

# Dana-Star Optimal Learning Rate Scaling Law ( $\omega = 4.0$ , Top-K = 5, Dataset = fineweb\_100)

