Entailment Score Increment at (temperature, top-p) = (1.0, 1.0)**BLUR-NPO** 0.7 NPO NPO+ENT NPO+ENT+TMP **SimNPO** 0.6  ${\sf GradDiff}$ Increment of Best Entailment Score 0.5 Worst Leak@k Performance 0.4 0.3 Best Leak@k Performance 0.2 0.1 0.0 -0.116 32 64 4 128 Number of Generations (k)