

Agentic RL Training

Infra for agentic RL training

K8s + Docker to deploy train, rollout, and env with disaggregated resources.

train: use ray to allocate gpus for 5D parallelism

- receive trajectory (reasoning token + tool response, reward)
- sync model weights with rollout engine
- engine: megatron-lm / pytorch fsdp

rollout: use vllm / sglang router for generation

- receive tool output and reward from env
- send tool call request to env

env: use api / mcp as abstraction

- candidate tools: subagent, terminal (bash), knowledge graph

