Loss
Tokens seen
original_paper:overtrain
overtrain c4_original-d=1024_l=24_h=8-0.25
overtrain c4_original-d=1024_l=24_h=8-0.5
overtrain c4_original-d=1024_l=24_h=8-1.0
overtrain c4_original-d=1024_l=24_h=8-16.0
overtrain c4_original-d=1024_l=24_h=8-2.0
overtrain c4_original-d=1024_l=24_h=8-4.0
overtrain c4_original-d=1024_l=24_h=8-8.0
overtrain c4_original-d=512_l=8_h=4-0.25
overtrain c4_original-d=512_l=8_h=4-0.5
overtrain c4_original-d=512_l=8_h=4-1.0
overtrain c4_original-d=512_l=8_h=4-16.0
overtrain c4_original-d=512_l=8_h=4-2.0
overtrain c4_original-d=512_l=8_h=4-32.0
overtrain c4_original-d=512_l=8_h=4-4.0
overtrain c4_original-d=512_l=8_h=4-8.0
overtrain c4_original-d=576_l=24_h=8-0.25
overtrain c4_original-d=576_l=24_h=8-0.5
overtrain c4_original-d=576_l=24_h=8-1.0
overtrain c4_original-d=576_l=24_h=8-16.0
overtrain c4_original-d=576_l=24_h=8-2.0
overtrain c4_original-d=576_l=24_h=8-32.0
overtrain c4_original-d=576_l=24_h=8-4.0
overtrain c4_original-d=576_l=24_h=8-8.0
overtrain c4_original-d=96_l=8_h=4-0.5
overtrain c4_original-d=96_l=8_h=4-1.0
overtrain c4_original-d=96_l=8_h=4-16.0
overtrain c4_original-d=96_l=8_h=4-2.0
overtrain c4_original-d=96_l=8_h=4-32.0
overtrain c4_original-d=96_l=8_h=4-4.0
overtrain c4_original-d=96_l=8_h=4-8.0
overtrain c4_original-open_lm_1b-1.0
overtrain rpj-d=1024_l=24_h=8-0.25
overtrain rpj-d=1024_l=24_h=8-0.5
overtrain rpj-d=1024_l=24_h=8-1.0
overtrain rpj-d=1024_l=24_h=8-16.0
overtrain rpj-d=1024_l=24_h=8-2.0
overtrain rpj-d=1024_l=24_h=8-32.0
overtrain rpj-d=1024_l=24_h=8-4.0
overtrain rpj-d=1024_l=24_h=8-8.0
overtrain rpj-d=512_l=8_h=4-0.25
overtrain rpj-d=512_l=8_h=4-0.5
overtrain rpj-d=512_l=8_h=4-1.0
overtrain rpj-d=512_l=8_h=4-16.0
overtrain rpj-d=512_l=8_h=4-2.0
overtrain rpj-d=512_l=8_h=4-32.0
overtrain rpj-d=512_l=8_h=4-4.0
overtrain rpj-d=512_l=8_h=4-8.0
overtrain rpj-d=576_l=24_h=8-0.25
overtrain rpj-d=576_l=24_h=8-0.5
overtrain rpj-d=576_l=24_h=8-1.0
overtrain rpj-d=576_l=24_h=8-16.0
overtrain rpj-d=576_l=24_h=8-2.0
overtrain rpj-d=576_l=24_h=8-32.0
overtrain rpj-d=576_l=24_h=8-4.0
overtrain rpj-d=576_l=24_h=8-8.0
overtrain rpj-d=96_l=8_h=4-0.5
overtrain rpj-d=96_l=8_h=4-1.0
overtrain rpj-d=96_l=8_h=4-16.0
overtrain rpj-d=96_l=8_h=4-2.0
overtrain rpj-d=96_l=8_h=4-32.0
overtrain rpj-d=96_l=8_h=4-4.0
overtrain rpj-d=96_l=8_h=4-8.0
overtrain rpj-open_lm_1b-1.0
overtrain rw_original-d=1024_l=24_h=8-0.25
overtrain rw_original-d=1024_l=24_h=8-0.5
overtrain rw_original-d=1024_l=24_h=8-1.0
overtrain rw_original-d=1024_l=24_h=8-16.0
overtrain rw_original-d=1024_l=24_h=8-2.0
overtrain rw_original-d=1024_l=24_h=8-32.0
overtrain rw_original-d=1024_l=24_h=8-4.0
overtrain rw_original-d=1024_l=24_h=8-8.0
overtrain rw_original-d=512_l=8_h=4-0.25
overtrain rw_original-d=512_l=8_h=4-0.5
overtrain rw_original-d=512_l=8_h=4-1.0
overtrain rw_original-d=512_l=8_h=4-16.0
overtrain rw_original-d=512_l=8_h=4-2.0
overtrain rw_original-d=512_l=8_h=4-32.0
overtrain rw_original-d=512_l=8_h=4-4.0
overtrain rw_original-d=512_l=8_h=4-8.0
overtrain rw_original-d=576_l=24_h=8-0.25
overtrain rw_original-d=576_l=24_h=8-0.5
overtrain rw_original-d=576_l=24_h=8-1.0
overtrain rw_original-d=576_l=24_h=8-16.0
overtrain rw_original-d=576_l=24_h=8-2.0
overtrain rw_original-d=576_l=24_h=8-32.0
overtrain rw_original-d=576_l=24_h=8-4.0
overtrain rw_original-d=576_l=24_h=8-8.0
overtrain rw_original-d=96_l=8_h=4-0.5
overtrain rw_original-d=96_l=8_h=4-1.0
overtrain rw_original-d=96_l=8_h=4-16.0
overtrain rw_original-d=96_l=8_h=4-2.0
overtrain rw_original-d=96_l=8_h=4-32.0
overtrain rw_original-d=96_l=8_h=4-4.0
overtrain rw_original-d=96_l=8_h=4-8.0
overtrain rw_original-open_lm_1b-1.0