**Introduction:**

According to The [Atlantic](https://www.theatlantic.com/technology/archive/2016/05/how-many-stories-do-newspapers-publish-per-day/483845/), The New York Times publishes more than 150 articles a day and more than 250 on Sundays. The Wall Street Journal publishes about 240 stories every day. Other websites, like Buzzfeed, publish more than 6,000 stories every month.With the amount of information made available to us, it is no surprise that summarization has become a very popular task in machine learning/Natural Language Processing (NLP).

**Overview:**

Text summarization is the task of automatically generating a shorter version of a document while retaining its most important information. The task has received much attention in the natural language processing community.  Since it has immense potential for various information access applications. Examples include tools which digest textual content (e.g., news, social media, reviews), answer questions, or provide recommendations. The summarization model could be of two types:

* Extractive Summarization — Is akin to using a highlighter.  We select sub segments of text from the original text that would create a good summary
* Abstractive Summarization — Is akin to writing with a pen. Summary is created to extract the gist and could use words not in the original text. This is harder for machines to do

**Metrics:**

The performance of a text summarization system is measured by its ROUGE score. ROUGE score measures the overlap between predicted and ground truth summary.

**BERT:**

BERT is a powerful model that has proven effective on a variety of NLP tasks. BERT’s key technical innovation is applying the bidirectional training of Transformer, a popular attention model, to language modelling. Its success shows that a language model which is bidirectionally trained can have a deeper sense of language context and flow than single-direction language models. Here is an excellent link to learn more about [BERT](https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270).

BERT can also be used for next sentence prediction. The model receives pairs of sentences as input and learns to predict if the second sentence in the pair is the subsequent sentence in the original document. During training, 50% of the inputs are a pair in which the second sentence is the subsequent sentence in the original document. While in the other 50% a random sentence from the corpus is chosen as the second sentence.

**Facebook/bart-large-cnn:**

BART is a transformer encoder-encoder (seq2seq) model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder. BART is pre-trained by (1) corrupting text with an arbitrary noising function, and (2) learning a model to reconstruct the original text.

BART is particularly effective when fine-tuned for text generation (e.g. summarization, translation) but also works well for comprehension tasks (e.g. text classification, question answering