Ratio of hallucinated tokens over the course of training (lin-log) 0.96 0.95 0.94 0.93 0.92 0.91 0.90 0.89 100 1000 10000 100000

steps

Ratio