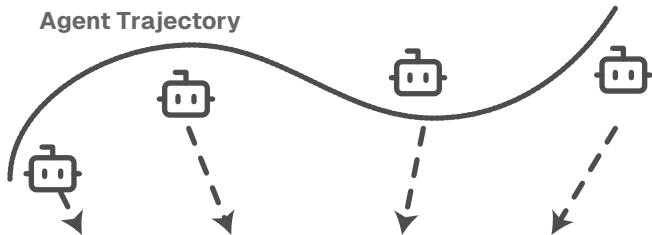


Agent Trajectory



**Goal-Oriented
Evaluation**



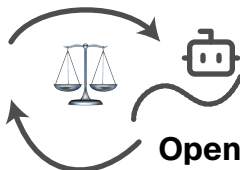
**Manually Designed
Behavior Evaluation**

Ours



**AutoLibra
Metric Induction**

Evaluation



Extraction



**Open-Ended
Feedback**