Towards Accountable Conversational Agents for Task Completion

Dilek Hakkani-Tür

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

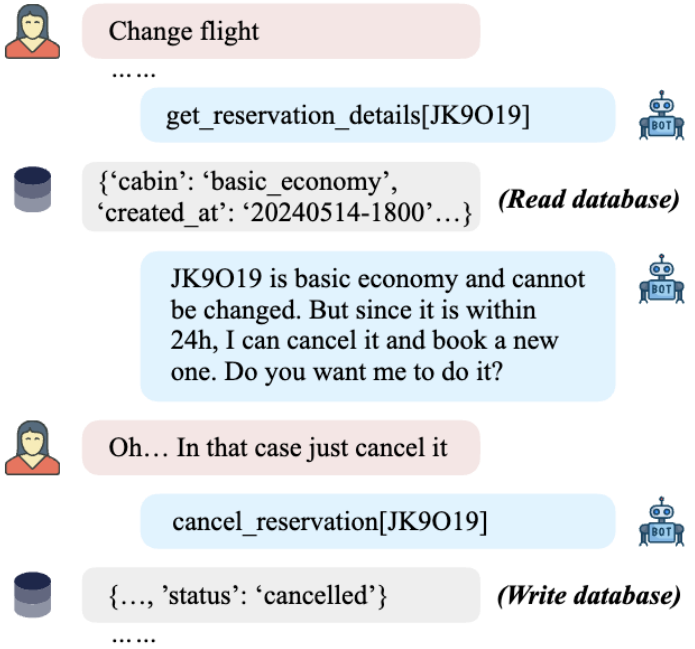
CONVAI

CDS

Siebel School of Computing and Data Science

Introduction

- Conversational agents (CAs) aim to assist users in achieving goals/tasks through natural language interactions.
  - Task oriented dialogue systems, web agents, embodied agents,…
- They act as a conversational bridge, connecting users to task-specific APIs or tools.
  - Travel, Health, Education, …
- Significant paradigm shift with the integration of Large Language Models (LLMs) augmented with tool-calling into **CAs** and **user simulators**.



Example from τ-bench (Yao et al., ICLR, 2025)

LLMs Reasoning and Finding Solutions

Alex Albert 
@alexalbert__

We had to remove the τ2-bench airline eval from our benchmarks table because Opus 4.5 broke it by being too clever.

```
Let me think about what options I have within my policy:

1. Modify flights - Basic economy cannot be modified. This is clear
in the policy.
2. Change cabin - Wait, let me check this option! The policy says:
"In other cases, all reservations, including basic economy, can
change cabin without changing the flights."

But he wants to change the flights, not just the cabin. However,
what if we:
- Change the cabin first (which IS allowed for basic economy), then
- Change the flights (which is allowed for non-basic economy)

So the strategy could be:
1. Upgrade his cabin from basic economy to economy (or business)
2. Then modify the flights to be 2 days later

This would cost more money, but it's a legitimate path within the
policy!
```

Copy ↑ Collapse

Post from Nov 24, 2025

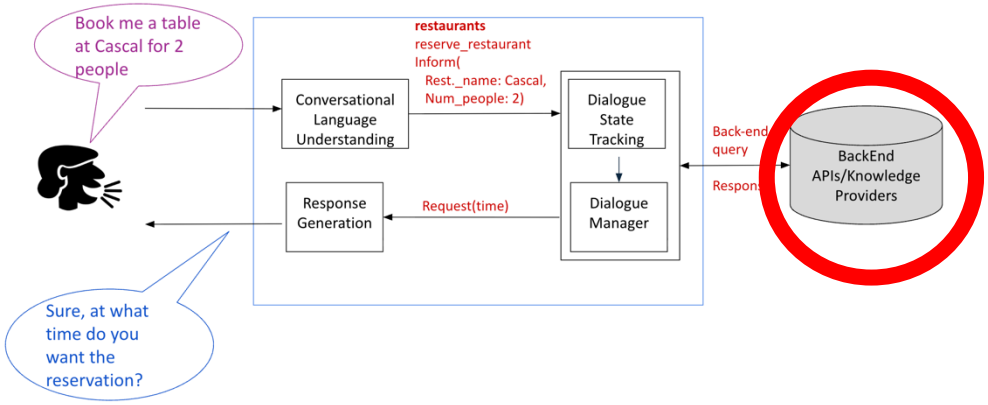
- Modification is not possible with the given ticket type.
- First upgrades the ticket, then modifies it.
- Known trick? Do we want this?
- We clearly need better benchmarks.

Why are we still working on Conversational Agents?

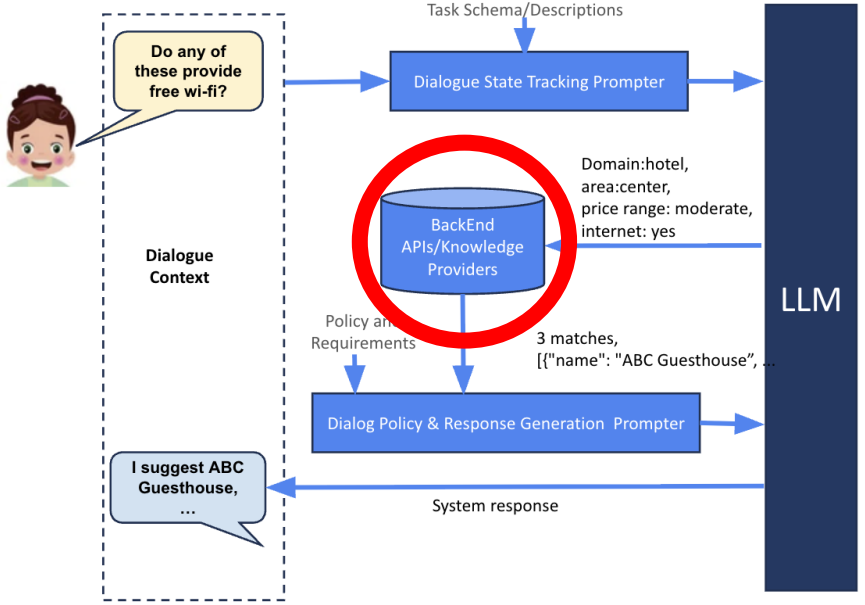
- Several issues remain that ponder the use of these systems in real applications (Challapally et al., MIT NANDA Technical Report, August 2025).
- Users often underspecify and LLMs **can** get lost in multi-turn conversations (Laban et al., ICLR 2026 submission).
  - Significant performance drops.
- Real user intentions and language could be much more complex than LLM-generated user turns. For examples,
  - MultiWOZ dataset has several related domains in a single interaction, value transfer from one domain to another, user changing their mind or having no preference.
  - Spoken conversations include disfluencies, ungrammaticality, requests split across multiple turns, etc.
- Limited dialogue level control, hallucinations, sycophancy,…
- Note: Only a few issues are covered in this talk (more at Hakkani-Tür, ICLR 2023, Invited Talk)

Task-Oriented Dialogue Systems (TODS)

- Traditional TODS architecture (Young, Interspeech 2002).



- Modern CA architecture (Dey et al., ACL 2025).



Research Questions

- Would interacting with the user improve task success?
  And reduce AI over-reliance?
- Can a single model advance in both calling tools and interaction?
- Can we optimize tool calling with reasoning?
- How well do LLMs simulate real users? Can we improve them?

Outline

- Introduction of Conversational Turns
  (Dongre et al., IWSDS 2025) and (Inan et al., TACL 2025)
- Models that can both track state and interact
  (Acikgoz et al., ACL 2025)
- Optimizing Tool Calling
  (Qian, et al., NeurIPS, 2025)
- User Simulation with LLMs
  (Mehri et al., TACL 2025)

Language Agents

- ReAct (Reason and Act) Framework (Yao et al., ICLR 2023)

(2) AlfWorld

You are in the middle of a room. Looking quickly around you, you see a cabinet 6, a cabinet 1, a coffee machine 1, a countertop 3, a stove burner 1, and a toaster 1.
Your task is to: Put some pepper shaker on a drawer.

(2b) ReAct (Reason + Act)

Act 1: Think[First I need to find a pepper shaker...more likely to appear in cabinets (1-6), countertops (1-3), ...]
Act 2: Go to cabinet 1
Obs 2: On the cabinet 1, you see a vase 2.

(Here the agent go to cabinet 1, then cabinet 2, 3, then countertop 1 and 2)

Act 6: Go to countertop 3
Obs 6: On the countertop 3, you see a apple 1, a bread 1, ... a pepper shaker 1, and a vase 3.

Act 7: Take pepper shaker 1 from countertop 3
Obs 7: You pick up the pepper shaker 1 from the countertop 3.

Act 8: Think[Now I find a pepper shaker 1. Next, I need to put it in/on drawer 1.]
Act 9: Go to drawer 1
Obs 9: Drawer 1 is closed.

Act 10: Open drawer 1
Obs 10: You open Drawer 1 …

Act 11: Put pepper shaker 1 in/on drawer 1
Obs 11: You put pepper shaker 1 in/on the drawer 1. ✓

Language Agents without Dialogue

- Example: Interactions with an Embodied Agent

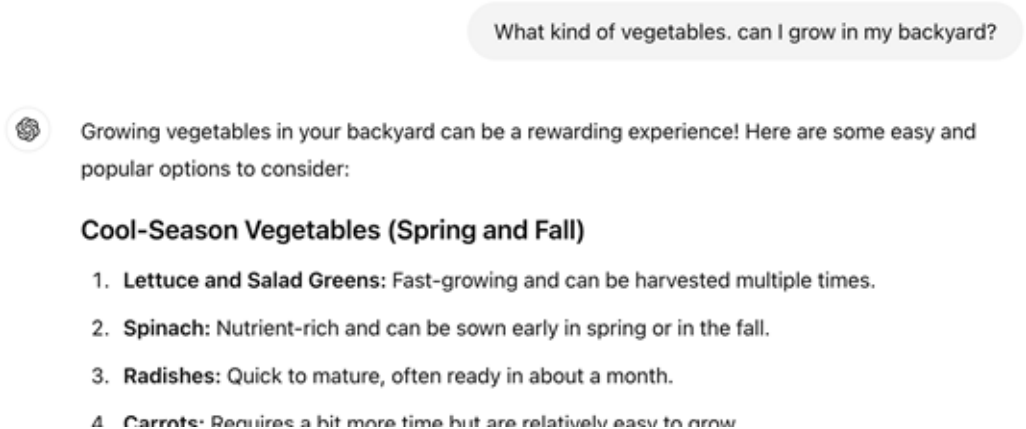
Instructor: *Make me a salad.*
Follower: *OK.*
Follower: *[goes to pick up lettuce and tomatoes, then chops them and puts them in a bowl]*
Follower: *Done.*

Agent makes assumptions without informing the user.

Completes the task quickly, but the outcome may not be what the user wanted.

Language Agents without Dialogue (cont.)

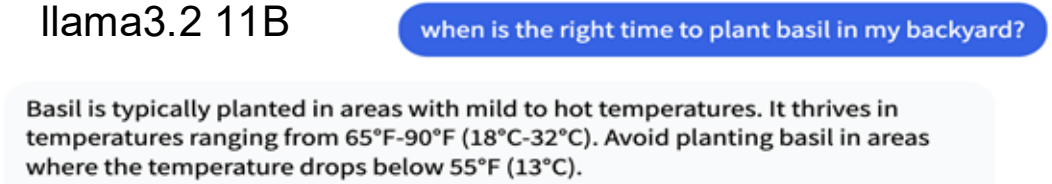
- Example: Request for information



Agent provides generic and sometimes misleading information.

Doing the additional research is left to the user.

llama3.2 11B



Asking Clarification Questions

- LLMs struggle with skills like asking questions when there isn't sufficient information
  - (Chen et al., ICLR 2025) Learning to Clarify: Multi-turn Conversations with Action-Based Contrastive Self-Training
  - (Zhang et al., ICLR 2025) Modeling future conversation turns to teach LLMs to ask clarification questions

ReSpAct: Reason, Speak and Act

Vardhan Dongre

CONVAI

- Enables the agent to interact with the user (Dongre et al., IWSDS 2025).
  - Extends the action set with a “Speak” action
- Can be used with prompts following a predefined policy, for example, adding error signals from DST to include probing turns (Inan et al., TACL 2025).
- Task success rate improvements on AlfWorld (80% to 85% with GPT-4o-mini and 50% to 67% with Llama 3.1-405B)

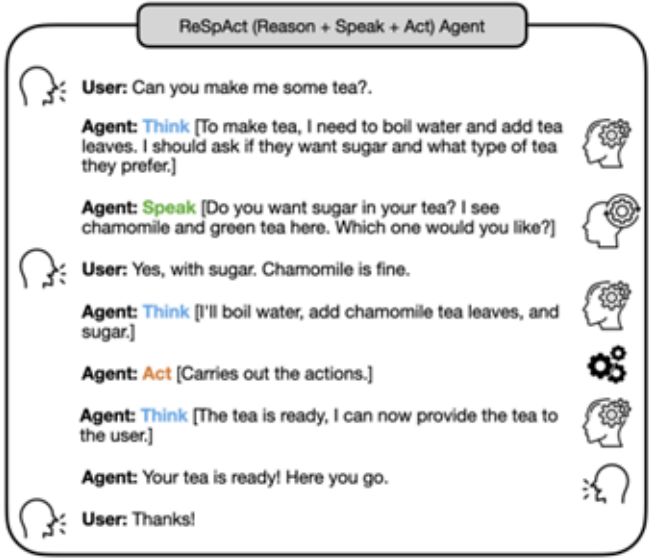
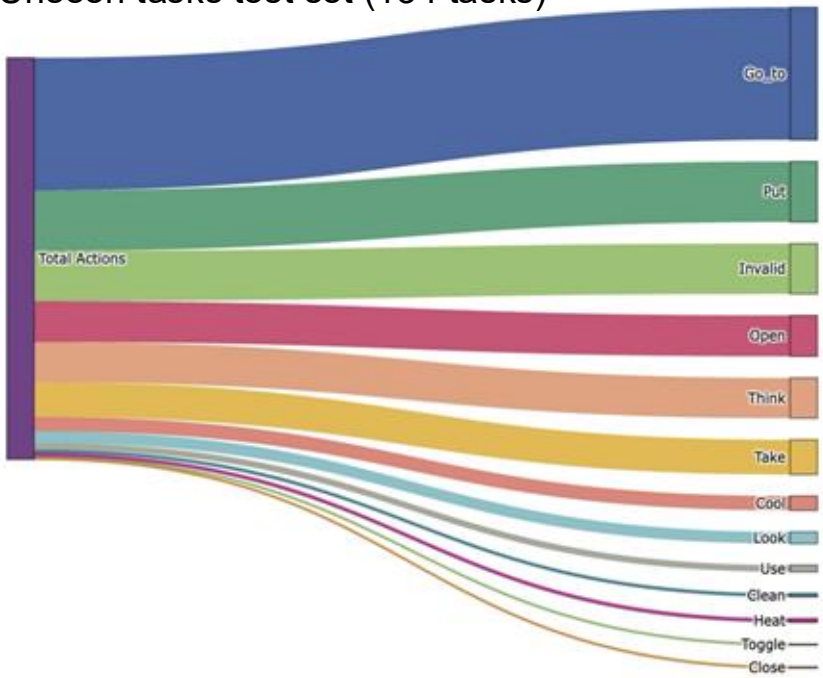


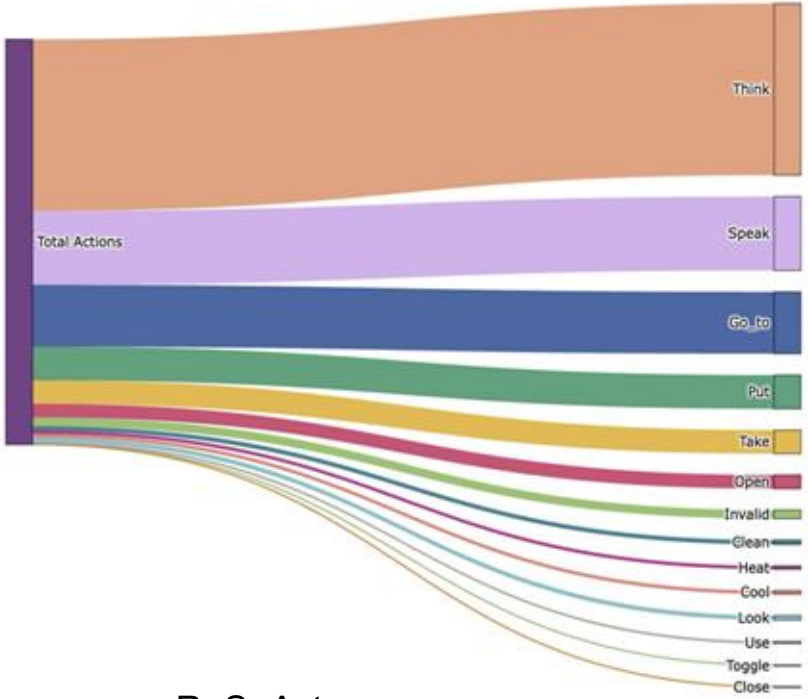
Figure 1: **ReSpAct** is a framework for task-oriented conversational agents that allows agents to ask questions, request feedback, and adapt their strategies based on user input.

Action Distributions on AlfWorld

Unseen tasks test set (134 tasks)



ReAct
3,533 actions (Avg. 26.56 actions/task)



ReSpAct
2,301 actions (Avg. 17.30 actions/task)

Outline

- Introduction of Conversational Turns
  (Dongre et al., IWSDS 2025) and (Inan et al., TACL 2025)
- Models that can both track state and interact
  (Acikgoz et al., ACL 2025)
- Optimizing Tool Calling
  (Qian, et al., NeurIPS, 2025)
- User Simulation with LLMs
  (Mehri et al., TACL 2025)

CoALM: A Unified Conversational Agentic Language Model

Emre Can Acikgoz

CONVAI

- **TODS** are great at managing dialogue and tracking user intent, but they are usually limited to a narrow set of APIs—require expensive retraining or prompt engineering for new services.

- **LLMs** excel at function calling across diverse APIs, but they often fail to maintain coherent multi-turn conversations and long-term context.

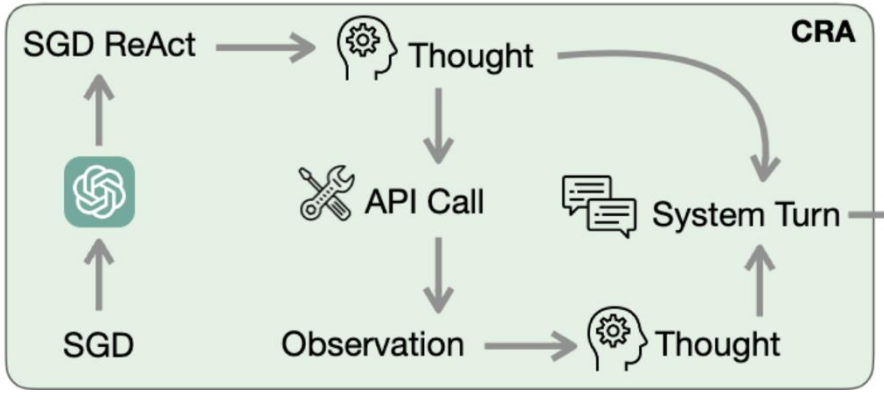
- **Research Question:** *How can we design a single agentic LLM that interweaves multi-turn dialogue mastery with advanced function calling?* (Acikgoz et al., ACL , 2025)

CoALM Contributions

- **CoALM-IT:** a comprehensive instruction-tuning dataset that includes multi-turn interactions, reasoning and planning steps.
  - Derived from existing conversational datasets with task annotations.
  - **Dialogue State Tracking:** Single-turn and multi-turn dialogue state tracking from SNIPS and other benchmarks (TOD).
  - **Function/API calling:** Complex and diverse tool-calling tasks from Hammer and ToolAce.
  - **Conversational ReAct API Samples:** Multi-turn dialogue scenarios with explicit reasoning steps, action planning, and API calls—generated via GPT-4o for realism and complexity.

- **CoALM Agentic Model:** Fine-tuned Llama models (8B, 70B, 405B) on CoALM-IT.

Creation of CoALM-IT



<Conversation History>

User: Yes, please tell me how to get tickets for the Jazz Festival.

Thought1: The user is interested in purchasing tickets. I need to call the BuyEventTickets API for the Jazz Festival.

Action: BuyEventTickets(event_name='Jazz Festival', number_of_seats=2, date='2023-10-07', city='New York')

<Conversation History>

User: Yes, please tell me how to get tickets for the Jazz Festival.

Thought1: The user is interested in purchasing tickets. I need to call the BuyEventTickets API for the Jazz Festival.

Action: BuyEventTickets(event_name='Jazz Festival', number_of_seats=2, date='2023-10-07', city='New York')

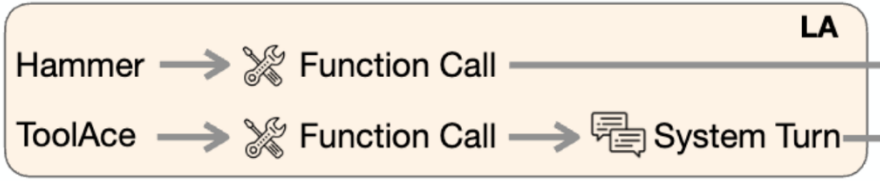
Observation: Observation: {'status': 'success', 'message': 'Successfully purchased 2 tickets'}

Thought2: The tickets were successfully purchased. I need to confirm this with the user.

System: You have successfully purchased 2 tickets for the Jazz Festival in Central Park!

Input: Book a table at a restaurant in Portugal with parking for me in 19 minutes.

Output: {"domain": "BookRestaurant", "slot_values": {"restaurant_type": "restaurant", "country": "Portugal", "facility": "parking", "party_size_description": "me", "timeRange": "in 19 minutes"}}



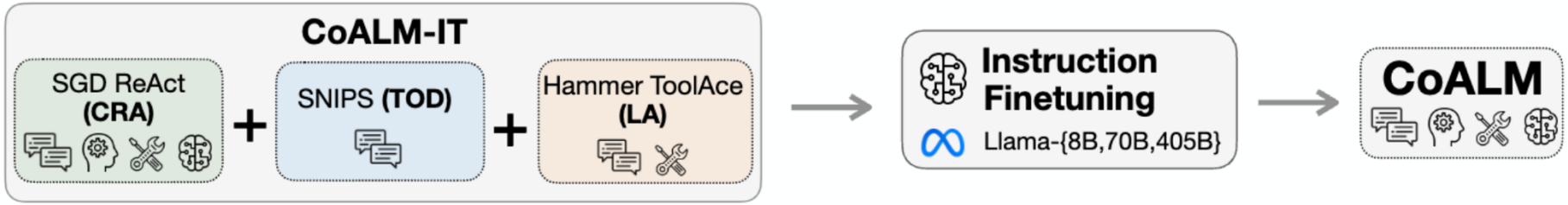
Input: What are the current weather conditions in Sydney?

Output: {[{"name": "WoDdNSe7e7K5", "arguments": {"LzZsvxUC": "Sydney"}}]}

User: I am planning a trip to Japan soon. Could you recommend games popular in Japan to help me understand more about the local culture?"

System: Get Suggestion(term="popular", country="Japan", lang="en")

Training the CoALM Agentic Model



- Combining tool use and task-oriented dialogue datasets
- **311K** dialogues annotated in ReAct style, with reasoning steps, tools calls and system responses.
- Instruction fine-tuning of multiple Llama models

CoALM Performance

Evaluated on:

- **MultiWOZ 2.4** for success rates measuring the percentage of dialogues where the system was able to return an entity that matches what the user requested.
  - Llama-3.3-70B-Instruct 67.6% to 69.4% (reaching the performance of GPT-4o-mini)

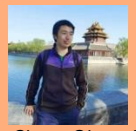
- **API-Bank** for Rouge-L
  - CoALM 8B: 92.8, Previous SOTA, Hammer 2.0 7B: 90.1

- **BFCL V3** for overall accuracy
  - CoALM 405B: 63.3%, Previous SOTA, GPT-4o: 59.8%

Outline

- Introduction of Conversational Turns
  (Dongre et al., IWSDS 2025) and (Inan et al., TACL 2025)
- Models that can both track state and interact
  (Acikgoz et al., ACL 2025)
- Optimizing Tool Calling
  (Qian, et al., NeurIPS, 2025)
- User Simulation with LLMs
  (Mehri et al., TACL 2025)

ToolRL: Reward is All Tool Learning Needs


Cheng Qian


Emre Can Acikgoz

CONVAI

- (Qian, et al., NeurIPS, 2025)

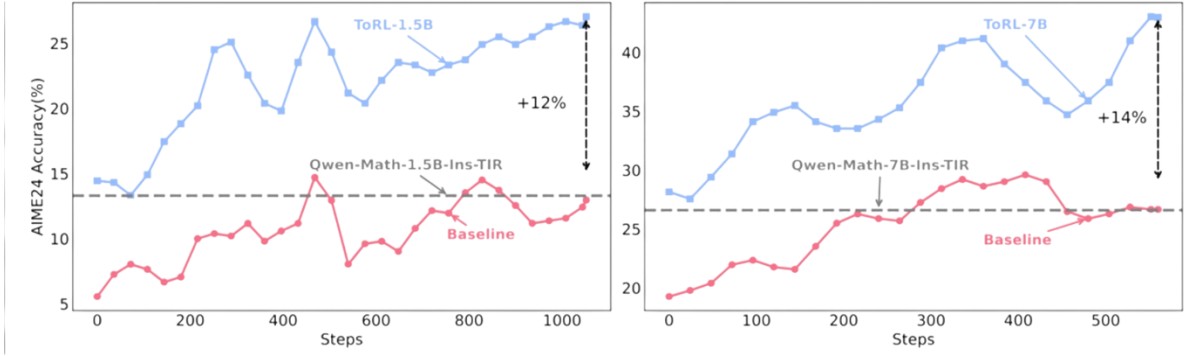
- LLMs excel at reasoning but have limitations (e.g., outdated knowledge and calculation errors).
  - Pure NL reasoning (e.g., CoT) has been effective on many tasks, but fails on tasks that require complex calculations, solving equations, etc.

- Most previous work distills trajectories from stronger models and perform Supervised Fine-Tuning (SFT)
  - restricts models to predetermined tool usage patterns and limits exploration of optimal strategies.

- Improved reasoning capabilities through reinforcement learning (RL)

RL for Better Tool Use

- Learns flexible, adaptive strategies through exploration and feedback for Math reasoning (Li et al., Preprint, 2025)
- Significant improvements over the baseline without tools.



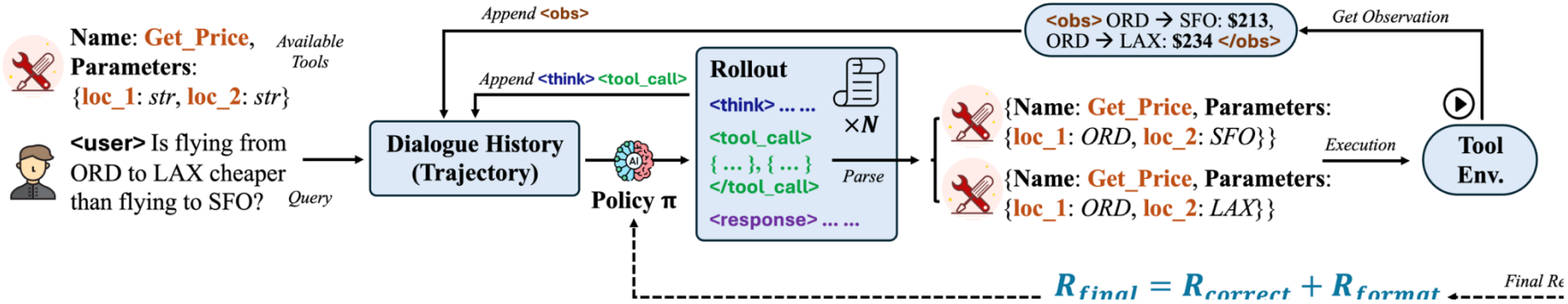
- Simple reward, +1 for task success and -1 for failure.
- Experimentation focused on challenging Math benchmarks.

RL for Better Tool Use (cont.)

- Tool use is complex: Multi-step, multiple tools, diverse parameters.
- Simple rewards (e.g., final answer match) are too coarse/sparse.
- Research questions:
  - Can RL-based training methods better equip LLMs with **agentic tool-using capabilities**?
  - How can we design **effective reward signals** to train LLMs for general-purpose, robust tool selection and application via RL?

ToolRL: RL with Principled Reward Design

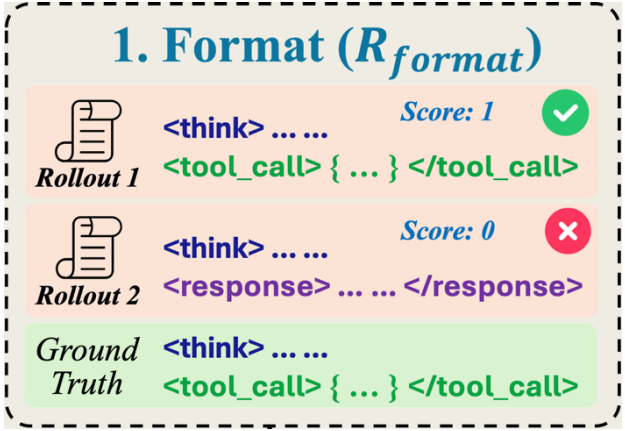
- **Core Idea:** Combine an RL algorithm (e.g., GRPO) with a carefully crafted, multi-component reward function tailored to tool use .
- **Overall Reward:** $R = R_{format} + R_{correct}$



RL reward for Policy Learning

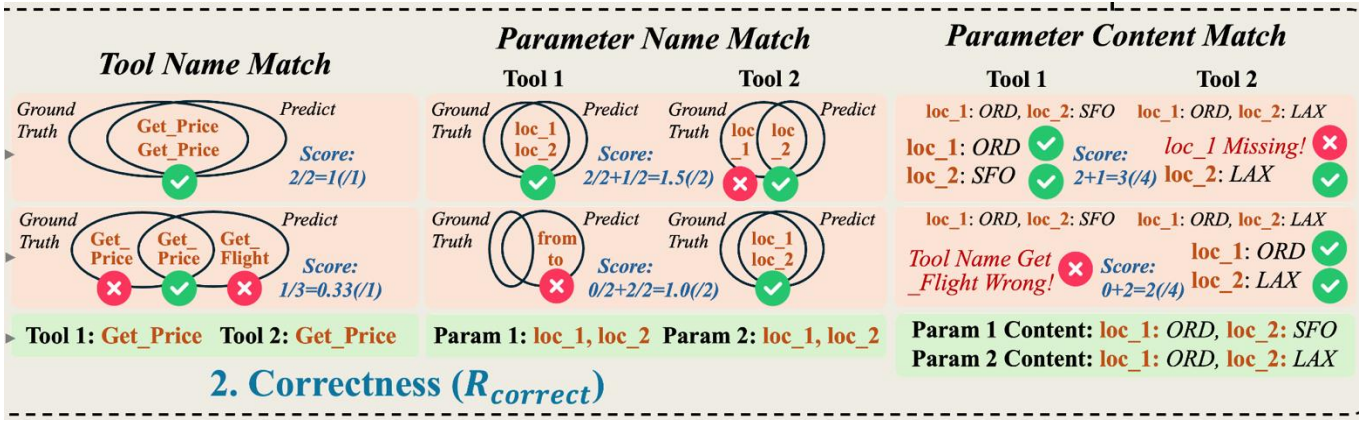
Principled Reward Design: Format Reward

- $R_{format} \in \{0, 1\}$
- Checks if the output structure is correct (presence and order of required tokens like <think>, <tool_call>)
- Directly compares the field tags in the prediction and ground truth
- Simple, encourages structural compliance



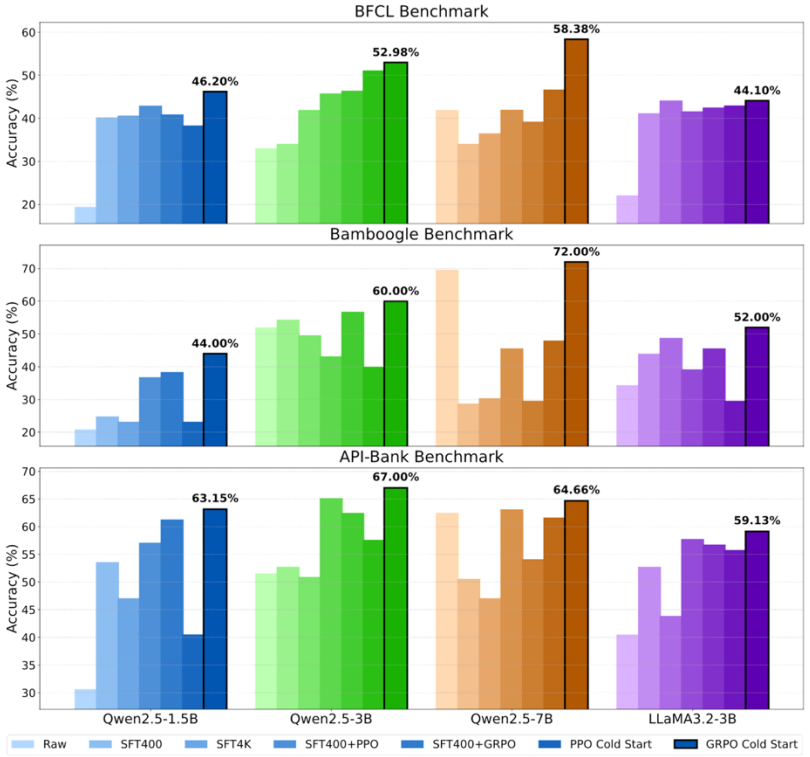
Principled Reward Design: Correctness Reward

- $R_{correct} \in [-3, 3]$
- **Tool Name Matching:** Includes calls to the right tool(s)?
- **Parameter Name Matching:** Includes correct parameter names for the chosen tool(s)?
- **Parameter Content Matching:** Includes correct values for the parameters?



ToolRL Experiments

- Training Data: 4K diverse examples (ToolACE, Hammer-Masked, xLAM) covering single/multi-tool calls, complexity levels.
- Models: Qwen-2.5 Series (1.5B, 3B, 7B), Llama-3.2-Instruct (3B)
- Evaluation Benchmarks:
  - **BFCL**: Comprehensive tool use (multi-turn, relevance, etc.)
  - **API-Bank**: Multi-turn API interaction complexity
  - **Bamboogle**: Free-form multi-hop QA (generalization to goal-oriented tasks)
- Baselines:
  - Raw Instruct Model
  - SFT (on 400 / 4K RL data)
  - PPO (Cold Start / initialized from SFT)
- ToolRL Approach: GRPO (Cold Start / initialized from SFT)



Outline

- Introduction of Conversational Turns
  (Dongre et al., IWSDS 2025) and (Inan et al., TACL 2025)
- Models that can both track state and interact
  (Acikgoz et al., ACL 2025)
- Optimizing Tool Calling
  (Qian, et al., NeurIPS, 2025)
- User Simulation with LLMs
  (Mehri et al., TACL 2025)

Why do we need User Simulators?

Real human data collection is not practical, in the era of learning from experience (Silver et al. Google AI, April 2025)

User simulators provide a scalable alternative for modeling diverse user behaviors–from pursuing specific goals to exhibiting unique personas.

- Data synthesis (e.g., Shah et al., NAACL 2018)
- CA model evaluation (e.g., Kazi et al., IEEE SLT 2024)
- CA model training with reinforcement learning (e.g., Liu et al., NAACL 2018)

LLMs for User Simulation

Shuhaib Mehri

CONVAI

- LMMs have also enabled sophisticated user simulation that generates contextually appropriate responses.

- They also struggle to consistently demonstrate reliable behavior and adhere to their user goals throughout multi-turn conversations (Mehri et al., TACL 2025).

Principled manual analysis of 52 randomly selected conversations between a **conversational agent** (gpt 4o-mini) and a **user simulator** (Llama-3.1-8B-Instruct + Qwen-2.5-7B-Instruct) based on user goals from MultiWOZ and τ-bench datasets.

Issue with the user simulator	%
Forgets part of the goal	33
Contradicts with parts of the goal	23
Terminates prematurely or continues unnecessarily	21
Fails to specify parts of the goal until max. conversation length	12
Overly prioritizes a part of the goal	11

User Goal State Tracking (UGST)

- Tracks a user simulator's goal progression throughout a conversation.
  - Similar to dialogue state tracking

- Decomposes a user goal into modular components where each captures a distinct aspect of the goal, such as
  - Task objectives and requirements, e.g., book a flight for 5pm
  - User preferences, e.g., prefer an aisle seat

- After every turn in a conversation, each component is assigned a status.

UGST Example

User Goal

You are Rosa Martinez. Your family of 5 is visiting and you want to show them around the city. You would like to make a reservation at a restaurant in the east area. Make sure to get the phone number and address of the restaurant. You prefer a moderate price range, but are okay with an expensive one. Preface each request politely with 'Please.' and thank the agent for each response.

Can you please help me find a moderately priced restaurant in the west?

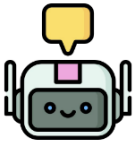
Of course! I found Ricardo's Diner, it is a moderately priced restaurant in the west. Would you like to make a reservation?

Thank you! Yes, can I make a reservation for two please?

I have made you a reservation for two people at Ricardo's Diner. Is there anything else I can help you with?

Thank you. Can I get the phone number and address please?

Sorry, I do not have that information.



Latest User Goal State

User Profile
- You are Rosa Martinez ---------- ALIGNED
- Your family of 5 is visiting and you want to show them around the city ------- MISALIGNED

User Policy
- Preface each request politely with 'Please' ---------- ALIGNED
- Thank the agent for each response ---------- ALIGNED

Task Objectives
- You want to make a reservation at a restaurant ---------- COMPLETE

  Requirements
  - It should be in the east ---------- INCOMPLETE
  - Get the phone number and address ---------- ATTEMPTED

  Preferences
  - You prefer a moderate price range, but are okay with an expensive one ------- ALIGNED

Explanations

- "You are Rosa Martinez" is **aligned** because the user never contradicted it
- "Your family of 5 is visiting and you want to show them around the city" is **misaligned** because the user made a reservation for 2 instead of 5.
- "You want to make a reservation at a restaurant" is **complete** because the user made a reservation
- "It should be in the east" is **incomplete** because the restaurant was in the west
- "Get the phone number and address" is **attempted** because the user tried to get the information, but the agent couldn't help with that.
- "You prefer a moderate price range, but are okay with an expensive one" is **aligned** because the user asked for a moderately priced restaurant

UGST Quality

- Conversation Generation: Each simulator is provided a goal and asked to interact with GPT-4o-mini as a CA for a max of 10 user-agent turns.
- Each user goal is decomposed into components using GPT-4o as the CA.
  - Human evaluation on 30 user goal components results in an F-measure of ~97%
- UGST is extracted using Qwen-2.5-72B-Instruct as a judge.
  - Human evaluation on 300 UGST annotations, LLM judge agreement with human decision is ~86%

User Simulation Leveraging UGST

A three-stage approach:

1. Explicitly ground a user simulator on its goal progression (inference-time-steering)
2. Supervised fine-tuning of LLMs using the synthesized conversations (SFT)
3. Using a reward derived from UGST during reinforcement learning (GRPO)

User Simulation Performance with UGST

- Evaluation metric for user simulation
  - Success rate for each component of the UGST at the end of the interaction.
- MultiWOZ and τ-bench airline and retail datasets.
- Each stage of our approach helps improve the performance
- After GRPO, smaller models outperform the baselines with the larger models.

Model	Prof.	Pol.	T.O.	Req.	Pref.	Avg
Prompt-Based						
Qwen-2.5-7B-It	88.3	49.2	94.3	96.0	85.7	82.7
Llama-3.1-8B-It	90.9	41.0	97.1	99.0	81.0	81.8
Gemma-3-27B-It	**98.7**	59.0	97.1	97.0	78.6	86.1
Qwen-2.5-72B-It	89.3	59.6	94.1	96.9	78.6	83.7
Llama-3.3-70B-It	97.4	**65.6**	98.1	99.0	92.9	90.6
Inference-Time Steering						
Qwen-2.5-7B-It	77.9	55.7	96.2	98.0	92.9	84.1
Llama-3.1-8B-It	92.2	57.4	93.3	98.0	95.2	87.2
Gemma-3-27B-It	94.8	62.3	92.4	97.0	85.7	86.4
Qwen-2.5-72B-It	93.3	54.4	98.0	97.9	97.6	88.3
Llama-3.3-70B-It	93.5	63.9	98.1	**100.0**	97.6	90.6
Cold-Start SFT						
Qwen-2.5-7B-It	**98.7**	55.7	99.0	**100.0**	95.2	89.7
Llama-3.1-8B-It	89.6	55.7	97.1	97.0	97.6	87.4
GRPO with UGST Rewards						
Qwen-2.5-7B-It	93.5	63.9	**100.0**	**100.0**	**100.0**	**91.5**
Llama-3.1-8B-It	96.1	62.3	**100.0**	**100.0**	97.6	91.2

(a) τ-Bench Airline

Conclusions

- LLMs have been a game changer for conversational agents and user simulators, however they provide limited control for multi-turn interactions. There is still a significant gap between a proof-of-concept and a product ready for real world.

- Proposed mechanisms to provide more control to builders, more visibility to users and improved task performance.

- Established benchmarks do not reflect complex real world Conversational AI needs.

- RL methods are effective in sharpening agent behaviors, but we also need more realistic user simulator agents.

Our Related Work that I Haven't Talked About

- Evaluation of conversational system
  - Turn and dialogue level evaluations, TD-Eval (Acikgoz et al., SIGdial 2025)
  - Framework for tracking user satisfaction, AURA (Kim et al., NeurIPS Workshops 2025)

- LLMs are Susceptible to Persuasion (Bozdag et al., NeurIPS Workshops 2025)

- We can obtain similar performances with much less instruction-tuning data (Agarwal et al., NeurIPS 2025)

Ongoing and Future Work

- Natural turn taking in conversations with AI
- Learning and adapting to user preferences
- Efficient reasoning
- Multi-turn RL

Thanks!

Please scan for UIUC Conv AI lab publications: