

$\Delta\text{Acc}(\text{FP8} - \text{FP32})$

CIFAR-10

PTQ FP8-E4M3

ScrewSet

GPTQ FP8-E4M3

ImageNet

