# Initial Project Overview

## Overview of project content and milestones

The aim of the project is to create and evaluate a method of determining if the headline of a news article appropriately reflects its content.

The project will have several milestones. Firstly, a literature review will be undertaken that both analyses existing solutions, as well as exploring the problem space more generally. Then a prototype solution will be created, after which it will be tested and refined. Finally, the results and efficacy of the solution will be analysed and conclusions drawn.

## The main deliverable(s)

The most important deliverable of this project will be an algorithm that is able to determine to what percentage a headline matches the content of an article.

## The target audience for the deliverable(s)

?? people in the field of sentiment analysis, news agencies, people who want to determine the trustworthyness(?) of the info they're reading

## The work to be undertaken

- literature will need to be reviewed
- data will need to be either collected and labelled, or an existing dataset found
- the algorithm will need to be created
- the results of the algo will need to be tested and analysed

## Additional information/knowledge required

I'm currently unsure how to approach the algorithm design and implementation of this project. I have a very rudimentary understanding of neural networks and sentiment analysis, so research will need to be conducted into these areas, and a decision made whether to use one or the other, a combination of both or something else entirely.

## Information sources that provide a context for the project

- *Detecting Incongruity Between News Headline and Body Text via a Deep Hierarchical Encoder, 2018* Uses two neural networks with hierarchical structures to determine how incongruint headlines and bodies in articles are

- *The effects of subtle misinformation in news headlines, 2014* Investigates how misinformation can be delivered via news headlines

- *Media-generated Shortcuts: Do Newspaper Headlines Present Another Roadblock for Low-information Rationality?, 2007* A manual content analysis that shows "considerable difference" between articles and headlines

## The importance of the project

Since the idea of news was created, people in power have used a range of techniques to manipulate it to their advantage. As headlines are the most prominent aspect of any article, they are a prime target for alteration and miscommunication. By building a tool that can progromatically detect any incongruity, misleading and potentially manipulative headlines can be identified.

## The key challenge(s) to be overcome

The first challenge is that of creating or obtaining a quality dataset. In order to have validity, the project will need a large collection of news articles and their headlines, with each being labelled with their incongruence. While it would be fairly trivial to automate the collection of news articles, labelling them would have to be done manually, which would be a very subjective and time-consuming process.

Another challenge is my lack of experience around building classifiers and machine learning. While I expect to do a lot of learning around this area during the literature review, I anticipate that the development aspect of the project may take a while as I get to grips with some paradigms and techniques.