**(temperature, top-p) = (0.8, 0.8)**

*Method* (y-axis): Original, RMU, GradDiff, SimNPO, NPO+ENT, NPO, Retrain, BLUR-NPO, LoUK

*Number of Generations $k$* (x-axis): 1, 2, 4, 8, 16, 32, 64, 128

Colorbar: *leak@$k$-ES* ranging from 0.0 to 1.0