

Unlearning Methods Comparison

36000 steps

72000 steps

108000 steps

143000 steps

Pythia 6.9B

Percent Memorized (%)

Pythia 2.8B

Perplexity

Perplexity

Perplexity

Perplexity

act

slim

durable_agg

random_greedy

hc

durable

random