Fine-tuning: Task-Specific Heads on BERT

**Pre-trained BERT
(shared)**

Classification

[CLS] → Linear → Softmax

**Token Classification
(NER)**

Each token → Linear → Labels

Question Answering

Span prediction
(start/end logits)