



Don't Adapt Small Language Models for Tools; Adapt Tool Schemas to the Models

Jonggeun Lee* Woojung Song* Jongwook Han Haesung Pyun Yohan Jo†
Graduate School of Data Science, Seoul National University



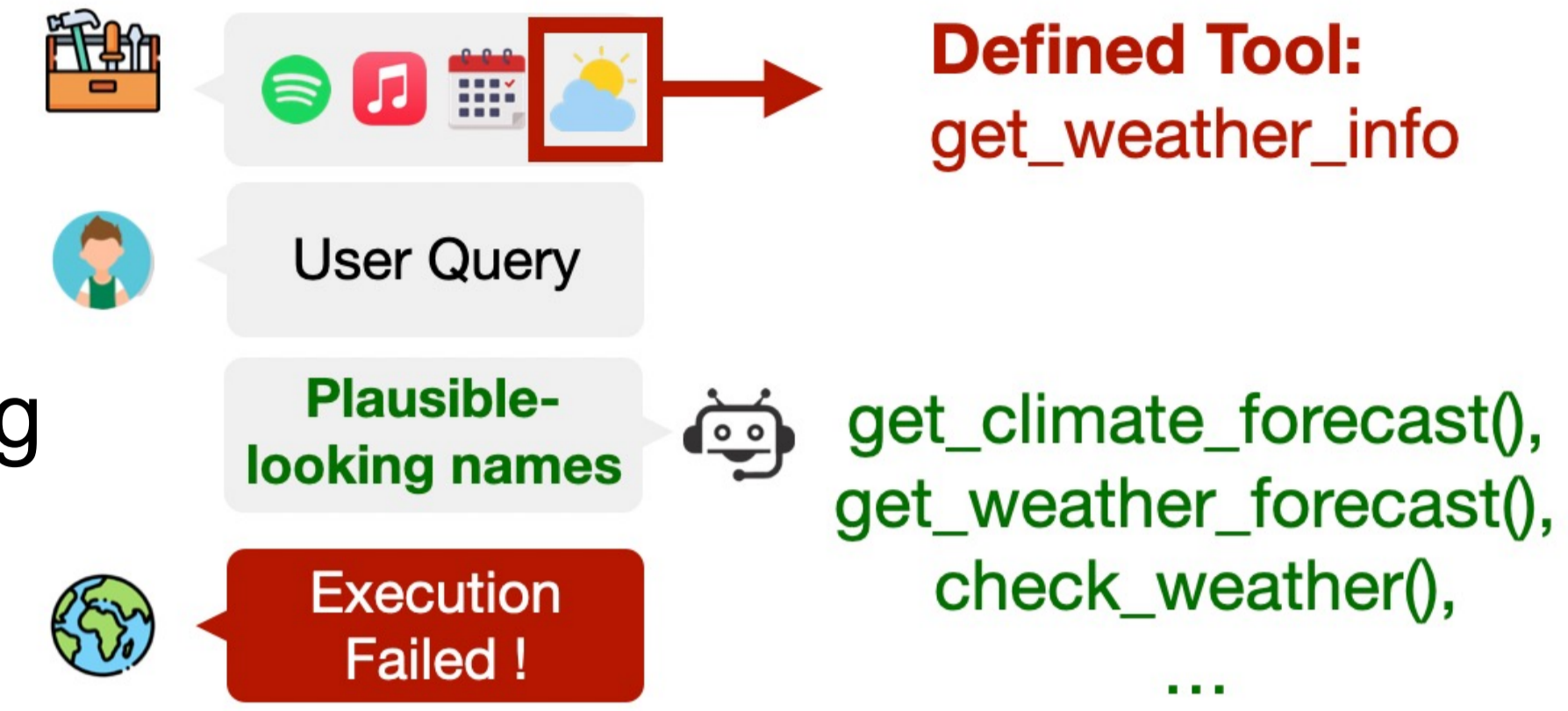
Motivation

- For efficiency, SLMs ($\leq 8B$) are increasingly deployed as sub-agents in multi-agent systems, but they struggle with tool use.
- A common failure is **schema misalignment**: even with the right tool in context, **SLMs hallucinate plausible-looking tool names** absent in the context, falling back on **internalized naming conventions**.

💡 Rather than training models to fit unfamiliar schemas (e.g., SFT, RL), we **adapt the tool schema to fit the model's knowledge**.

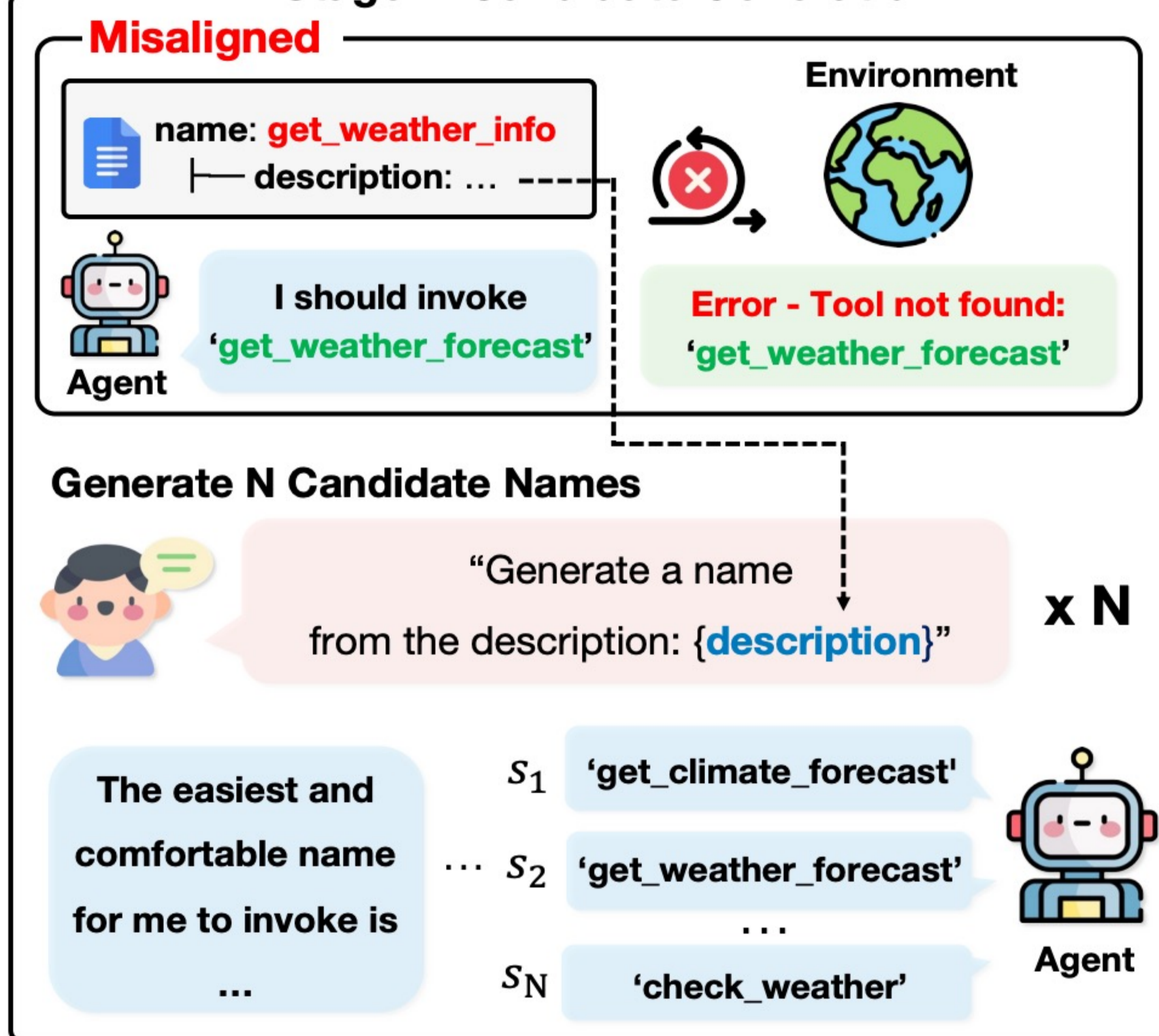
💡 We introduce **PA-Tool**, a **training-free** method that renames tool components (i.e., tool names and parameter names) using **peakedness** — a signal that measures how deeply a name is internalized.

Schema Misalignment
Query: I'm planning a trip to San Diego in July and need to check the **weather forecast**.

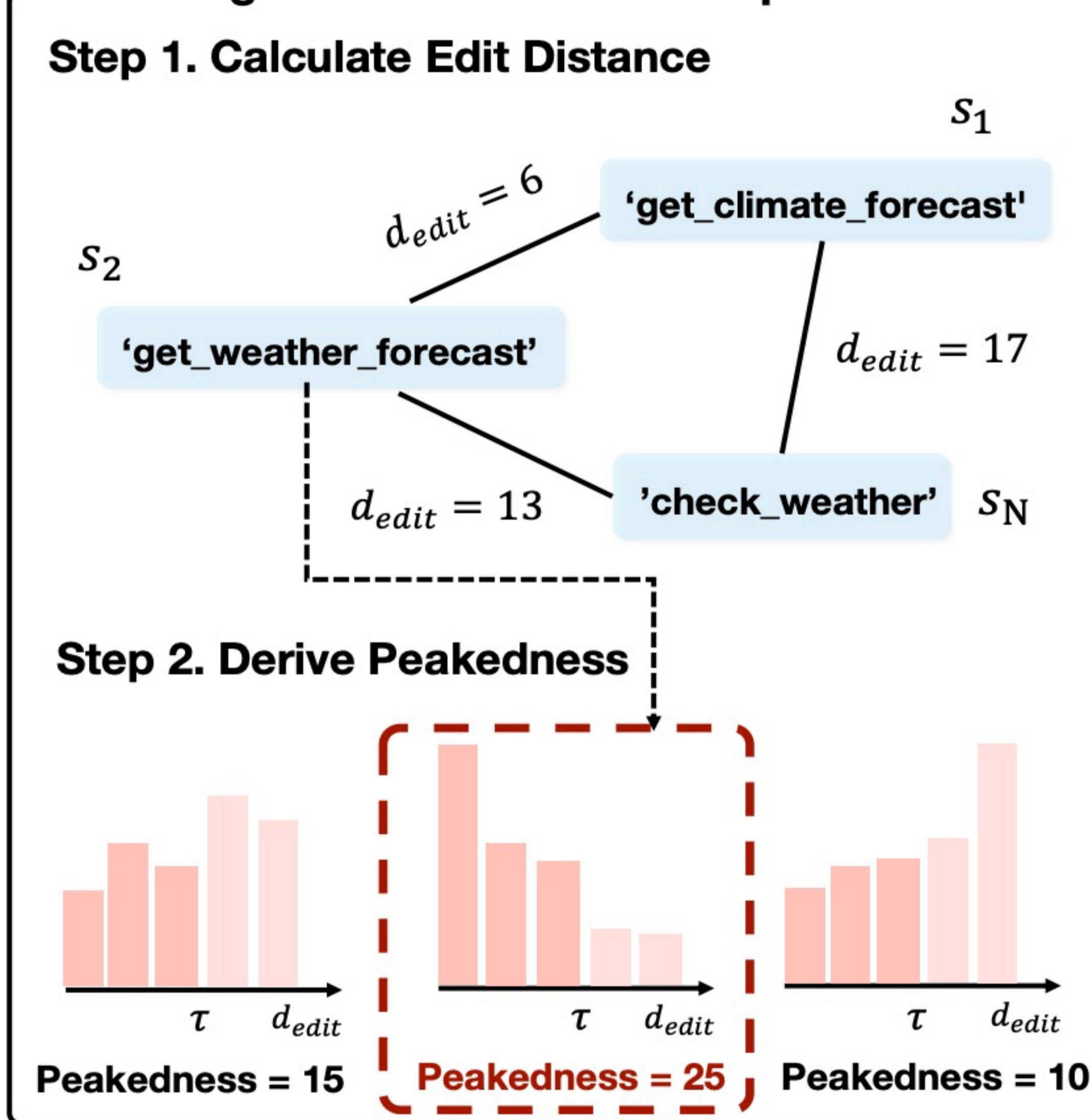


PA-Tool

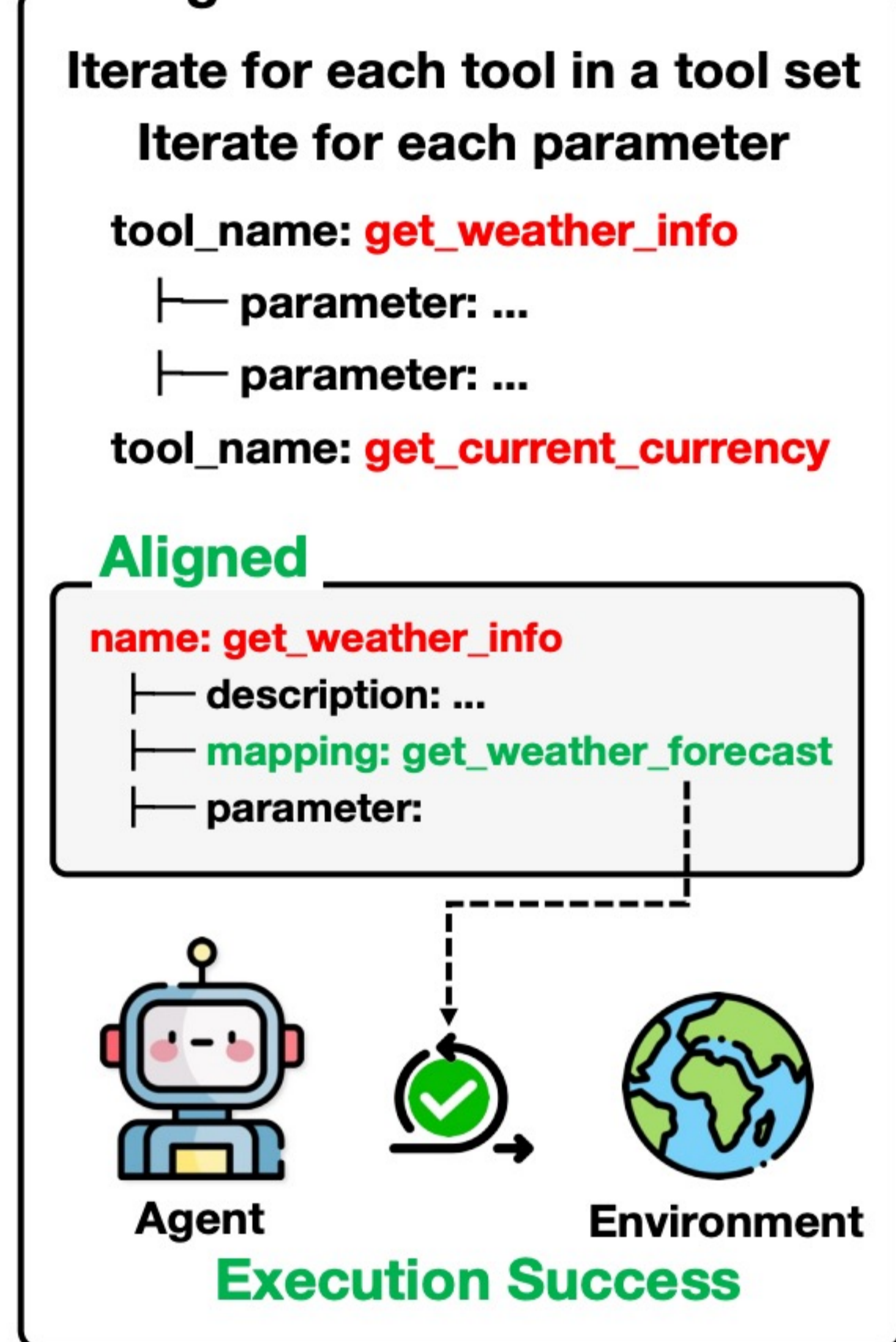
Stage 1. Candidate Generation



Stage 2. Peakedness Computation



Stage 3. Schema Selection



Similarity Threshold: $\tau = \alpha \cdot \ell_{\max}$

Peakedness: $\phi(s_i) = \sum_{j \neq i} \mathbb{I}(d_{\text{edit}}(s_i, s_j) \leq \tau)$

Renamed Schema: $s^* = \arg \max_{s_i \in \mathcal{C}} \phi(s_i)$

Experiments

Experimental Setup

- Model:** Llama3.1-8B (full results in the paper)
- Benchmarks:** MetaTool, RoTBench
- Metric:** Acc (%) on tool selection & parameter identification
- PA-Tool:** 32 candidates, $t = 0.4$, $\alpha = 0.2$

Model	MetaTool				RoTBench	
	Similar	Scenario	Reliability	Multi-tool	Tool.	Param.
Llama-3.1-8B						
Base	61.5	73.9	53.5	78.7	58.1	17.1
Greedy	64.6	72.9	51.5	78.9	63.8	18.1
MostFreq	68.8	79.3	66.4	85.7	66.7	18.1
PA-Tool	70.4	79.9	66.0	88.3	68.6	18.1

*RoTBench results show single-turn performance only.

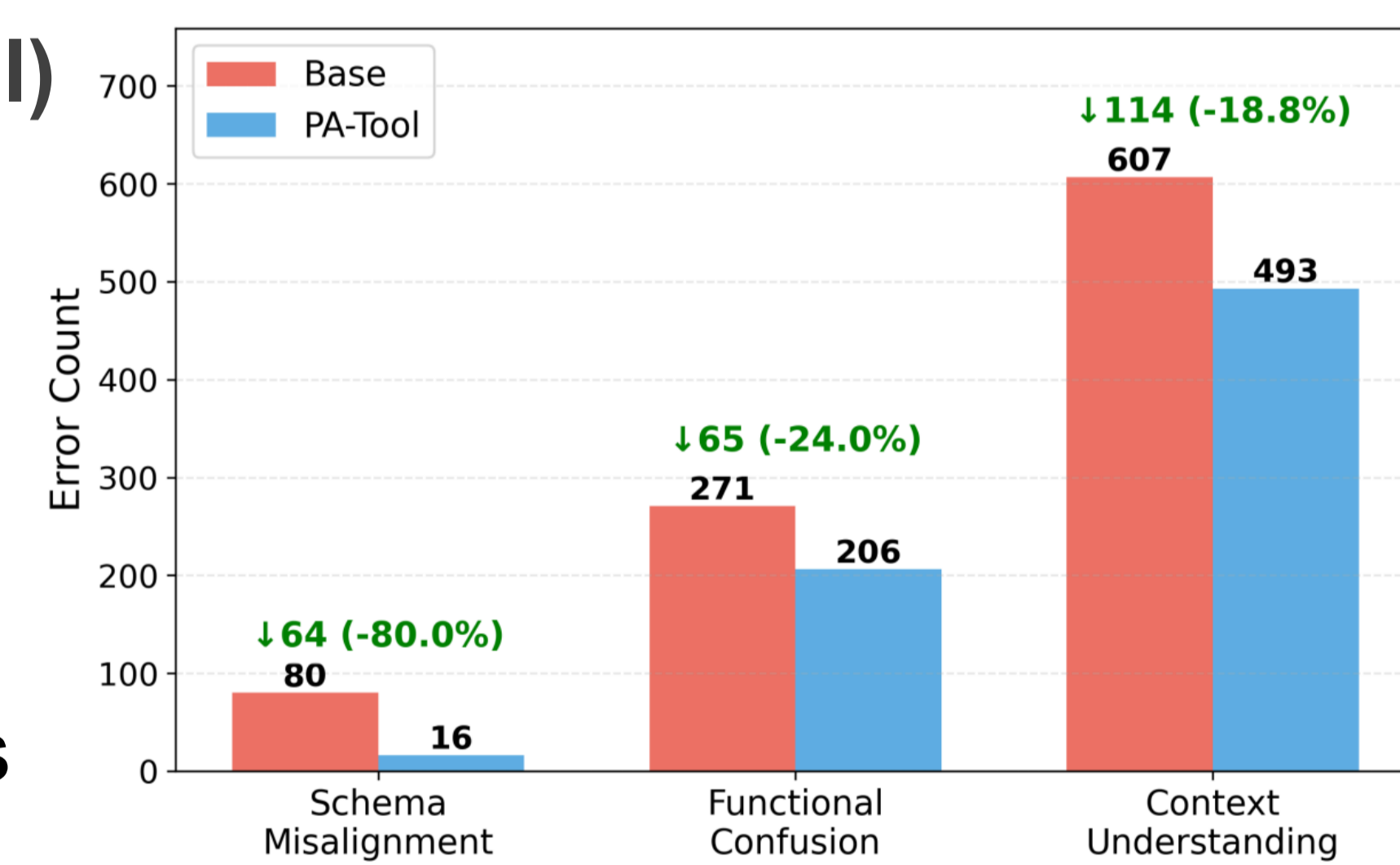
PA-Tool vs. SFT

- How does training-free PA-Tool compare to **fine-tuning**, and what happens when we combine the two?

Model	MetaTool				RoT Bench
	Similar	Scenario	Reliability	Multi-tool	Tool.
Llama-3.1-8B					
Base	65.2	71.1	55.1	79.8	58.1
PA-Tool	72.7	75.9	67.2	89.9	68.6
SFT	71.2	77.6	57.1	82.8	61.0
SFT+PA-Tool	72.7	80.8	57.1	89.9	65.7

Error Analysis (Base vs. PA-Tool)

- Where do the gains come from?
- Does PA-Tool really fix schema misalignment?



Human Evaluation

- Are the renamed names still readable to human developers?

Dimension	Rating		Preference	
	Orig	PA	Orig	PA
Understand.	2.72	3.41	10.2	52.3
Func. match	2.78	3.44	10.4	50.9

Original	PA-Tool	Description
PA-Tool improves clarity		
Figlet	ascii_converter	Convert text into ASCII fonts
ad4mat	track_traffic	Monetize traffic via tracking links
universal	web_analyzer	Access web pages, analyze PDFs, etc.
PA-Tool reduces clarity		
ProductComparison	compare_options	Compare product options
StrologyTool	astro_services	Provide astrology services
ShoppingAssistant	cart_qr_generator	Manage cart and display QR codes

★ Takeaway

★ Adapt Tool Schemas to the Models!

💡 **Tool component names matter to SLMs**
SLMs rely on tool names, not just descriptions.

💡 **Align names with knowledge**

Rename schemas to what the model already knows, using peakedness. No training needed, no risk of catastrophic forgetting.