

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

