HoVer, Qwen3 8B **Optimization Method 52.5 Baseline** MIPROv2 **50.0 GRPO GEPA** 47.5 **GEPA+Merge Validation Performance** 9 45.0 9 42.5 **Test-set Performance** 40.0 **37.5** 35.0 **5000** 20000 10000 **15000** 25000

Number of Rollouts