--gpus 4

--train\_batch\_size 4

--eval\_batch\_size

--max\_source\_length 128

--max\_target\_length 25

--val\_max\_target\_length 25

--test\_max\_target\_length 25

--label\_smoothing 0.1

--dropout 0.1

--attention\_dropout 0.1

--weight\_decay 0.001

--adam\_epsilon 1e-08

--max\_grad\_norm 0.1

--lr\_scheduler polynomial

--learning\_rate 3e-05

--num\_train\_epochs 10

--warmup\_steps 500

--gradient\_accumulation\_steps 4

--distributed\_retriever ray

--num\_retrieval\_workers 4