

# Generalization Gap (Testset vs. Best Validation Score) on PUPA for GPT-4.1

