ToolACE-R: Model-aware Iterative Training and Adaptive Refinement for Tool Learning

Xingshan Zeng, Weiwen Liu°, Xu Huang, Zezhong Wang, Lingzhi Wang, Liangyou Li,
Yasheng Wang, Lifeng Shang, Xin Jiang, Ruiming Tang, Qun Liu
zeng.xingshan@huawei.com; wwliu@sjtu.edu.cn

Motivation

Tool learning enables Large Language Models (LLMs) to invoke external tools (APIs, databases, services) to solve complex tasks. With tools, LLMs can: 1) Access up-to-date information; 2) Perform precise computations; 3) Interact with third-party systems.

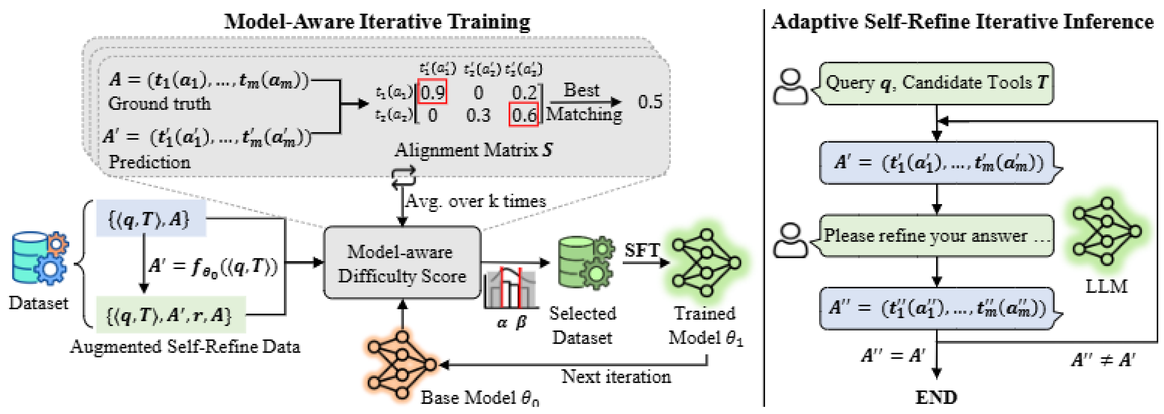
Most existing approaches rely on customized fine-tuning to teach models how to call tools.

Challenge 1: Which Training Data Is **Appropriate**?
- Synthetic data (mostly produced by strong LLMs) **mismatches model capability**: samples beyond the model's knowledge may cause hallucinations or degraded performance.

Challenge 2: Model Potential **Underexplored**.
- **Fully exploiting** existing data (e.g. self-refinement /augmentation).
- **Adaptive** scaling **test-time computation** in tool learning.
  - Simple queries often need no refinement;
  - Complex queries benefit from multiple iterations.

ToolACE-R

A unified pipeline to stimulate model potential in tool learning.



Overview of ToolACE-R Framework

Model-Aware Iterative Training

- Model-aware Difficulty:
  - Key Idea: measure difficulty by how well the current model can **reproduce the reference tool calls** before training.
  - **Optimal bipartite matching** to align tool calls.

$$S_{ij} = \begin{cases} \frac{|a_i \cap a'_j|}{|a_i \cup a'_j|} & \text{if } t_i = t'_j \\ 0 & \text{otherwise} \end{cases} \qquad O(A, A') = \frac{\sum_{(i,j)\in\mathcal{M}} S_{ij}}{\max(m, m')}$$

Parameter-level alignment score Action-level alignment score

$$D(\langle q, T\rangle, A) = 1 - \frac{1}{k}\sum_{l=1}^{k} O(A, A'^{(l)})$$

Difficulty score with k runs

- Data Augmentation with Self-Refinement:
  - For what? Teach the model:
    - **How to refine**: refine problematic output A' to groundtruth A.
    - **When not to refine**: retain $A' = A$.

$$\{< \boldsymbol{q}, \boldsymbol{T} >, \boldsymbol{A}\} \rightarrow \{< \boldsymbol{q}, \boldsymbol{T} >, \boldsymbol{A'}, \boldsymbol{r}, \boldsymbol{A}\}$$

"Please refine your answer…"

- Iterative Training:
  - Data selection: select data samples with $\alpha < D < \beta$.
  - Align with Model's **evolving capability**.

Adaptive Self-Refine Iterative Inference

- Iteratively refine its own tool call output.
- Stopping criteria: two consecutive iterations are identical.

Experiments

- Settings:
  - **Base model:** LLaMA-3.1-8B-Instruct (trained with LoRA)
  - **Dataset:** ToolACE-Full (select with model-aware difficulty)

- ToolACE-R achieves **over 4.5 average score improvement**, compared to ToolACE-8B.

Models	Non-Live				Live				Overall		
	Simple	*Multiple*	*Parallel*	*Multiple Parallel*	*Simple*	*Multiple*	*Parallel*	*Multiple Parallel*	*Non-live*	*Live*	*Overall*
GPT-4o-2024-11-20	77.17	95.00	93.50	85.00	84.50	79.30	87.50	70.83	87.67	80.24	83.96
GPT-4.1-2025-04-14	**80.50**	94.00	93.00	87.50	**85.66**	76.54	**93.75**	75.00	88.75	78.46	83.60
GPT-4o-mini-2024-07-18	80.08	90.50	89.50	87.00	81.40	76.73	**93.75**	79.17	86.77	77.87	82.32
Gemini-2.5-Pro-Exp-03-25	67.50	91.00	83.50	73.00	79.46	64.96	81.25	79.17	78.75	68.17	73.46
DeepSeek-V3	78.67	95.50	91.00	**91.50**	83.72	**82.15**	81.25	62.50	89.17	**82.09**	85.63
Llama3.1-70B-Inst	77.92	96.00	94.50	**91.50**	78.29	76.16	87.50	66.67	89.98	76.53	83.26
Llama3.1-8B-Inst	71.50	93.50	86.50	86.00	73.26	68.95	56.25	50.00	84.37	69.28	76.83
Qwen2.5-7B-Inst	70.50	94.00	90.50	82.50	74.81	71.89	62.50	62.50	84.38	72.17	78.28
ToolACE-8B (FC)	76.67	93.50	90.50	89.50	73.26	76.73	81.25	70.83	87.54	76.02	81.78
Hammer2.1-7B (FC)	78.08	95.00	93.50	88.00	76.74	77.40	81.25	70.83	88.65	77.20	82.92
xLAM-2-8b-fc-r (FC)	73.08	93.50	87.00	84.00	74.81	66.29	56.25	50.00	84.40	67.51	75.95
ToolACE-R (FC)	79.75	**97.00**	**95.50**	**91.50**	82.95	81.10	**93.75**	**87.50**	**90.94**	81.72	**86.33**

Table 1: Accuracy comparison on BFCL (Last updated on 2025-04-25). The best results in each category are marked in **bold**. The second best results are underlined. FC indicates the models are fine-tuned for function calling.

- **Consistent improvement** across multiple benchmarks.
- Model-aware data selection contributes most.

Models	ACEBench	APIBank	ToolAlpaca
GPT-4o	**87.00**	77.16	83.87
GPT-4o-mini	81.63	73.10	83.87
Llama3.1-8B-Inst	50.63	60.41	79.03
ToolACE-R	83.88	75.89	85.48
+ Adaptive SR	84.00	**81.22**	**88.71**

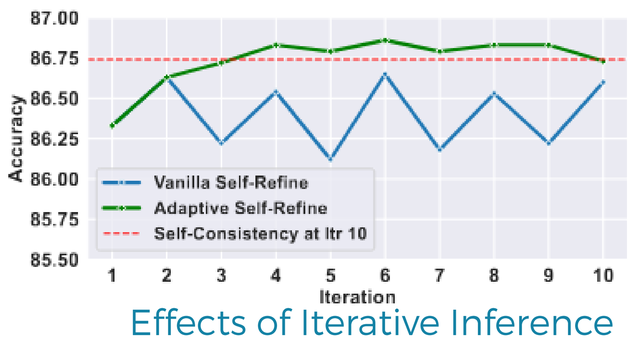
Table 2: Accuracy comparison on ACEBench, API-Bank and ToolAlpaca. "SR" is short for "Self-Refine".

Models	Non-live	Live	Overall
ToolACE-R + Adaptive SR	**91.32**	**82.34**	**86.83**
ToolACE-R	90.94	81.72	86.33
- Data Selection	88.96	81.13	85.04
- SR Data	88.75	80.09	84.42
Llama3.1-8B-Inst (base)	84.37	69.28	76.83

Table 3: Ablation Study of our proposed modules on BFCL. "SR" is short for "Self-Refine".

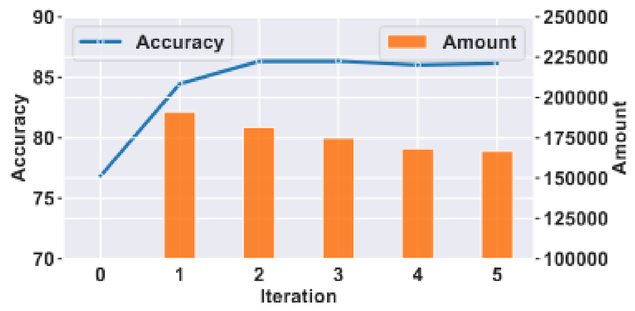
Iterative Effectiveness

- Iterative Inference
  - Adaptive self-refine exhibits **more stability**.
  - Consistent better performance compared to naïve self-consistency.



Effects of Iterative Inference

- Iterative Training
  - Most improvement locates in the **first two iteration**.
  - Training sample amount decreases monotonically.



Effects of Iterative Training

More Observations

- Selecting thresholds (α & β): not sensitive.
- Work well even k=1 when computing difficulty score.
- Difficulty score computed with other models cannot well align with current model.

Conclusion

- We propose ToolACE-R, a unified pipeline for tool learning:
  - Model-aware iterative training: a new metric for model-aware difficulty & augmented with self-refinement.
  - Adaptive self-refinement: iterative inference framework.
- A new exploration on model potential beyond data synthesis in tool learning.
- Consistent improvement, as well as generalizability and robustness across diverse settings.

Team-ACE

AAAI 2026: The 40th Annual AAAI Conference on Artificial Intelligence, Singapore, Jan 20-27, 2026