DS-1.5B: Convergence (Avg accuracy vs Steps)

Target (AdamW best @0–200): 0.364
AdamW: 180 | ISO: 40
Saved: 140 (77.8% fewer)

Training efficiency (77.8% fewer steps)

AdamW
ISO (ours)