INTRO

* What is our task?
* What have people done on this – not solved just like what people think of nlp/nature of the fake vs real (the theory of fakeness)
* Why have we selected our data and what is its relevance to our problem? (how does it capture our problem) – should this be in the data selection section?
* Background/related work
  + How have people solved this problem in the past
  + Mention techniques commonly used for solving NLP problems (which we have also used) – maybe reference seminal papers here
    - Find references for SVM as gold standard for this task
    - Find reference for CNN and LSTM being the best for our task/NLP
  + Theory of embeddings and why it is necessary to transform textual data.
  + How do various embeddings differ? Probably don’t need to be this specific in the intro
  + Papers indicating relevant embeddings to your task
* Our intuition when solving our problem (reference previous work in addition to integrating context of your specific problem). what

Actual text:

In a world of COVID-19 and an increasingly polarised and unstable political climate, fake news has become a real and credible threat to citizen liberty and health as it continues to disseminate insidious disinformation within a public sphere that increasingly turns to online social-media for news sources \cite{nyi}\cite{rajan}\cite{shu}. Within politics specifically it is known to re-enforce confirmation bias of hyper-partisan views, encourage increased political polarisation, and undermines modern-day democracy \cite{lee}\cite{rose}.

The task of detecting fake news falls under the umbrella of natural language processing (NLP). NLP is a branch of machine learning which addresses the extraction of semantics and syntactic structure from human language. \textbf{this section needs references, will ask laura- also possible include "intuition when solving our problem", not quite sure what this means}. Embeddings were utilised as part of NLP; as they facilitate mapping of text to low-dimensional, learned continuous vector representations that capture meaningful relationships within language for classification \cite{koe}\cite{ruiz}. Those utilised were chosen based on their appropriateness for the task, and referenced success in related studies, this included TF-IDF \cite{smith}, GloVe.

Fake news, as a relevant and challenging topic that could benefit from analysis, has a myriad of academic studies on methods of detecting fake news utilising machine learning \cite{asr}\cite{con}\cite{shu}\cite{smith}. In our report, previous studies and results will be taken into consideration to compare statistical machine learning with deep learning techniques. Methods used were those that have been proven to be successful in previous studies; thus for statistical models, a Na\"{i}ve Bayes classifier was used as a baseline \cite{con} and Support Vector Machine (SVM) as the best performing statistical machine learning model for fake news detection \cite{con} \cite{smith}. Furthermore, two variants of deep learning were evaluated for this task: Convolutional Neural Networks (CNN), as they are cited as being appropriate for longer text classification \cite{asr}, and Recurrent Neural Networks (RNN), as they have been found to be especially effective in sequential text classification \cite{elvis}\cite{shu}\cite{smith}. In order to accurately judge each model and their relevance within the context of society, accuracy, precision, recall, and the generalisability of each model have been considered as evaluation metrics.