1. According to the National Institute of Mental Health (NIMH), in the year 2017, it was estimated nearly 46.6 million Americans suffer from a mental illness. That translates to roughly 1 in 5 Americans. In the same year, it was also estimated that roughly 11 million Americans have experienced at least one episode of severe and debilitating major depression. Of those who reported they live with a mental illness, only 42.6% received treatment; and nearly 35% of those experiencing a major depressive episode did not receive treatment. Although accessibility to mental healthcare is slowly improving, out-of-pocket costs of treatment, stubborn social and cultural stigmas, and an unmet demand for mental health professionals remain as huge barriers for the most vulnerable groups to find viable treatment options.

   However, an interesting phenomena occurred alongside with the explosive emergence of social media--as global social networks grew, people became more open about sharing their experiences, and giving emotional validation and support to their fellow netizens. In other words, people began hacking a way through the social network landscape to establish support systems which were otherwise limited in accessibility, and/or availability in the "offline" space. This resulted in the massive proliferation of textual data which document the cognitive processes of individuals who may, or may not, have a mental illness, and certainly of those who are going through powerful experiences seeking validation and support. I believe this data presents an opportunity to serve a chronically underserved population, and is also an underrated source for insights in mental health.

   The goal of my project is to use natural language processing to detect common cognitive distortions, as defined by David D. Burns. For the purpose of this project, detecting even one cognitive distortion would suffice. I plan to use Reddit's api to extract text data from subreddits relating to a mental health disorder. I am aware of three attempts done in academia that sought to develop a cognitive distortion detection system. I will use the insights gathered from those attempts to help guide this project.

   The reason for choosing Reddit as my data source is because it is publicly accessible, and mental health subreddit communities usually have large followings with thousands of posts that date back for years. This is a double-edged approach. On one hand, there is a large and accessible dataset available for my usage; on the other, this dataset will be messy, owing to a lack of controls and manipulations of the platform. I expect the most challenging portion of this project will be cleaning the data, and defining cognitive distortions in code.

2. Create an image processing/natural language processing algorithm to recognize text from a book. Use one's smartphone to quickly capture an image of a page, then use an AR interface to point the camera to a computer screen, then transfer the transcribed text file onto a digital format.

3. Create a Yelp data metric for reviewing restaurants' sustainability scores. Living in a COVID era business world, it appears plastic companies are doing better than ever. Time to stick it to the man by convincing restaurants that buying sustainably = profit, and plastics will cost them in the long run.