# CS 410 (Fall 2023): Final Project Proposal

Team CAHJ

1. **What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.**

| **Name** | **NetID** |
|---|---|
| Christian Opperman (captain) | cjo6 |
| Aaditya Murthy | amurthy7 |
| Himangshu Das | hdas4 |
| Jinfeng Wu | jinfeng4 |

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

   Our group has chosen to do a project in the *Intelligent Browsing* topic. Specifically, we would like to create a program that allows users to search for topics in videos; the search would provide a ranked list of transcript timestamps that relate to the user's search query, allowing the user to browse for the portion of the Coursera video(s) that cover the topic of interest.

   Additionally, time permitting, we would like to expand our search extension with ChatGPT capabilities, allowing the user to use generative AI and LLMs to get information about their search topic potentially not covered in the video.

   Ideally, these functionalities would be packaged together into a Chrome extension that a user could use from their web browser.

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

   We intend to use the web search techniques from class, including inverse indexing, ranking functions, and probabilistic language models, in our project.

4. **How will you demonstrate that your approach will work as expected?**

   The primary test of whether our approach works as expected will be if we are able to create a functioning web browser extension that correctly scrapes, indexes, and searches the Coursera video transcripts. Because search and ranking is an inexact science, there is no exact metric that we can use to scientifically prove accuracy, but we will rely on user feedback and judgement for returned transcript sections to see how our program is working.

5. **Which programming language do you plan to use?**

   We plan to use Python for the indexing and processing of user searches and the ChatGPT integration. We intend to use JavaScript for the Chrome extension.

6. **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.**

   The main tasks we envision are:
   - Coursera login and authentication integration (10 hours)
   - Coursera video transcript scraping and indexing (10 hours)
   - Search function buildout (10 hours)
   - ChatGPT integration (10 hours)
   - Chrome extension and UI buildout (40 hours)