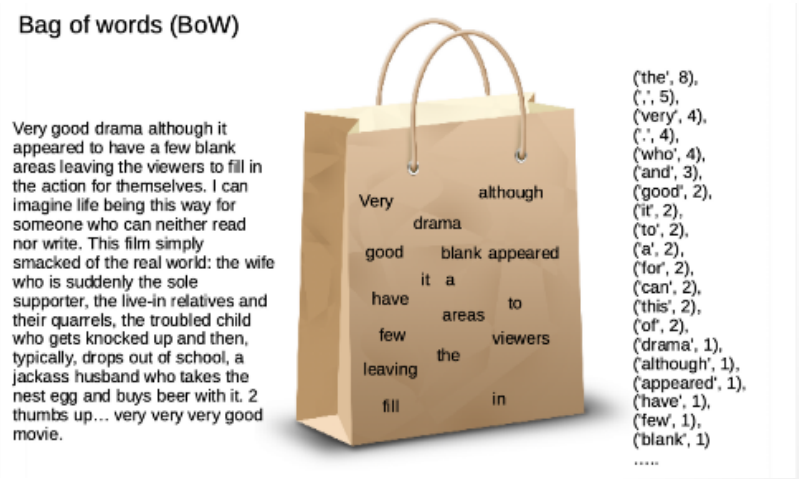
BAG-OF-WORDS APPROACHES

- A burgeoning literature in finance and accounting that use textual analysis(reviews by Li 2010, Loughran & McDonald 2016; Bockhay et al. 2022)

- The majority of these studies rely on NLP algorithms that assume a bag-of-words structure and use one-hot encoding, which do not consider **word contexts**
  - Dictionary approaches
  - The naïve Bayes classifications
  - Topic modelling techniques
  - Support Vector Machine &
    Random Forest
  - Measures of textual features such as:
    - Readability; salience or concreteness
    - Similarity



(LARGE) LANGUAGE MODELS (LLM)

- A language model is a probability distribution over sequences of words
  - Language understanding, translation, generation; question answering

- Huge amount of textual data, substantial computing resources & time to train and apply
- Intuitive algorithms; notoriously opaque
- Deep neural (Transformer) architecture

A TWO-STEP PROCESS TO USE LLM: AN ILLUSTRATION

- Pre-training: predict masked word (15%) & next sentence (50%)
- Fine-tuning: further train pre-trained model for downstream tasks (e.g., classification/labeling)

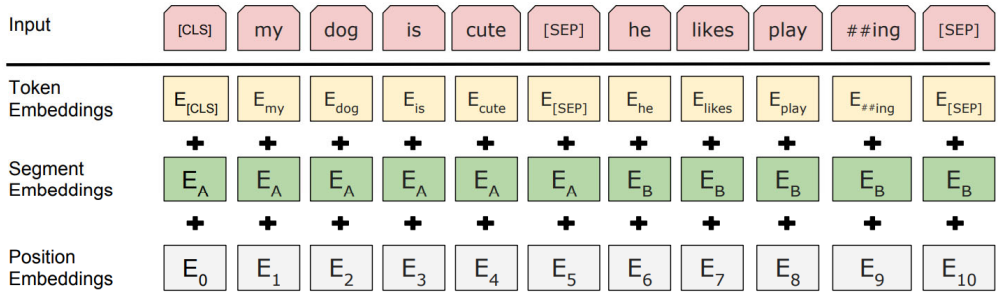
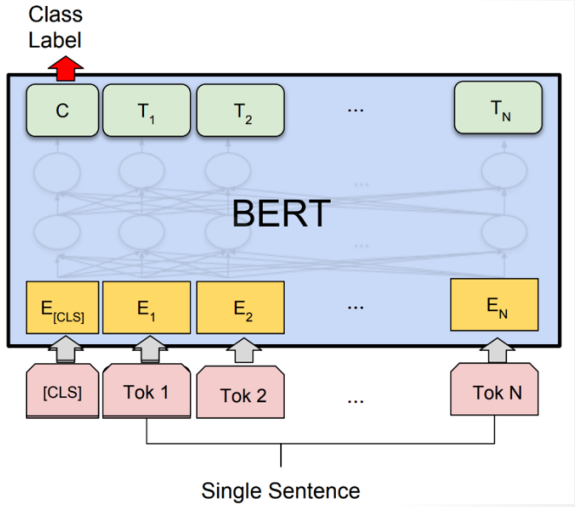
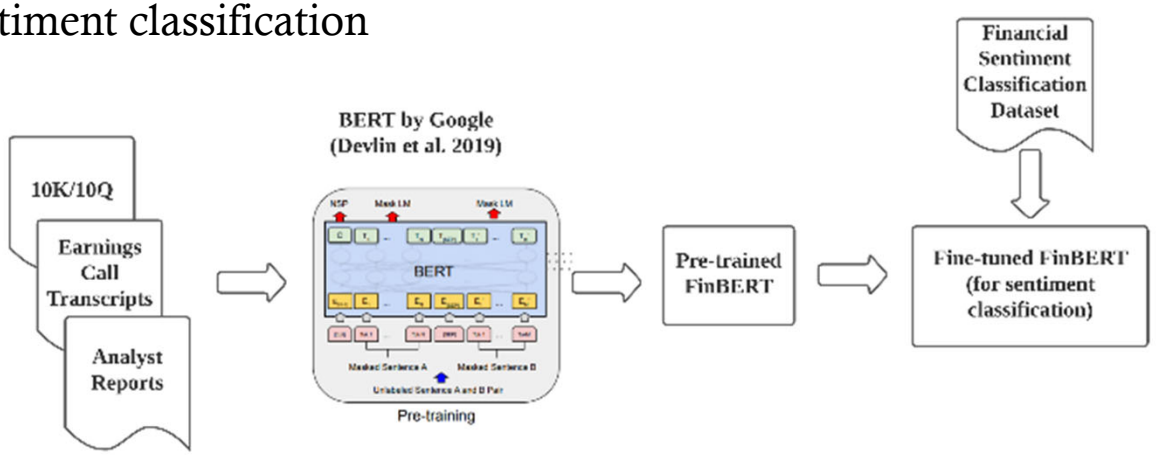


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

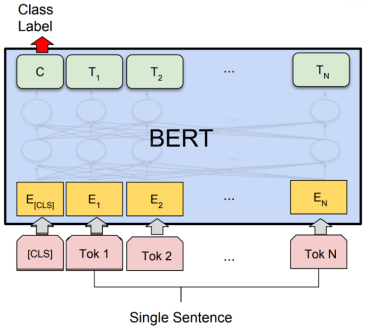


FINBERT – A LARGE LANGUAGE MODEL FOR EXTRACTING INFORMATION FROM FINANCIAL TEXT

- For Sentiment classification

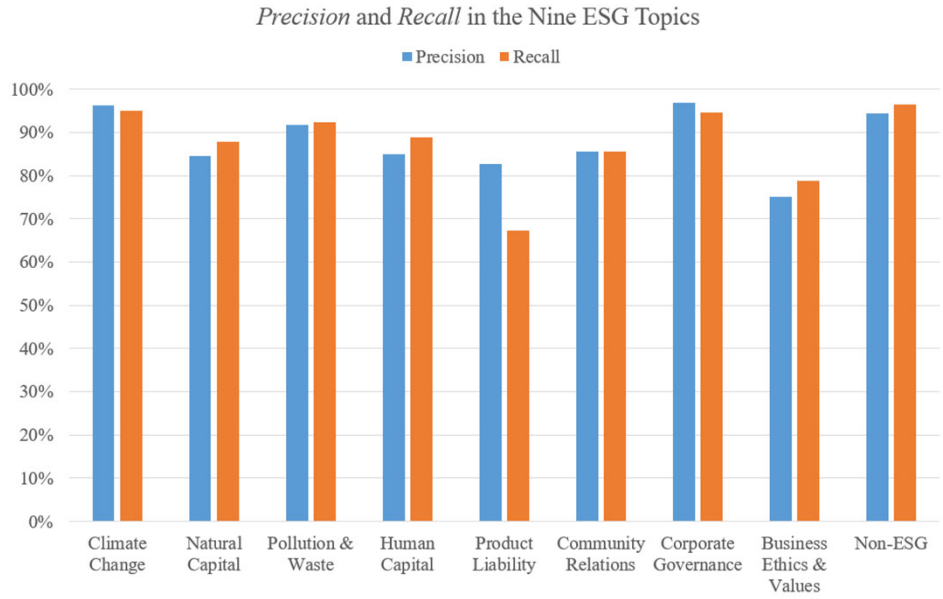


- Also fine-tuned for ESG classification; all models including the pre-trained model and various fine-tuned models available at https://www.allenhuang.org/coding.html
- What can you do with FinBERT?
  1. Fine-tune the pre-trained model for other tasks (need labels)
  2. Directly use a fine-tuned model



FINBERT DEMO (ESG CLASSIFICATION)

- Fine-tune sample is 16,857 sentences from 55 S&P firms in 11 GICS sectors



- Demo: A Google Colab code to demo FinBERT fine-tuned for ESG topics
  - Link available at https://www.allenhuang.org/coding.html
  - Or direct link: https://tinyurl.com/finbertdemo