BERT: Pre-training of Deep Bidirectional Transformers forLanguage Understanding

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

### [Swag: A large-scale adversarial dataset for grounded commonsense inference](https://arxiv.org/abs/1808.05326)

<https://arxiv.org/pdf/1808.05326.pdf>

COMMONSENSEQA: A Question Answering Challenge TargetingCommonsense Knowledg

<https://arxiv.org/pdf/1811.00937.pdf>

Commonsense Reasoning for Natural Language Understanding:A Survey of Benchmarks, Resources, and Approaches

<https://arxiv.org/pdf/1904.01172.pdf>

Explain Yourself!Leveraging Language Models for Commonsense Reasoning

<https://arxiv.org/pdf/1906.02361.pdf>