Entailment Score Increment at (temperature, top-p) = (0.8, 0.2)**BLUR-NPO** 0.7 NPO NPO+ENT NPO+ENT+TMP **SimNPO** 0.6 GradDiff Increment of Best Entailment Score 0.5 0.4 0.3 0.2 0.1 Worst Leak@k Performance 0.0 Best Leak@k Performance -0.116 32 4 64 128 Number of Generations (k)