IO:S1 Attn Ratio Across Checkpoints (pythia-160m-alldropout)

Metric Value vs Training Checkpoint

Attention Head
- Layer 7-Head 6
- Layer 7-Head 9
- Layer 7-Head 10
- Layer 8-Head 1
- Layer 8-Head 2
- Layer 8-Head 5
- Layer 8-Head 8
- Layer 8-Head 9