ToolkenGPT: Augmenting Frozen Language Models with Massive Tools via Tool Embeddings

Shibo Hao, Tianyang Liu, Zhen Wang, Zhiting Hu
{s5hao, til040, zhw085, zhh019}@ucsd.edu

Background

The original price of MacBook Air is $1580. Can you help me purchase it when it gets 10% off?

Sorry, but this is beyond my capabilities as a language model…

LLMs fail to help people with daily tasks, due to their functional limitations, e.g., accurate math calculation, updated world knowledge, taking real-world actions, etc.

Imagine if we can **connect LLMs with tools** seamlessly…

<multiply> (1580, 90%) → 1422

<price> (“MacBook Air”) → $1390

<purchase> (“MacBook Air”) → Success.

Previous works fine-tune LLMs or prompt LLMs (in-context learning) to call APIs.

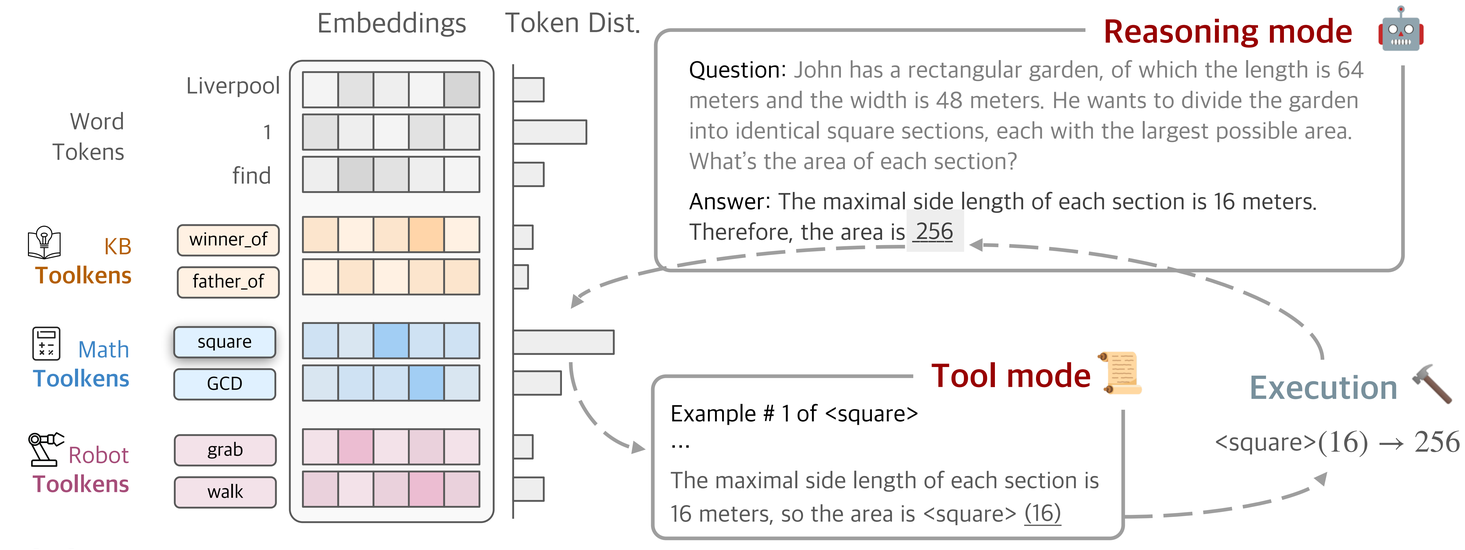
	FT	ICL
• **Frozen LMs**: No need for costly tuning	✗	✓
• **Massive Tools**: Works with a large tool set	✓	✗
• **Plug & Play**: Flexible to add / delete a tool	✗	✓
• **Accuracy** : Learn deep semantics of tools	✓	✗

Framework

Our approach **represents each tool as a token (“toolken”)** and learns an embedding for it.

Inference:

- Reasoning mode - the LLM predicts the next token, considering word tokens and plugged-in toolkens jointly
- Tool mode - Once a toolken is predicted, the LLM is prompted to complete the arguments using ICL.
- Tool Execution - The external tool processes the call, and the results are sent back to the reasoning mode.



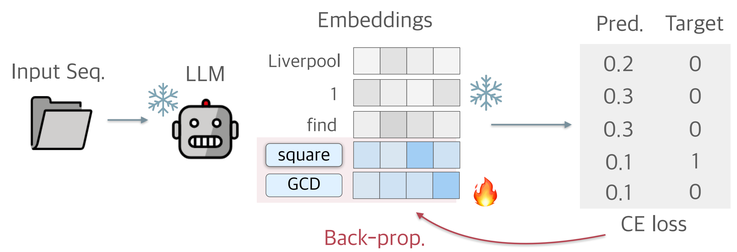
Training

Objective - Next token/toolken prediction

Input seq. s: The area is 2 5 6 square feet

Target seq. s': The area is <square> [mask] [mask] square feet

Inside one training step:



Disentangled Representation of tools → **Plug-and-play**

No gradients flow through LLM → **As fast as LLM inference**

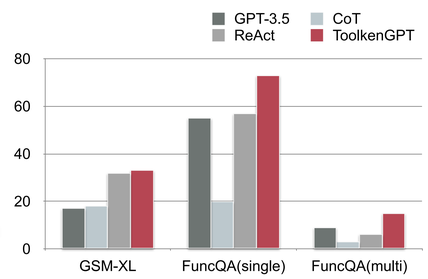
Data source: Demonstration / Synthetic data (self-instruct)

Experiments

1. Numeric Reasoning

w/ math operator

- Outperforms other tool learning baselines, especially better at **more complex math tools**.
- **Beats GPT-3.5** with LLaMA-33B

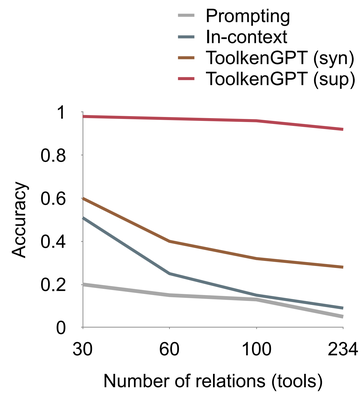


2. Knowledge-based Question Answering

w/ knowledge base

Question: Which team is the winner of 2005-06 FA CUP?
Answer: winner_of (2005-06 FA CUP) → Liverpool

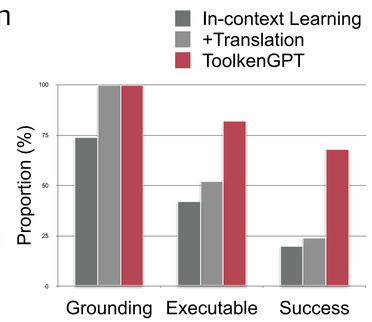
- ToolkenGPT with only **synthetic data** beats all baselines
- Scales to **> 200 tools**



3. Embodied Plan Generation

w/ robot action

- Naturally solved the **grounding** problem in embodied planning
- Higher success rate due to **deeper understandings** of tools learned from data



arXiv GitHub