

Pre-trained model

Quantization

Training data

Retraining/Finetuning

Quantized model