# **Appendix**

## SLIDE 5

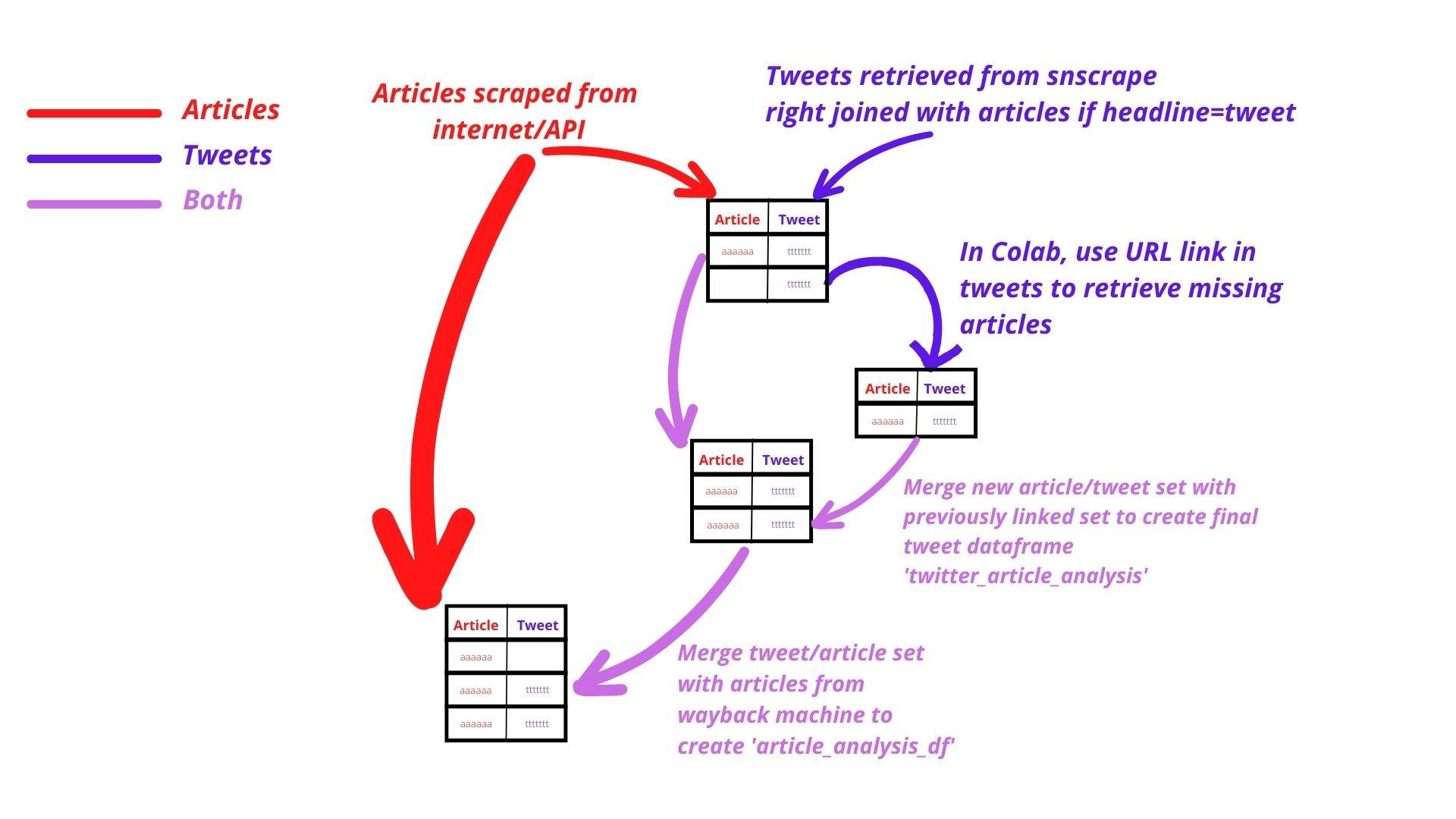
Graphic shows a BertViz visualisation of the words of a sequence as encoded by one head in one layer of the transformer model. Q.K gives us the attention for any word in the sequence against any other word in the sequence, focussing on one group of linguistic features.

## SLIDE 6

Free-to-use python library [snscrape](https://github.com/JustAnotherArchivist/snscrape), was used to scrape the tweets. It returns the tweet text, URL links within the tweet and metrics such as likes and retweets.

For those able to afford $1500, the articles can all be retrieved from the [News Api](https://newsapi.org/) module. This example instead used the Internet [Wayback Machine](https://archive.org/web/) to retrieve the links to articles hosted on the outlet's homepage, going back to 2013.

These URLs were then used to scrape the actual article and retrieve the text, headline, category, and date. Guardian articles were scraped using its free api.

The article and twitter datasets were then linked as below…

## SLIDE 7

## The graphic shows the fundamental idea behind transformer models. The original paper can be found here <https://arxiv.org/pdf/1706.03762.pdf>

## SLIDE 9

The dataframes of 10,000 predicted headlines, and 10,000 predicted tweets can be found on GitHub <https://github.com/jimp93/Capstone-API> under the filenames *headline\_pred\_df* and *tweet\_pred\_df*

## SLIDE 10

Graphic shows the relationship between the predicted word ‘members’ in the headline and words from the article text using the Bertviz module.

This relationship shows the ‘attention’ paid by one ‘attention head’ in one level of the decoder to various attention heads in the same level in the encoder (as shown by the coloured bars).

Each attention head captures a different semantic or syntactic quality, which the model learns for itself (ie, we don’t tell the model any semantic or grammar rules).

The intensity of the coloured bars shows the strength of the linguistic relationships – represented by the bar – between the words

## SLIDE 11

Rouge scores are a way of measuring the closeness of two text sequences. The different metrics focus on the appearance of common word blocks in both sequences, with each metric being a different length word block.

## SLIDE 13

The importance of each word in the prediction, and resulting visualisation, is calculated using SHAP values. Here is an explanation of how they work…

<https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30>

## SLIDE 15

The keywords in predicting the number of retweets and the visualisation are generated using LIME values. Here is an explanation of how they work…

<https://towardsdatascience.com/lime-explain-machine-learning-predictions-af8f18189bfe>

## SLIDE 16

The displayed word vectors are calculated using PCA, which collapses the multidimensional word vectors into the 3 most important planes so we can visualise using tensorflow’s tensorboard.

The word vectors at the moment are generated by the retweet model, which is problematic as tweets are quite short, so not as many instances of the words to train on, and also because the target of the model is to predict retweets, so the words are not only clustered by their semantic similarities but by whether they appear in retweets with similar virality.

This introduces an ‘impurity’ if we are trying to understand the vocabulary purely in terms of its usage, but does open up interesting potential ways to visualise use of vocabulary in certain specialised environments.

## SLIDE 20

The summarizer models are being improved by training different models for each different type of article, so it can learn the linguistic relationships found in news pieces separately from those found in opinion pieces and other non-news pieces. Instead of having one ‘Jack of all trades’ model, the specialised models should produce better summaries.

The summarizer models are pre-trained on a corpus. We could do with diversifying this corpus to ensure inbuilt biases don’t become permanently entrenched.

Transformer models are being developed to replace the current retweet predictor and opinion-score generator. They promise to deliver more accurate predictions and deeper analytical insights.

The retweet model will later on incorporate image data drawn from the photograph attached to the tweet, as that is likely to be a large driver of virality.

Named entity recognition techniques will be incorporated into the opinion-news scorer as we will be able to isolate certain people, organisations, places and analyse how they affect the style of reporting. We will offer a separate sentiment analysis tool so that we can instantly check the negative/positive reporting associated with these entities for each outlet.

The incorporation of Part of Speech tagging into the opinion-score model should also improve performance and deepen analytical insight.

Other models in development include a headline congruence tester and summarizers trained on specialist publications, so journalists can immediately extract the important points out of technical reports from medical and scientific publications, financial institutions etc. The lesson of Covid, where journalists were asked to quickly produce copy on vaccine efficacy studies etc, highlights how valuable this tool would be

# **Sources**

1. <https://news.gallup.com/poll/394283/confidence-institutions-down-average-new-low.aspx>
2. <https://github.com/jessevig/bertviz>
3. <https://doi.org/10.48550/arXiv.1706.03762>
4. <https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30>
5. <https://towardsdatascience.com/lime-explain-machine-learning-predictions-af8f18189bfe>