

How to train BERT with an Academic Budget

<https://arxiv.org/abs/2104.07705v2>

(used for training pipeline and architectural design guidelines; accepted at EMNLP)

XLNet: Generalized Autoregressive Pretraining for Language Understanding

<https://arxiv.org/abs/1906.08237v2>

(uses Accuracy as metric; accepted at NeurIPS)

Long Short Term Memory (LSTM)

<http://www.bioinf.jku.at/publications/older/2604.pdf>

Bidirectional Encoder Representations for Transformers (BERT)

<https://arxiv.org/abs/1810.04805>

Centralised Kernel Alignment (CKA)

<https://arxiv.org/abs/1905.00414>

<https://github.com/yuanli2333/CKA-Centered-Kernel-Alignment>

Attention Is All You Need (Transformers)

<https://arxiv.org/abs/1706.03762>

CLIP

<https://arxiv.org/abs/2103.00020>

DALLE

<https://arxiv.org/abs/2102.12092>

SentenceBERT

<https://arxiv.org/abs/1908.10084>

<https://www.sbert.net>