**Supervised learning**

Paper title: The NLP Cookbook: Modern Recipes for Transformer based Deep Learning Architectures

How about transfer learning for our project? Can we use an existing model like GPT-II or BERT?

Or less expensive models like Knowledge distillation, MobileBERT, TinyBERT, etc.?

Process longer sequences: Transformer-XL, Longformer, BigBird?

Various models and their descriptions provided at the end.

**Semi-supervised**

Another paper for semi-supervised learning

DOI: 10.1109/ICACCI.2018.8554593