1. Name of Team and Members
   * Introduce – Emily, Shubhi and Linghan
2. The Issue/ Problem We’re Working On
   * Fakenewschallenge.org - Explore how artificial intelligence technologies could be leveraged to combat fake news
   * The term “fake news” was [almost non-existent in the general context and media providers prior to October 2016](http://signalmedia.co/media-monitoring-blog/fake-news-dissected-media-monitoring/) but times have changed and I would not be surprised if you have heard the term being used today, in the news, the radio or just in the street.
   * Fake news is a term that has been used to describe very different issues, from satirical articles to completely fabricated news and plain government propaganda in some outlets. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems with huge ramifications in our society. However, in order to start addressing this problem, we need to have an understanding on what Fake News is. Only then can we look into the different techniques and fields of machine learning (ML), natural language processing (NLP) and artificial intelligence (AI) that could help us fight this situation.
   * Since the 2016 presidential election, one topic dominating political discourse is the issue of “Fake News”. A number of political pundits claim that the rise of  significantly biased and/or untrue news influenced the election, though a study by researchers from Stanford and New York University concluded otherwise. Nonetheless, fake news posts have exploited Facebook users’ feeds to propagate throughout the internet.
   * The data science community has looked to respond to this problem through a Kaggle competition called the “Fake News Challenge”
3. What Makes It Interesting or Hard.  Your approach to the problem and what AI techniques/algorithms you will explore, develop, implement to address the problem
   * What makes this interesting and hard is even the definition of “fake news”
     1. False connection
     2. False context
     3. Manipulated content
     4. No intention to cause harm but potentially fool
     5. Misleading content
     6. Imposter content
     7. Fabricated content: 100% false
   * Approach: ML based faked news recognition. We are using two data sets, the fakes news dataset and real news. Then abstract linguistic features using tools such as NLTK. Next we will implement Naïve Bayes to predictive fake news vs real news by training the supervised learning model.
     1. Import packages such as pandas, numpy, csv.
     2. Read the csv file and display the data set (use python notebooks)
     3. Remove erroneous null and NAN values
     4. Calculate the length of news to check the length of fake and real news
     5. Data Preprocessing – NLTK tokenization
        1. Tokenize and normalize text
        2. Lemmatization and vectorization (count vectorizer vs tfidf vectorizer)
     6. Feature extraction
     7. Training a model to detect fake
     8. Build & evaluate
   * Dataset 1 (fake) - <https://www.kaggle.com/mrisdal/fake-news>
     1. Includes: author, title, language, site url, country, domain rank, spam score, type
   * Dataset 2 (real) – allsides.com
     1. Articles are categorized by topic: environment, economics, etc. and by political lean (left, center, right)
     2. Does by web crawling / scraping to get thousands of articles
     3. Data set is comprised of 10558 total articles with their headlines and full body text and their labels (real vs fake). The data is located here in this [github repo](https://github.com/GeorgeMcIntire/fake_real_news_dataset).
   * Try other supervised machine learning algorithms such as SVM, Logistic regression and KNN
4. How it can be evaluated
5. How we are dividing responsibilities