
One Tool Is Enough: Reinforcement Learning of LLM Agents for Repository-Level Code Navigation

Zhaoxi Zhang^{1*} Yitong Duan² Yanzhi Zhang² Yiming Xu¹ Zhixiang Wang¹ Kun Liang¹ Yang Li³
Jiahui Liang³ Deguo Xia³ Jizhou Huang³ Jiyan He² Yunfang Wu¹



北京大學
PEKING UNIVERSITY



北京中关村学院
Zhongguancun Academy



1. Motivation

- LLM agents have proved strong ability on real-world software engineering (SWE) tasks, but most methods are training-free or distillation-based. Training an agent without a stronger teacher model requires attention.
- Reinforcement learning has emerged as a powerful method for agentic training. However previous studies have shown that RL can not be applied without distillation.
- We change the tool set of the coding agent, and demonstrate that this design is efficient of RL agentic training.

Tool Design

- We propose RepoNavigator, which has only one tool.
- Jump: jumping to the definition of a given symbol. This operation is normal in modern coding IDEs (VScode, Pycharm, etc)

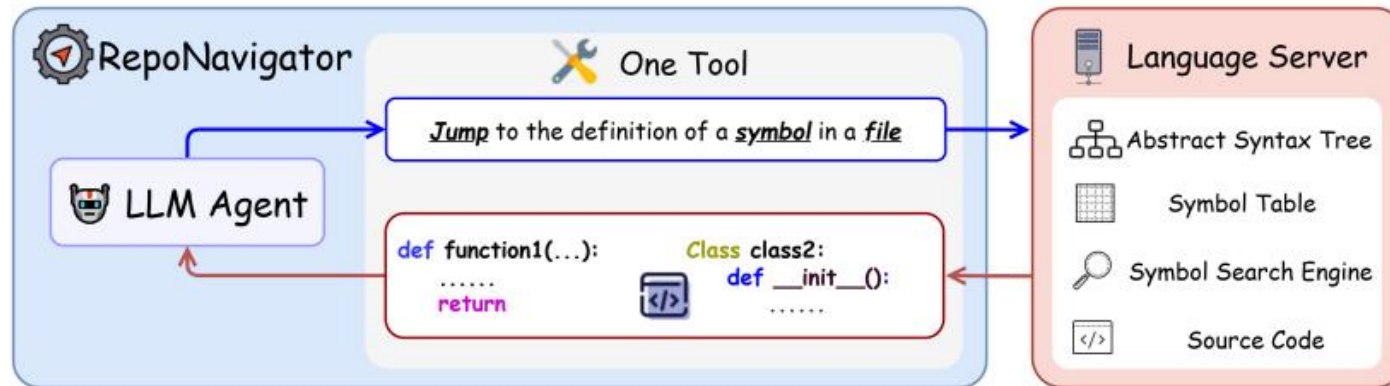
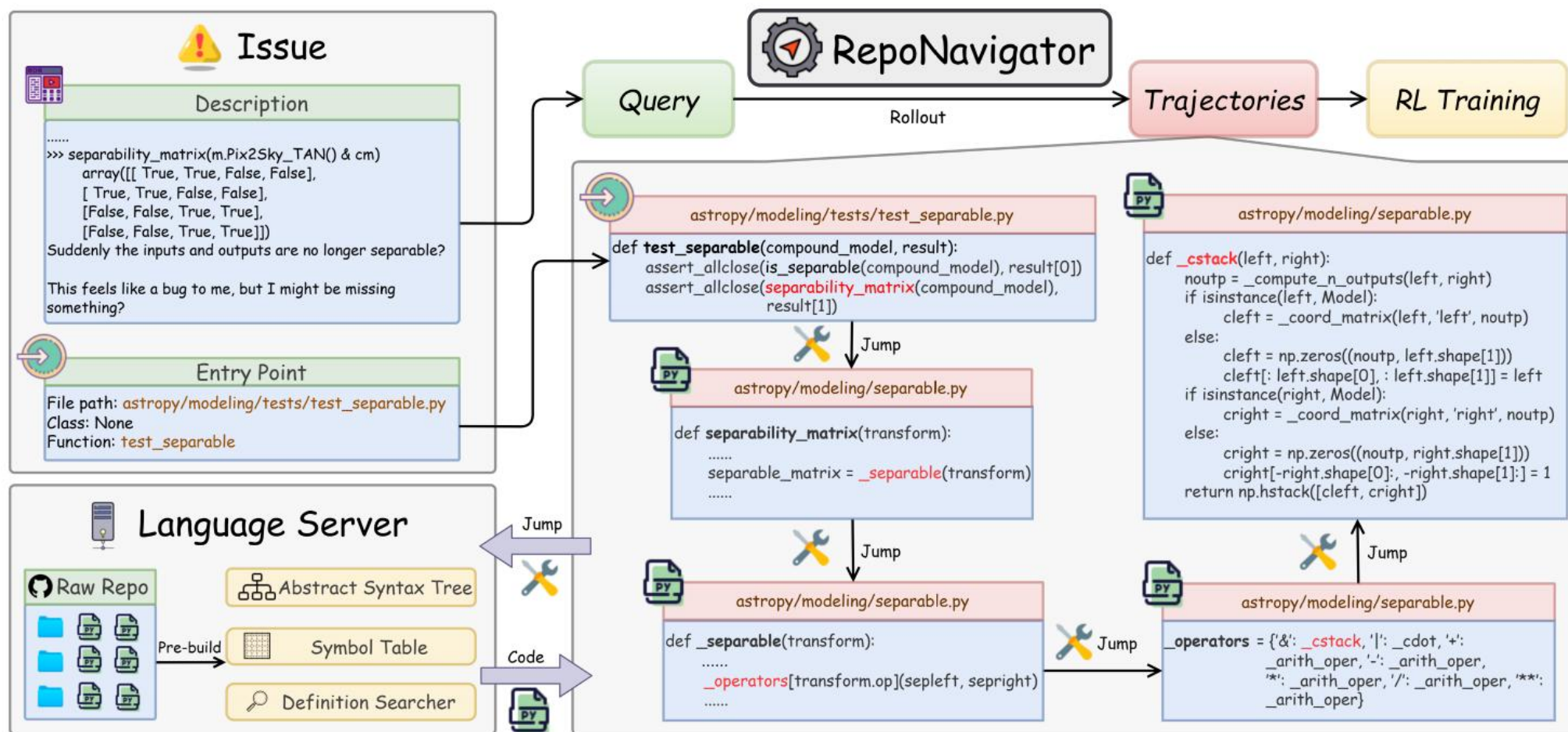


Figure 1: Illustration of an LLM navigating through a code repository. The LLM is equipped with a single yet powerful tool: `jump`, implemented via a language server.

2. Overview



3. GRPO Training

$$\mathcal{L}^{\text{GRPO}}(\theta) = \mathbb{E}_{(s_t, a_t) \sim \pi_{\theta_{\text{old}}}} \left[\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{\text{old}}}(a_t | s_t)} \hat{A}_t - \beta D_{\text{KL}}(\pi_{\theta_{\text{old}}}(\cdot | s_t) \| \pi_{\theta}(\cdot | s_t)) \right]$$

$$R(\hat{Y}, Y^*, \tau) = \lambda \cdot \text{DICE}(\hat{Y}, Y^*) + (1 - \lambda) \cdot S(\tau)$$

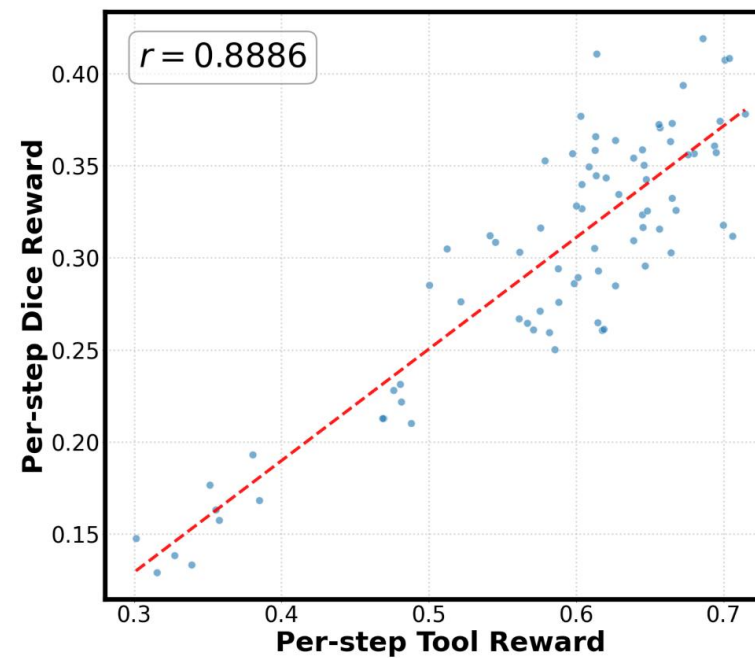
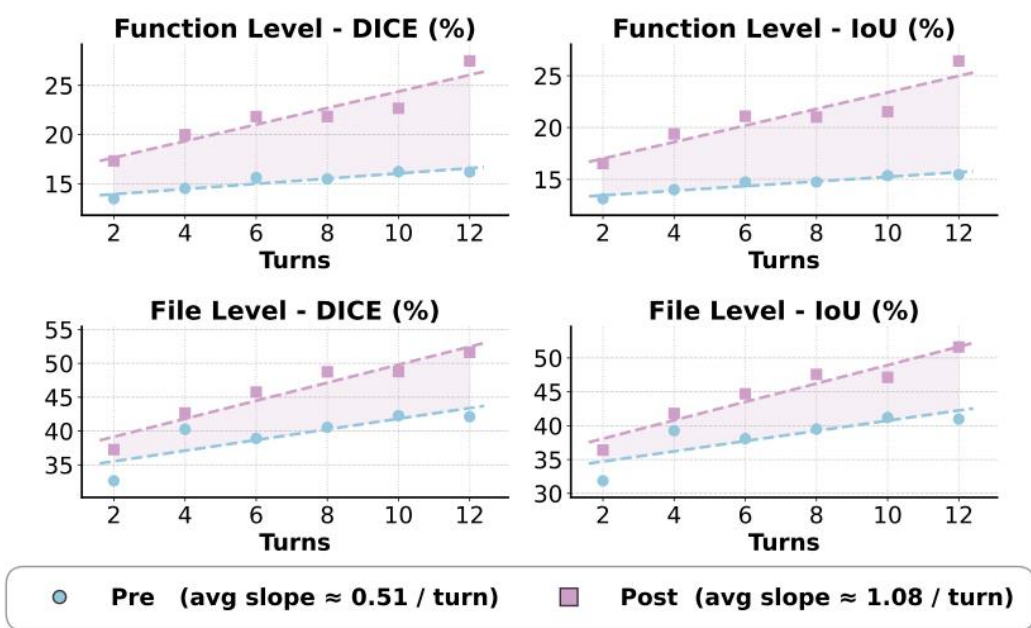
$$\text{DICE}(\hat{Y}, Y^*) = \frac{2 \times |\hat{Y} \cap Y^*|}{|\hat{Y}| + |Y^*|}$$

4. Experiments

Agent Pipeline	Model	Function-level				File-level			
		Recall	Precision	Sample-F1	IoU	Recall	Precision	Sample-F1	IoU
<i>closed-source Models</i>									
RepoSearcher	Claude3.7-Sonnet	66.80	19.90	28.30	17.89	89.71	21.04	33.15	20.67
RepoNavigator	Claude3.7-Sonnet	31.03	34.43	31.72	30.22	72.26	75.95	73.01	71.37
RepoNavigator	GPT5-chat	30.42	34.56	31.17	29.67	58.17	61.87	58.88	57.33
RepoNavigator	Claude4.5-Sonnet	43.97	45.76	43.62	41.31	80.68	81.92	79.94	77.49
<i>Qwen2.5-7B</i>									
Agentless	Training-free	24.92	12.93	15.31	11.74	63.01	19.32	27.82	18.85
CoSIL	Training-free	<u>29.30</u>	8.98	12.90	8.07	<u>70.12</u>	17.90	27.39	17.42
LocAgent	Training-free	17.62	11.71	12.71	10.31	60.96	34.88	40.67	33.33
OrcaLoca	Training-free	27.70	<u>20.29</u>	<u>21.70</u>	<u>17.92</u>	48.04	<u>48.65</u>	<u>47.36</u>	<u>45.77</u>
RepoSearcher	Training-free	7.86	8.08	7.48	6.85	48.61	21.15	27.36	20.39
RepoSearcher	Distillation+GRPO	63.26	19.24	27.37	17.59	84.11	19.97	31.64	19.57
RepoNavigator	Training-free	15.89	17.46	16.19	15.46	42.36	43.23	42.12	40.97
RepoNavigator	GRPO	26.69	30.34	27.49	26.43	50.62	53.83	51.63	50.62
<i>Qwen2.5-14B</i>									
Agentless	Training-free	25.20	14.30	16.14	12.28	75.65	19.76	29.88	19.30
CoSIL	Training-free	48.61	13.40	19.81	12.12	78.35	18.10	28.79	17.72
LocAgent	Training-free	35.62	13.32	17.71	12.32	71.42	31.66	40.77	30.64
OrcaLoca	Training-free	29.92	20.98	22.77	18.92	52.17	52.15	50.93	48.72
RepoSearcher	Training-free	26.13	11.96	14.35	10.60	74.77	18.80	28.79	18.15
RepoNavigator	Training-free	27.96	<u>25.77</u>	<u>25.58</u>	<u>23.00</u>	59.00	<u>56.68</u>	<u>56.39</u>	<u>53.74</u>
RepoNavigator	GRPO	31.02	30.08	29.23	26.84	61.60	58.97	58.90	56.36
<i>Qwen2.5-32B</i>									
Agentless	Training-free	40.79	24.07	27.33	22.08	78.93	25.60	35.38	24.96
CoSIL	Training-free	55.38	14.85	22.11	13.52	83.50	19.34	30.77	18.93
LocAgent	Training-free	46.79	16.29	21.48	14.18	79.39	34.18	44.18	33.24
OrcaLoca	Training-free	39.14	<u>25.59</u>	<u>28.72</u>	22.89	59.57	59.51	58.11	55.62
RepoSearcher	Training-free	18.83	19.05	18.10	16.79	66.60	39.81	45.81	37.93
RepoSearcher	Distillation+GRPO	69.50	20.29	29.11	18.23	89.33	20.27	32.93	20.35
RepoNavigator	Training-free	28.11	28.19	27.12	<u>25.16</u>	63.05	62.75	<u>61.67</u>	<u>59.28</u>
RepoNavigator	GRPO	33.71	37.19	34.09	32.30	67.29	70.76	67.75	65.75

Agent Pipeline	Model	Function-level				File-level			
		Recall	Precision	Sample-F1	IoU	Recall	Precision	Sample-F1	IoU
<i>Qwen2.5-7B</i>									
Agentless	Training-free	12.82	6.94	8.05	5.73	39.41	13.15	18.89	12.35
CoSIL	Training-free	8.64	3.33	4.58	2.87	26.64	8.47	12.11	7.70
LocAgent	Training-free	1.01	0.02	0.65	0.40	12.16	0.17	10.81	8.93
LocAgent	Distillation	13.76	1.61	2.66	1.48	37.44	5.11	8.39	5.00
RepoSearcher	Training-free	1.07	0.93	0.97	0.86	4.91	1.64	2.30	1.63
RepoNavigator	Training-free	9.84	<u>14.65</u>	<u>10.67</u>	<u>9.20</u>	30.50	<u>37.24</u>	<u>31.86</u>	<u>28.82</u>
RepoNavigator	GRPO	12.33	21.26	14.29	12.02	36.36	48.13	39.74	36.36
<i>Qwen2.5-14B</i>									
Agentless	Training-free	10.49	6.75	7.41	5.28	41.42	13.42	19.02	12.37
CoSIL	Training-free	10.73	4.67	5.96	3.94	34.31	9.97	14.81	9.30
LocAgent	Training-free	6.22	0.13	3.65	2.65	15.58	0.21	11.69	9.53
RepoSearcher	Training-free	2.79	1.38	1.69	1.14	17.37	5.17	7.60	4.84
RepoNavigator	Training-free	<u>14.36</u>	<u>19.74</u>	<u>15.27</u>	<u>12.00</u>	<u>43.57</u>	<u>54.52</u>	<u>46.06</u>	<u>41.07</u>
RepoNavigator	GRPO	16.05	25.25	18.06	14.58	46.85	58.64	49.72	45.14
<i>Qwen2.5-32B</i>									
Agentless	Training-free	11.08	7.31	7.98	5.80	43.07	13.89	20.07	13.11
CoSIL	Training-free	<u>14.03</u>	6.00	7.67	4.92	48.87	14.03	20.95	13.27
LocAgent	Training-free	8.72	0.17	4.30	2.90	25.73	0.38	19.77	16.50
RepoSearcher	Training-free	2.52	2.46	2.31	1.75	9.00	2.52	3.81	2.41
RepoNavigator	Training-free	13.96	<u>20.25</u>	<u>15.36</u>	<u>12.87</u>	<u>50.24</u>	<u>63.24</u>	<u>53.48</u>	<u>48.50</u>
RepoNavigator	GRPO	18.13	29.44	20.72	17.16	53.49	68.69	57.57	52.44

Role of Tool Reward



Influence on issue resolution

Table 3: Downstream influence: We use Qwen2.5-14B-Instruct on SWE-bench_Verified to evaluate issue resolution.

Agent Pipeline	Func-IoU(%)	Resolved(%)
Agentless	12.28	10.12
LocAgent	12.32	13.01
RepoSearcher	10.60	12.47
RepoNavigator	23.00	14.74
RepoNavigator+RL	26.84	15.03

5. Hint: Building Fewer and More Capable Tools

- Nowadays agents have multiple tools, and they generally have great effects.
- However, if a task can be finished with fewer tools, excessive tools will only impede the performance.
- Excessive tools are noises that may lead the agent astray.
- In SWE tasks, using one “Jump” tool is enough

Table 4: We change the tool set of RepoNavigator and present the function-level IoU (%) on Qwen2.5-7B-Instruct. From the table, we demonstrate that excessive tools do not boost RepoNavigator’s performance.

Jump	GetClass	GetFunc	GetStruc	IoU
✓	✓	✓	✓	13.71
✓	✓	✓	✗	21.44
✓	✗	✗	✓	24.00
✓	✗	✗	✗	24.28

Thank you!