

## Question Answering

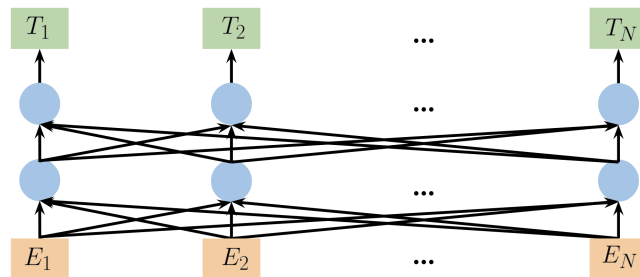
- ✓ **Video:** Week 3 Overview  
6 min
- ✓ **Reading:** Week 3 Overview  
10 min
- ✓ **Video:** Transfer Learning in NLP  
7 min
- ✓ **Reading:** Transfer Learning in NLP  
10 min
- ✓ **Video:** ELMo, GPT, BERT, T5  
7 min
- ✓ **Reading:** ELMo, GPT, BERT, T5  
10 min
- ✓ **Video:** Bidirectional Encoder Representations from Transformers (BERT)  
4 min
- ✓ **Reading:** Bidirectional Encoder Representations from Transformers (BERT)  
10 min
- ▶ **Video:** BERT Objective  
2 min
- 📖 **Reading:** BERT Objective  
10 min
- ▶ **Video:** Fine tuning BERT  
2 min
- 📖 **Reading:** Fine tuning BERT  
10 min
- ▶ **Video:** Transformer: T5  
3 min
- 📖 **Reading:** Transformer T5  
10 min
- ▶ **Video:** Multi-Task Training Strategy  
5 min
- 📖 **Reading:** Multi-Task Training Strategy  
10 min
- ▶ **Video:** GLUE Benchmark  
2 min
- 📖 **Reading:** GLUE Benchmark  
10 min
- ▶ **Video:** Question Answering  
2 min
- 📖 **Reading:** Question Answering  
10 min
- 📄 **Lab:** SentencePiece and BPE  
2h
- 📖 **Reading:** Content Resource  
10 min

## Assignment

# Bidirectional Encoder Representations from Transformers (BERT)

You will now learn about the BERT architecture and understand how the pre-training works.

- Makes use of transfer learning/pre-training:



There are two steps in the BERT framework: pre-training and fine-tuning. During pre-training, the model is trained on unlabeled data over different pre-training tasks. For fine tuning, the BERT model is first initialized with the pre-trained parameters, and all of the parameters are fine-tuned using labeled data from the downstream tasks. For example, in the figure above, you get the corresponding embeddings for the input words, you run it through a few transformer blocks, and then you make the prediction at each time point  $T_i$ .

- Choose 15% of the tokens at random: mask them 80% of the time, replace them with a random token 10% of the time, or keep as is 10% of the time.
- There could be multiple masked spans in a sentence
- Next sentence prediction is also used when pre-training.

The next video will talk about the BERT objective.

✓ Complete

Go to next item

