

(temperature, top-p) = (1.0, 0.2)

Method

Original

GradDiff

SimNPO

NPO+ENT

NPO

Retrain

BLUR-NPO

Number of Generations  $k$

1 2 4 8 16 32 64 128

Best Entailment Score

1.0

0.8

0.6

0.4

0.2

0.0

