IFBench, Qwen3 8B **Optimization Method** 38 **Baseline** 80 MIPROv2 \star **GRPO 36 GEPA 75 GEPA+Merge Validation Performance** 34 Ň **Test-set Performance** \star **32 65 30 60** 28 **5000** 20000 10000 **15000** 25000 0 **Number of Rollouts**