BERT (large) GLUE CoLA on 1 GPU

Validation accuracy vs Step

- 4bs_1gpu_1vn (baseline)
- 32bs_1gpu_8vn