BERT Model Architecture for Legal Clause Extraction

Token Embeddings

Position Embeddings

Segment Embeddings

+

768 dimensions

Multi-Head Self-Attention 12	Feed-Forward Network 12
Multi-Head Self-Attention 11	Feed-Forward Network 11
Multi-Head Self-Attention 10	Feed-Forward Network 10
Multi-Head Self-Attention 9	Feed-Forward Network 9
Multi-Head Self-Attention 8	Feed-Forward Network 8
Multi-Head Self-Attention 7	Feed-Forward Network 7

Classification Head
(Clause Types)

Clause Predictions
& Probabilities