NLP/ASR

Unit 5: Advanced Topics in ASR/NLP

5.4.1

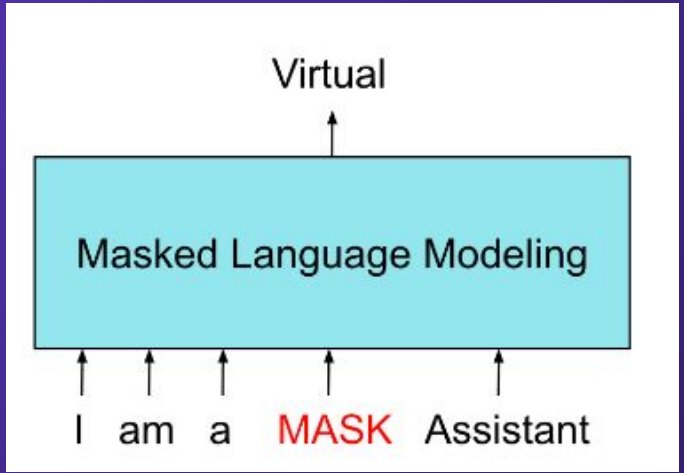
State-of-the-art Models

Advanced transformer models

TIL-AI
TODAY I LEARNED AI

Masked Language Model (MLM)

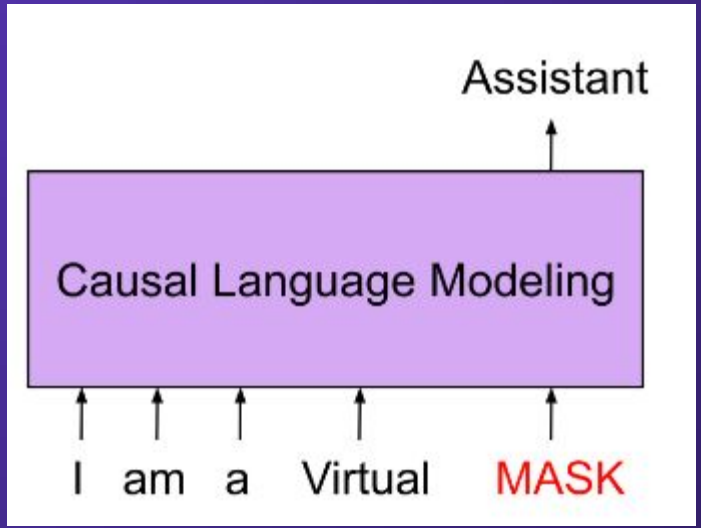
- During training, MLM randomly replaces a percentage of words in a text with special "MASK" tokens
- The model's objective is to predict the original masked words
- It learns by analyzing the surrounding context from both directions (before and after the masked word)
- Applications: NER, Sentiment Analysis
- Example: BERT



TIL-AI
TODAY I LEARNED-AI

Causal Language Model (CLM)

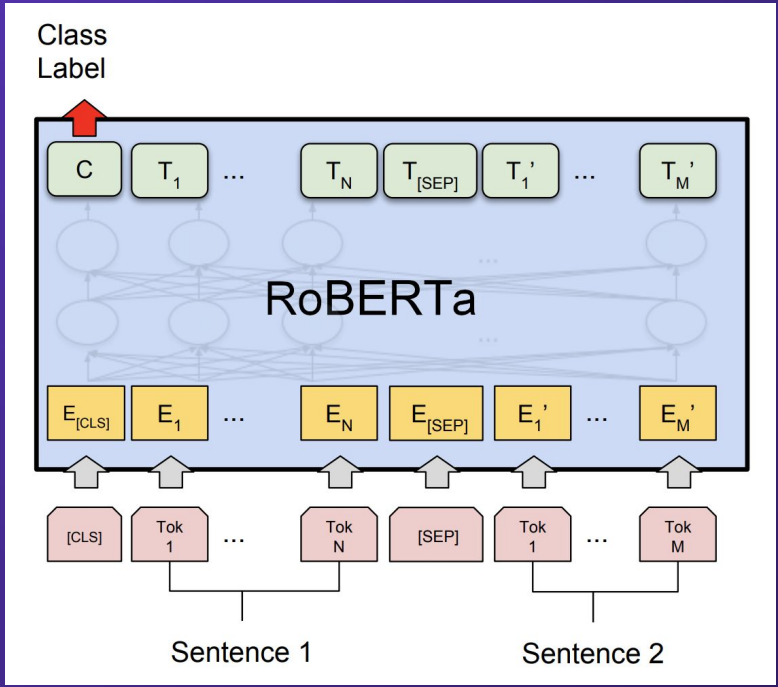
- CLMs are trained to predict the next word in a sequence
- Model can only access words seen previously (to the left)
- Mimics a left-to-right, sequential generation process
- Application: Text generation
- Example: GPT



TIL-AI
TODAY I LEARNED AI

RoBERTa

- RoBERTa uses significantly more data and trains for longer periods than BERT
- The masking pattern is changed in every training epoch, ensuring the model doesn't overfit to specific patterns
- By maximizing the model’s exposure to diverse linguistic patterns, RoBERTa achieves higher levels of robustness and performance



(https://huggingface.co/docs/transformers/en/model_doc/roberta)

Benefits of RoBERTa

- RoBERTa surpasses original BERT performance on key benchmarks including GLUE, RACE, and SQuAD.
- Although employing more data, RoBERTa maintains comparable computational requirements to BERT

ELECTRA

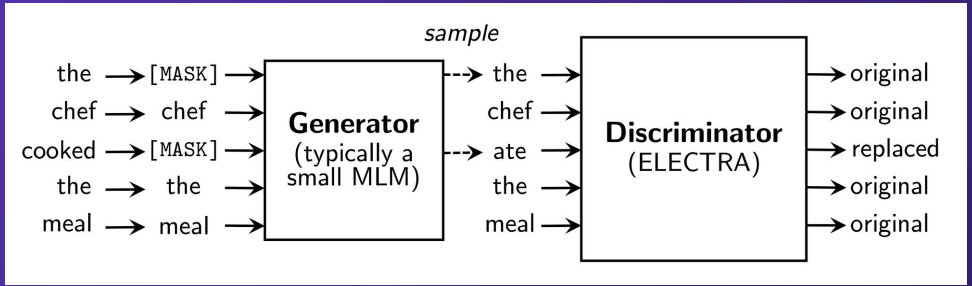
- ELECTRA stands out by using a sample-efficient pre-training method called replaced token detection. Unlike other models that predict masked words, ELECTRA discriminates whether a token is replaced by a generator model, enhancing training efficiency

- Two Transformer Models:
  - Generator: A smaller masked language model tasked with replacing tokens in the input sequence.
  - Discriminator: The main ELECTRA model, trained to detect if a token has been replaced by the generator

- Adversarial Training: Similar to GANs (Generative Adversarial Networks), the generator tries to deceive the discriminator, leading to both models improving over time

Benefits of ELECTRA

- ELECTRA achieves competitive or better results than BERT while using fewer parameters and less training data
- The replaced token detection task forces ELECTRA to focus on nuanced contextual understanding, often resulting in better disambiguation



- ELECTRA avoids some potential issues of Masked Language Modeling, such as the model only learning from a fraction of the input sequence during pre-training

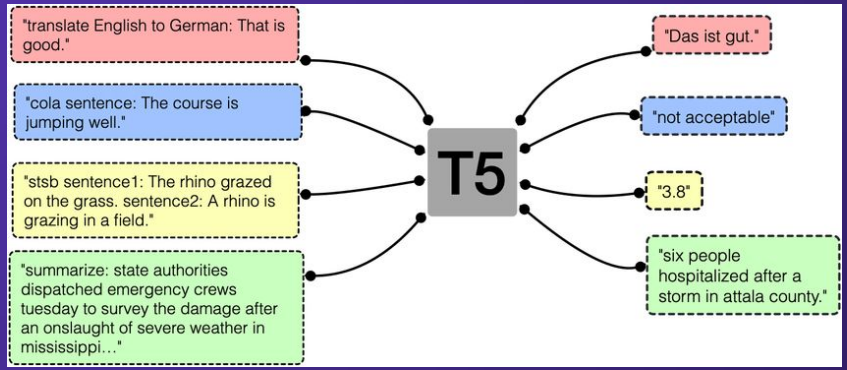
(https://huggingface.co/docs/transformers/en/model_doc/electra)

TIL-AI
TODAY I LEARNED-AI

T5 (Text-to-Text Transfer Transformer)

- T5 adopts a unified “text-to-text” framework. It formulates all NLP tasks as a text-to-text problem, where both the input and output are treated as text strings

- Encoder-Decoder Architecture: T5 employs the classic transformer encoder-decoder structure for both input understanding and output generation



(https://huggingface.co/docs/transformers/en/model_doc/t5)

TIL-AI
TODAY I LEARNED AI

Benefits of T5

- T5 seamlessly performs various NLP tasks like translation, question answering, summarization, and different classification tasks
- A single model architecture and its pre-training data enable it to be fine-tuned for diverse tasks
- Text-to-text removes complexities from task-specific models, simplifying the NLP landscape