

ORIGINAL RoBERTa training (400 examples): cross-entropy loss

