1. Why does BERT succeed?

Compared with the original RNN and LSTM, BERT can perform concurrent execution, while extracting the relationship features of words in the sentence, and can extract the relationship features at multiple different levels, and then more fully reflect the semantics of the sentence.

1. Can we leverage BERT in clinical natural language processing?

No, because BERT need a huge amount of data and computation power to train, and we do not have enough clinical natural language data.