



- Multi-agent framework (Ours)**
- Ensemble single-agent system with fine-tuning**
- Single-agent system with fine-tuning**
- Single GPT-4o without fine-tuning**