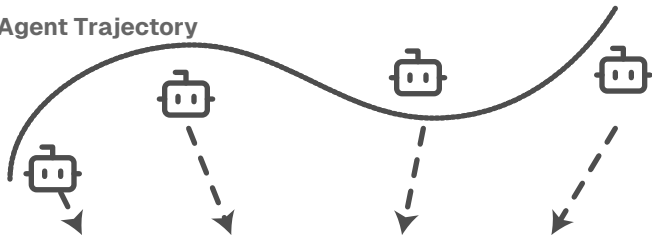


Agent Trajectory



**Goal-Oriented
Evaluation**



**Manually Designed
Behavior Evaluation**

Ours

Meta-Evaluation



**AutoLibra
Evaluation**



Extraction



**Open-Ended
Feedback**