

$d_w = 7000$ (Bert-tiny), 8500 (Bert-mini), 8000 (Bert-small), 4000 (Bert-medium), $d_s = 700$ (Electra-base-discriminator)
 $N = 10000$ $n = 2500$

