cited:

1. Level of "trustworthiness" based on a reference list of COVID-19 "authorities": Such a list might include institutions such as CDC, WHO, NIH, or Johns Hopkins, as well as fact

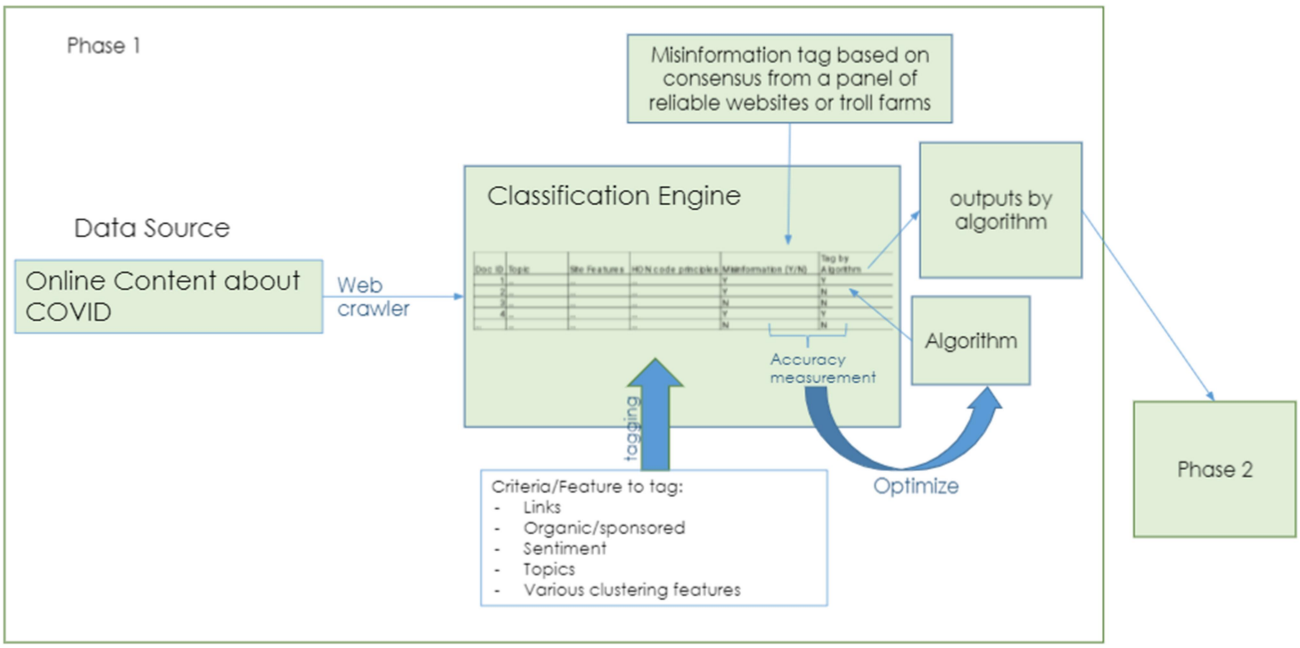
[5] Dongfang Gaozhao, "Flagging fake news on social media: An experimental study of media consumers' identification of fake news", Government Information Quarterly, Volume 38, Issue 3, July 2021.

checking organizations that are not topic specific, such as FactCheck.org, Media Bias/Fact Check, ReportersLab.org, Snopes.com, etc.

2. Level of "falsehood" based on a reference list of COVID-19 "sources of falsehoods": Such a list might include datasets such as ESOC[6], Newsguard's Coronavirus Misinformation Tracking Center[7], NCBI[8], etc. These variables ideally will correlate more with high levels of misinformation.

Useful features or promising independent variables may entail a combination of the following criteria:

1. Whether the search result is organic or sponsored: The presumption here is that sponsored search results are less likely to have an impartial point of view, thus tend to be false.
2. Analysis of the content possibly using sentiment tagging (e.g., words such "evil" or "manipulative").
3. Topic analysis (e.g., "vaccine" and "vampire", or "COVID" and "microchip") using simple bag-of-words approach or more advanced clustering methods.
4. Links: if a website refers to other trustworthy websites then it may be more reliable, and vice versa.
5. Clustering analysis using whatever additional inputs we can get.



Phase 2:

Create a Chrome browser extension that intercepts search results, routes them to the Python app created in Phase 1, waits for the marked-up results, then displays them in the browser by actually editing the HTML output of the search engine. In this way, disinformation flagging

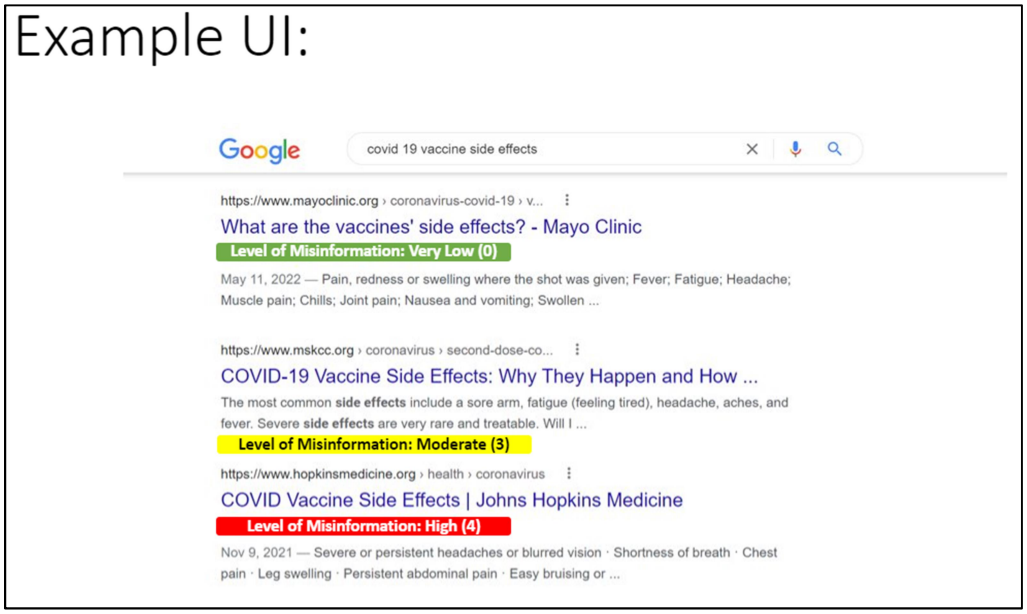
[6] https://esoc.princeton.edu/publications/esoc-covid-19-misinformation-dataset

[7] https://www.newsguardtech.com/special-reports/coronavirus-misinformation-tracking-center/

[8] "COVID-19 Rumor Dataset" at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8200409/

becomes a seamless part of the user's real-time browsing experience. Phase 2 would be started only if we have time after completing Phase 1, and likely would draw upon information at https://developer.chrome.com/docs/extensions/mv3/getstarted/ and elsewhere.

Simulate or "mock up" a marked-up search results screen and display the simulated search results mark-up in a separate browser tab or other window. This might be accomplished using Flask (https://flask.palletsprojects.com/) and could look something like this:



How will you demonstrate that your approach will work as expected?

In a formal study, a randomized study group could be provided with the completed browser extension and a set of queries to execute. This group's click-through behavior would be compared to a control group given the same set of queries but no browser extension. If users are browsing more trustworthy websites/links and less fake news than control groups, then it can prove that our browser add-in can help to hinder misinformation

Which programming language do you plan to use?

Python and readily available Python libraries.

Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

1. Create web crawler to scrape and extract source content on COVID-19 related data= 10-20 hrs
2. Manual classification of the data (this includes determine data features, topic relevance, data tagging, etc.) = 20-30 hrs
3. Develop an algorithm to predict COVID-19 data as misinformation or a source of truth = 20 hrs
4. Package algorithm into a chrome extension = 10-20 hrs
5. Develop user interface (UI) for the misinformation tags = 10-20 hrs.
6. Build report and presentation of findings and development= 10 hrs