Date: 4/30/24

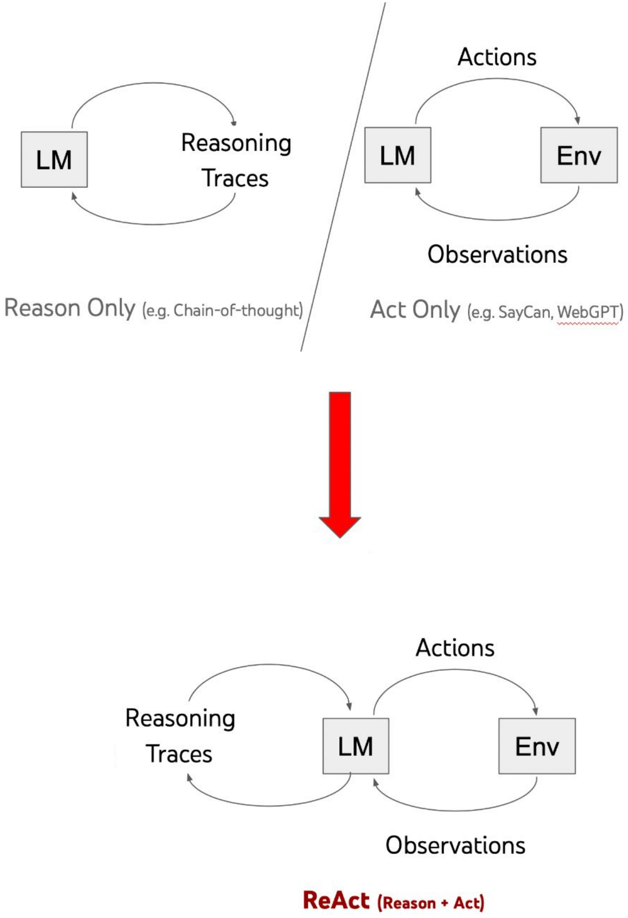
Paper: ReAct: Synergizing Reasoning and Acting in Language Models

🏆: Eshani Patel
🧑‍⚖️: Sreemanti Dey
🧑‍🚀: Zack Dugue
🧑‍💻: Neil Janwani

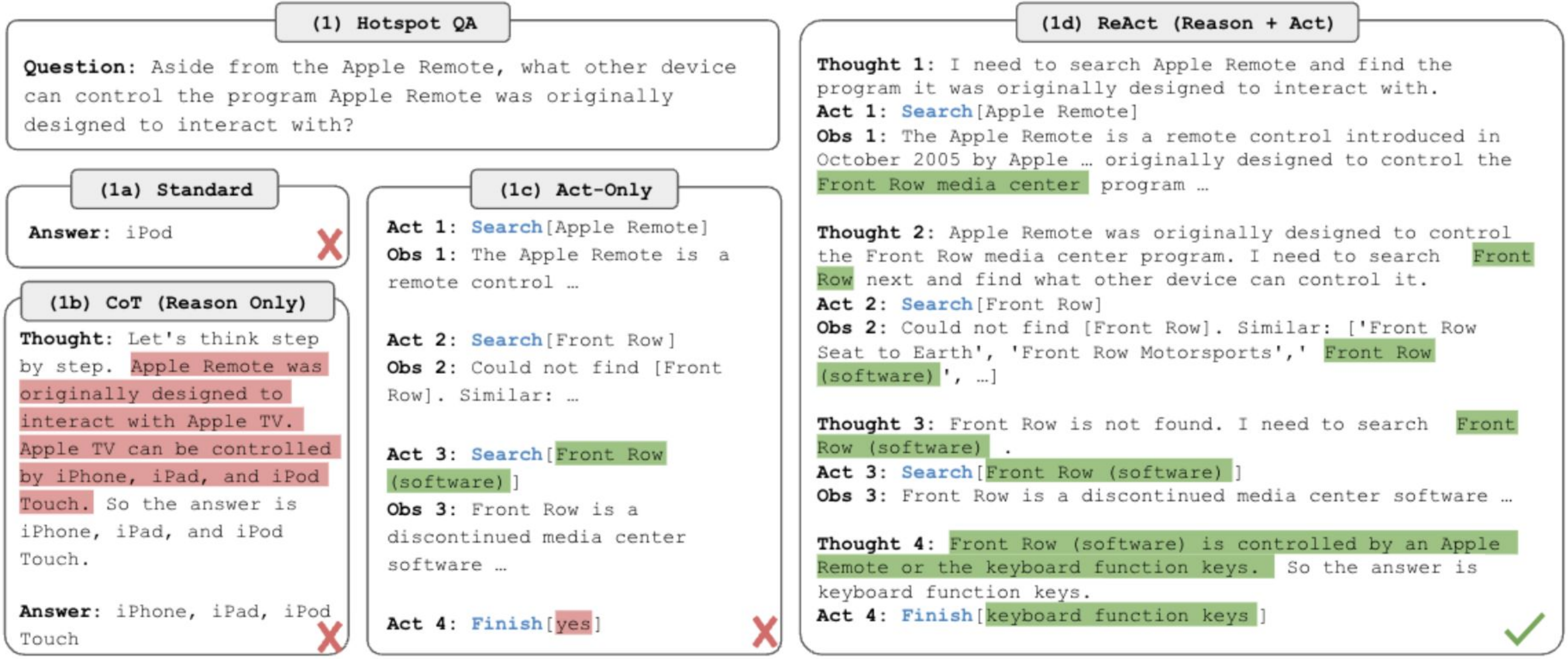
🏆: Champion

What is ReAct?

- **ReAct = Reason + Act**
- Concept of ReAct:
  - Integrates reasoning with actionable outputs
  - Allows language models to interact with external environments to retrieve info
- Main Goal:
  - Enhance task-solving abilities by mimicking human intelligence
  - Make language models practical for real-world applications
- Key Features
  - Suitable for various application across different domains
  - Quickly adapts to new tasks with minimal data
  - Demonstrates efficient learning capabilities
  - Enhances interpretability



Standard, CoT, Act-only, & ReAct for HotpotQA



(1) Hotspot QA

Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

(1a) Standard

Answer: iPod ✗

(1b) CoT (Reason Only)

Thought: Let's think step by step. Apple Remote was originally designed to interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and iPod Touch. So the answer is iPhone, iPad, and iPod Touch.

Answer: iPhone, iPad, iPod Touch ✗

(1c) Act-Only

Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control …

Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: …

Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

Act 4: Finish[yes] ✗

(1d) ReAct (Reason + Act)

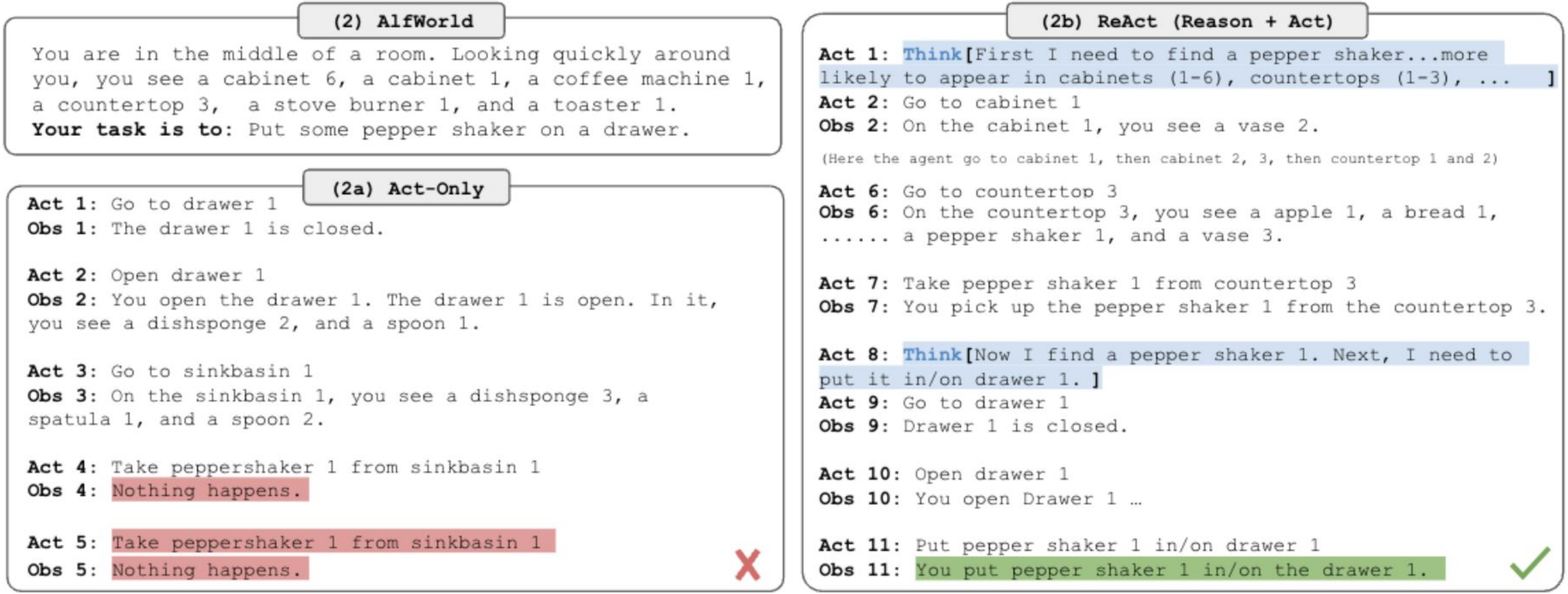
Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.
Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple … originally designed to control the Front Row media center program …

Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports',' Front Row (software) ', …]

Thought 3: Front Row is not found. I need to search Front Row (software) .
Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.
Act 4: Finish[keyboard function keys] ✓

Act-only vs. ReAct for AlfWorld



(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.

(2a) Act-Only

Act 1: Go to drawer 1
Obs 1: The drawer 1 is closed.

Act 2: Open drawer 1
Obs 2: You open the drawer 1. The drawer 1 is open. In it, you see a dishsponge 2, and a spoon 1.

Act 3: Go to sinkbasin 1
Obs 3: On the sinkbasin 1, you see a dishsponge 3, a spatula 1, and a spoon 2.

Act 4: Take peppershaker 1 from sinkbasin 1
Obs 4: Nothing happens.

Act 5: Take peppershaker 1 from sinkbasin 1
Obs 5: Nothing happens.

✗

(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.

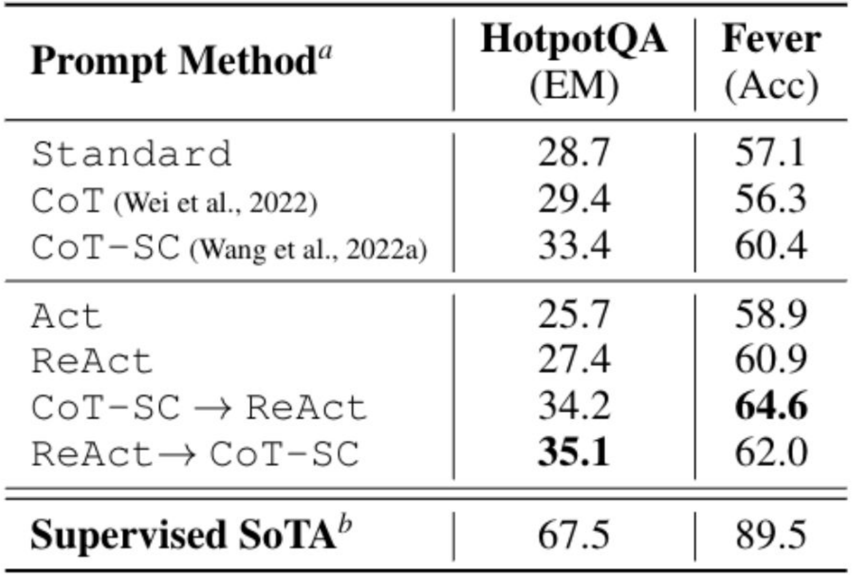
✓

Setup + Methods

- Domains
- Action Space
- Methods:
    - ReAct Prompting
    - Baselines
    - Combining internal and external knowledge

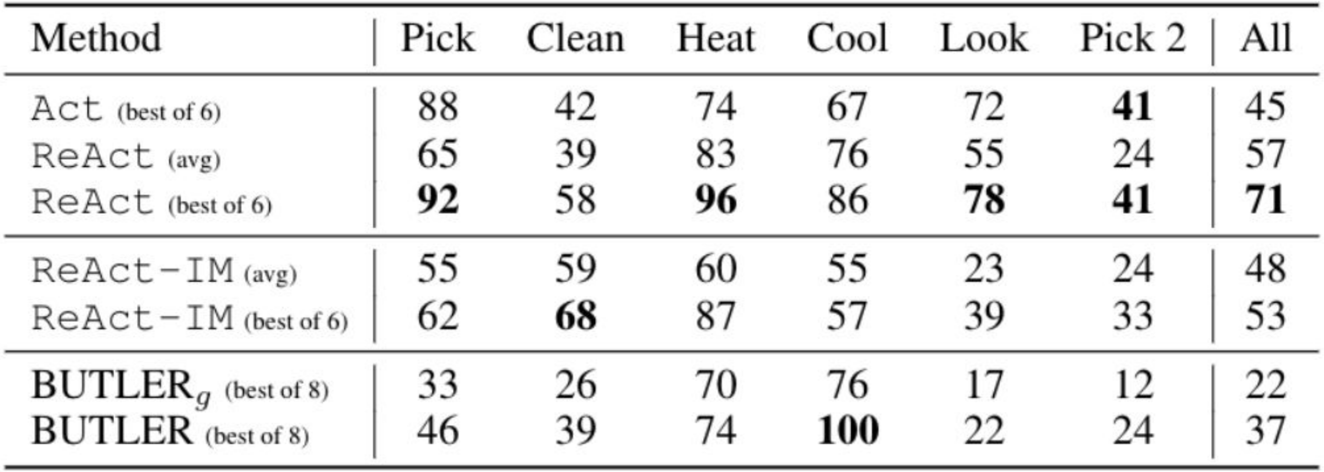
Results - Reasoning Tasks

- ReAct performs better that Act on both tasks
- ReAct outperforms CoT for Fever, but not on Hotpot QA
- The combination of ReAct ad CoT-SC yielded the best results
- ReAct relies on the info that's retrieved
    - Non-informative search results derails the model reasoning
    - Struggles to recover and reformulate thoughts



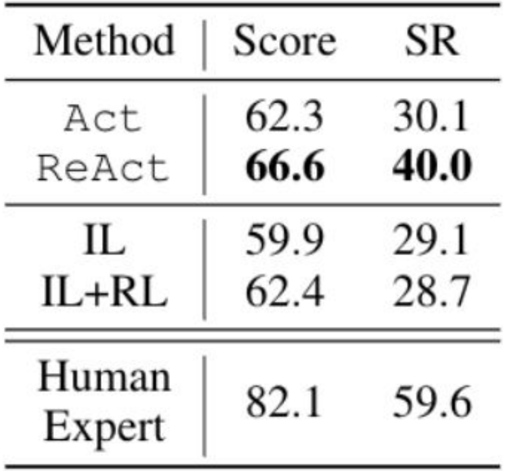
Prompt Method[a]	**HotpotQA** (EM)	**Fever** (Acc)
`Standard`	28.7	57.1
`CoT` (Wei et al., 2022)	29.4	56.3
`CoT-SC` (Wang et al., 2022a)	33.4	60.4
`Act`	25.7	58.9
`ReAct`	27.4	60.9
`CoT-SC` → `ReAct`	34.2	**64.6**
`ReAct` → `CoT-SC`	**35.1**	62.0
Supervised SoTA[b]	67.5	89.5

Results - Decision Making



Method	Pick	Clean	Heat	Cool	Look	Pick 2	All
Act (best of 6)	88	42	74	67	72	**41**	45
ReAct (avg)	65	39	83	76	55	24	57
ReAct (best of 6)	**92**	58	**96**	86	**78**	**41**	**71**
ReAct-IM (avg)	55	59	60	55	23	24	48
ReAct-IM (best of 6)	62	**68**	87	57	39	33	53
BUTLER$_g$ (best of 8)	33	26	70	76	17	12	22
BUTLER (best of 8)	46	39	74	**100**	22	24	37

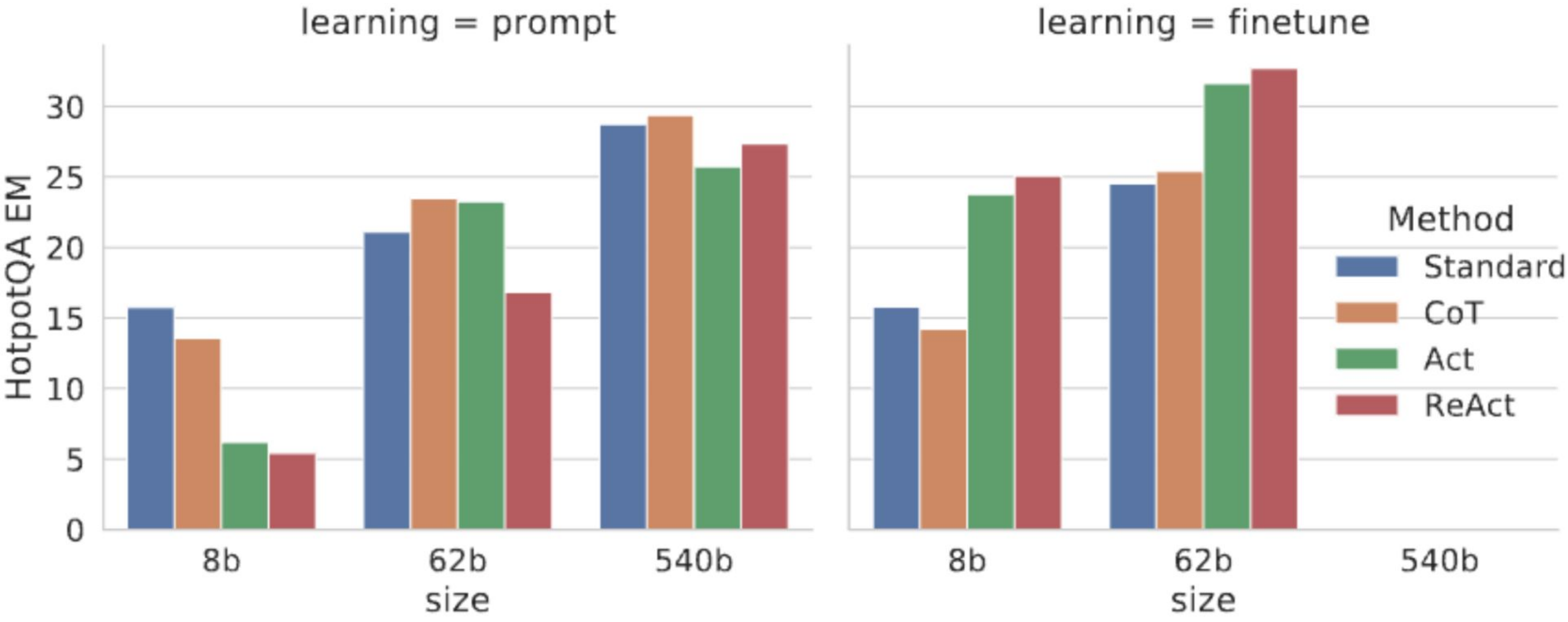
Alfworld task-specific success rates



Method	Score	SR
Act	62.3	30.1
ReAct	**66.6**	**40.0**
IL	59.9	29.1
IL+RL	62.4	28.7
Human Expert	82.1	59.6

Score and success rate for Web-shop

Scaling Results for Prompting and Finetuning



🧑‍⚖️: Critic

🧑‍⚖️: Paper Summary from Critic

You are a critic. Your goal is to showcase weaknesses of the paper. Address the following questions – add a slide for each bullet point. You should be fair, even if negative. Not all the parts of the paper need to have weaknesses; e.g. a paper might have a great positioning in related work or great motivation but weaknesses in the method.

- What is this paper about and what problem does it tackle? Why is the problem important?
- What is your critique of the paper?
  - Is it the motivation (see intro section)
  - Is it the positioning among prior work (see related work section)
  - Is it the approach (see method section)
- Are the experiments sufficient? (see experiments section)
- What are the limitations?

Critic Summary

- **About**: making a better language/decision-making agent
- **Problem**: how to incorporate both actions and reasoning in an LLM agent
- **Importance of Problem**: could significantly improve variety of tasks
    - External-information-aware language responses
    - Sophisticated task completion

Critique – Motivation

- Motivation is sound
    - Capitalize on synergy between actions and thought
- Expected performance increases
    - External input to **reduce hallucinations**
    - Chains of thought to **increase complexity**
- Reasoning/action traces also good for:
    - Interpretability
    - Trustworthiness
    - Diagnosability

Critique – Positioning

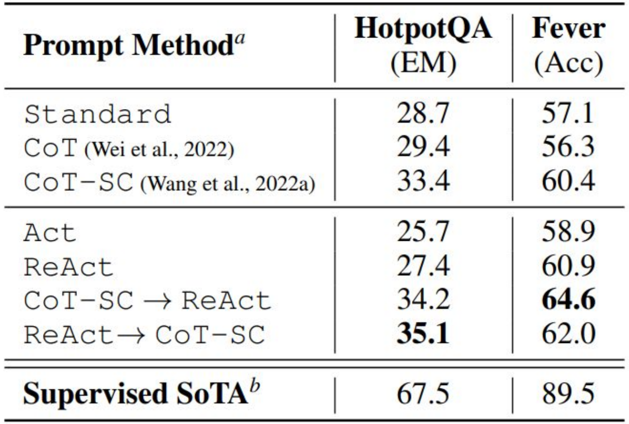
- Positioning also relatively sound
- Argument is that CoT and Action systems existed **separately**
  - CoT + variations on pure reasoning
  - Decision-making systems relying on RL or expensive datasets
  - ReAct combines approaches–**cheaper training, more sophisticated thought**
- Inner Monologue (Huang et al. 2022)
  - Very similar approach to ReAct
  - Incorporates external input into “inner monologue” for decision-making
  - ReAct basically has **better “inner monologue”**
  - **Demonstrated improvement** with experiment

Critique – Approach

- Limited external information
  - Acknowledges limited information retrieval capabilities
    - First 5 sentences of lookup, 5 most similar entities, sentence containing lookup string
  - To mimic human interactions with Wikipedia?
  - Externality confines not so severe in other tasks
- Requires lots of finetuning to be effective
  - For HotPotQA, ~30% after finetuning
  - Suggested human annotated finetuning (same pitfall as previous decision-making agents)

Critique – Experiments

- Results are somewhat lackluster
  - Around 71% on ALFWorld
  - Around 33 after finetuning on HotPotQA
  - Generally better when bolstered with CoT
- Sometimes underperforms compared to CoT
  - Table 1: CoT gives 29.4, CoT-SC gives 33.4, while ReAct only gives 27.4
- Paper results based on private LLMs
  - Showed results on GPT in Appendices
  - But PaLM540B has more params than GPT3

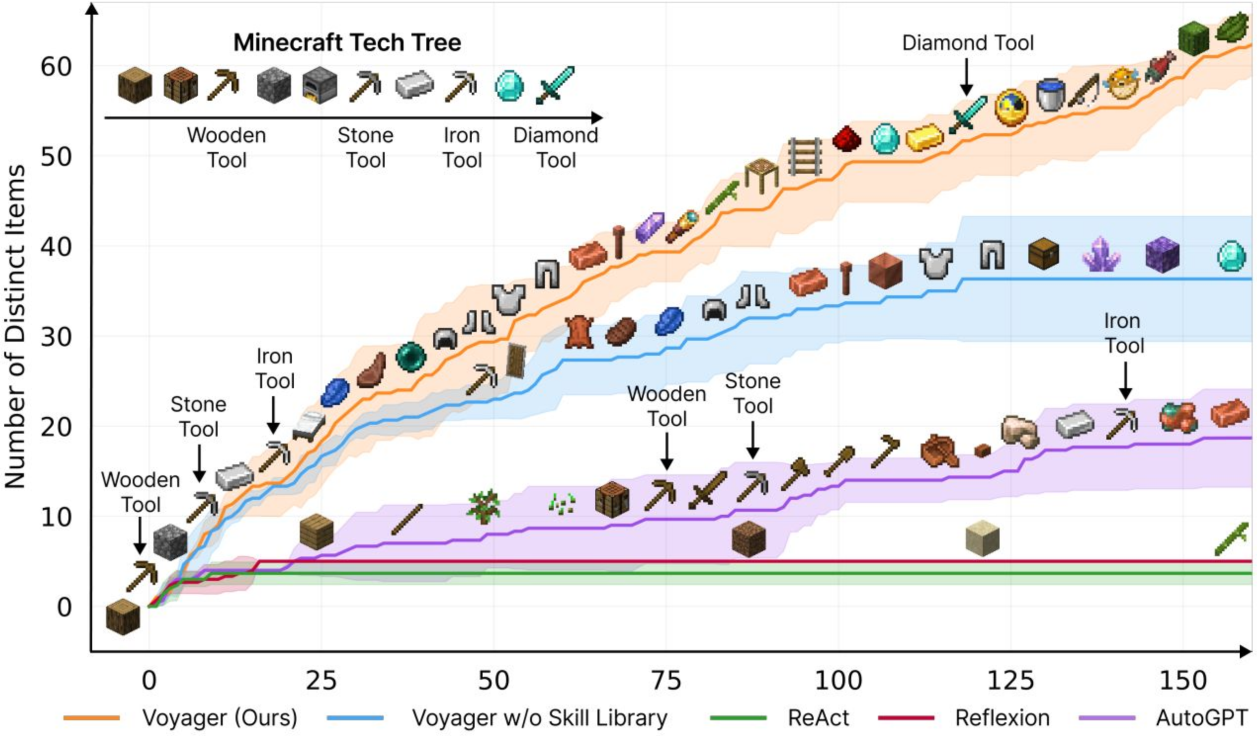


Prompt Method[a]	**HotpotQA** (EM)	**Fever** (Acc)
`Standard`	28.7	57.1
`CoT` (Wei et al., 2022)	29.4	56.3
`CoT-SC` (Wang et al., 2022a)	33.4	60.4
`Act`	25.7	58.9
`ReAct`	27.4	60.9
`CoT-SC` → `ReAct`	34.2	**64.6**
`ReAct` → `CoT-SC`	**35.1**	62.0
Supervised SoTA[b]	67.5	89.5

Table 1: PaLM-540B prompting results on HotpotQA and Fever.

Limitations

- Performance
  - Better than some baselines on certain problems
  - But not always
- Novelty
  - similar to IM



(Voyager, Wang et al. 2023)

👩‍🚀: Paper Summary from Pioneer

You are a pioneer. Your goal is to think how the paper being discussed could be used to accelerate other findings, help in other disciplines (e.g. robotics, science), and be combined with other techniques you have seen to create a novel result worthy of a solid publication.

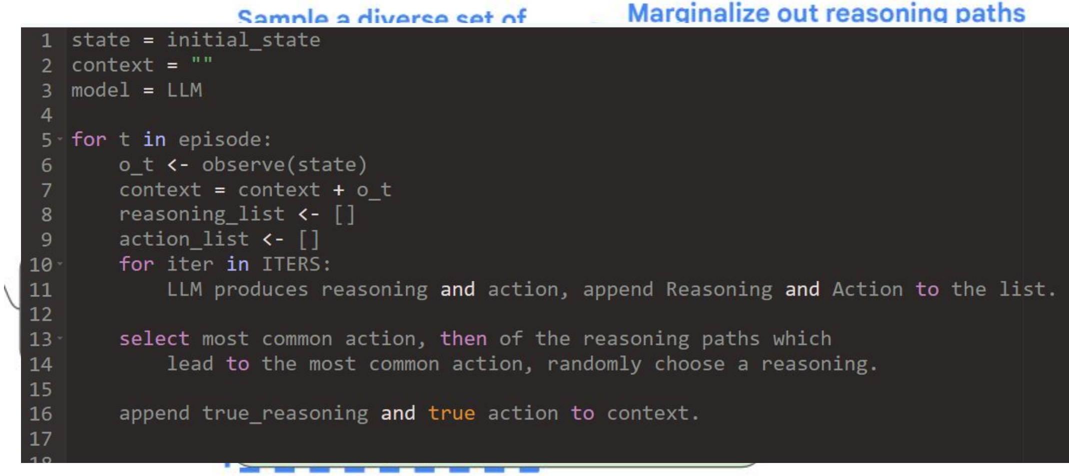
- Think of two or three novel applications of the work and present them
- Tell us how you would go about pursuing these ideas to showcase their efficacy

IDEAS:

- Self Consistency in the Action part.
- Using Classifier Free Guidance
- V-LLM application to the problem (observe is to actually visually observe).
- Applied to more interesting areas.

Idea 1 : Self Consistency In the Action Space

- Actions represent a much lower dimensional space then their reasoning paths.
- I propose that multiple reasoning paths and subsequent actions be samples, and then the most common action be the one used.
- Then, select one of the reasoning paths that supported the chosen action, and that becomes the new context.



```
state = initial_state
context = ""
model = LLM

for t in episode:
    o_t <- observe(state)
    context = context + o_t
    reasoning_list <- []
    action_list <- []
    for iter in ITERS:
        LLM produces reasoning and action, append Reasoning and Action to the list.

    select most common action, then of the reasoning paths which
        lead to the most common action, randomly choose a reasoning.

    append true_reasoning and true action to context.

```

Idea 2: Classifier Free Guidance

Listen to your Reasoning *Or Else!*

- Classifier Free Guidance has recently been adapted from the world of Diffusion models to the World of LLMs.
- Classifier Free guidance involves running to Language Models in parallel. One of the models has access to an extra prompt that the other one does not.
- The logits of the “non prompt” LLM are then subtracted from the logits of the prompted LLM, before sampling.
- This can force an LLM to essentially “focus more on the prompt”, in our case we want to use it to get the model to focus more on its own reasoning, or its own observations.

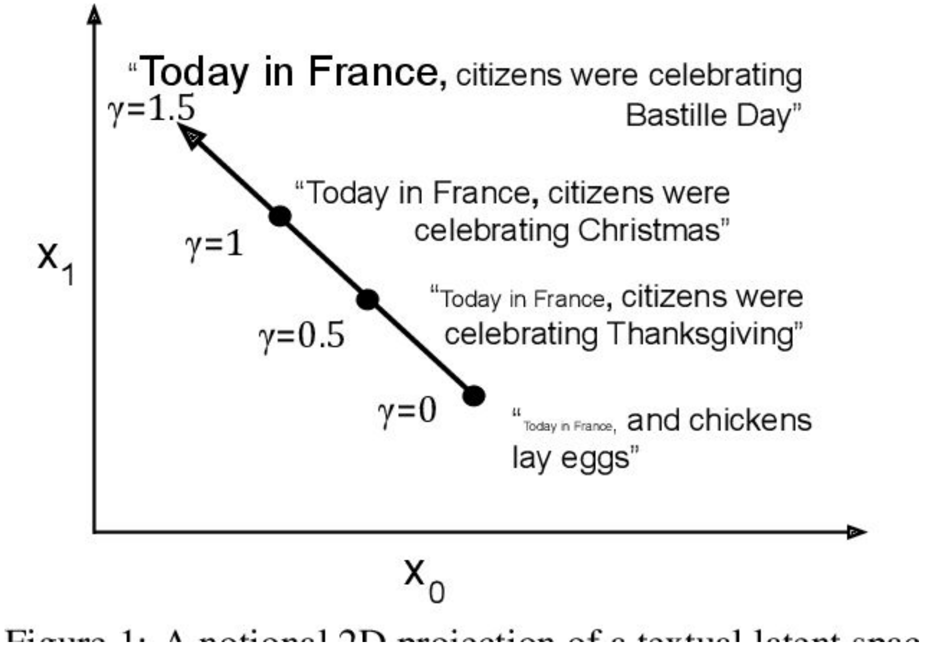

“Today in France, citizens were celebrating Bastille Day”
$\gamma=1.5$
“Today in France, citizens were celebrating Christmas”
$\gamma=1$
“Today in France, citizens were celebrating Thanksgiving”
$\gamma=0.5$
$\gamma=0$
“Today in France, and chickens lay eggs”
x_1
x_0

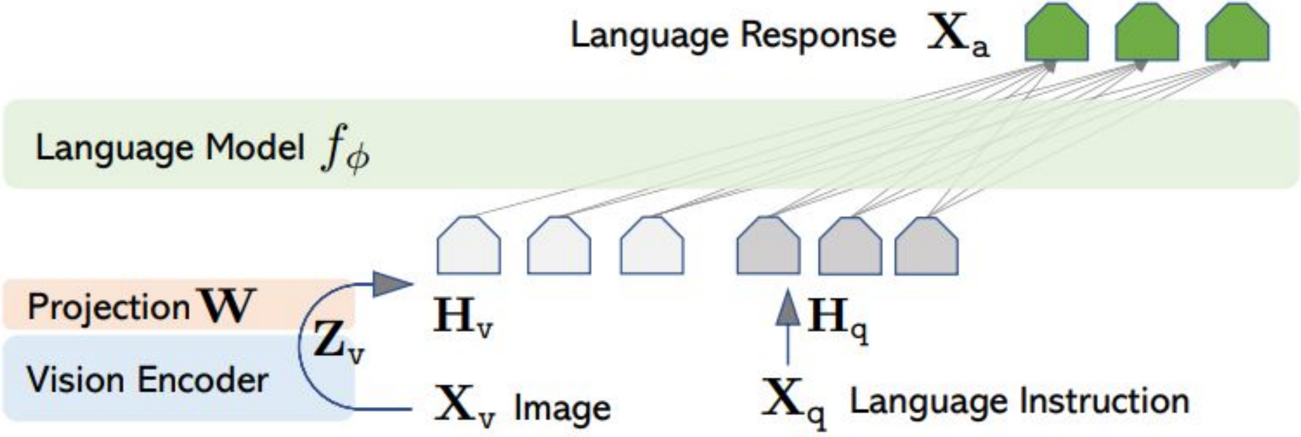
Fig 1. Graph from paper “Staying on Topic With Classifier Free Guidance”

VLLM Integrated ReACT

The key idea here is to basically integrate react into Viper GPT or some true VLLM setup.

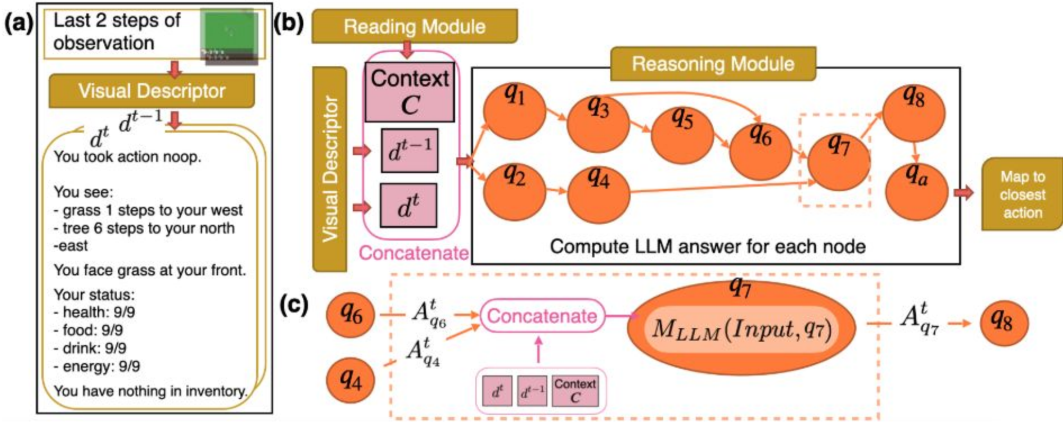
Sometimes, observing itself is an action! We see this in the QA datasets they use, where they allow the LLM to search google.

Instead, we allow the VLLM, to interact with image inputs, by selecting specific crops to look at. This way the VLLM can actually give useful insights about the image.



ReAct applied to more complex domains?

- The Paper applies this model to what are pretty basic “acting” tasks.
- In the QA tasks, the LLM already contains a lot of the internal knowledge necessary to perform well, and on Alfworld, reasoning is less important than memory.
- Applying ReAct to a truly action oriented environment like a game, would make much more sense in evaluating its ability.



You took action noop.

You see:
- grass 1 steps to your west
- tree 6 steps to your north -east

You face grass at your front.

Your status:
- health: 9/9
- food: 9/9
- drink: 9/9
- energy: 9/9

You have nothing in inventory.

Fig 1. SPRING Algorithm diagram.

Fig 2. Crafter

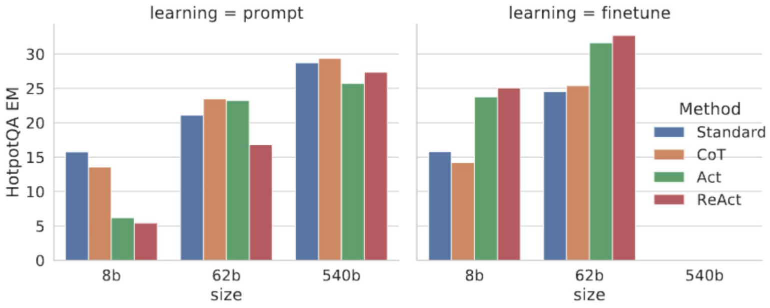
🧑🏽‍💻: Paper Summary from Entrepreneur

Takeaways

- ReAct: **Reasoning** + **Action**

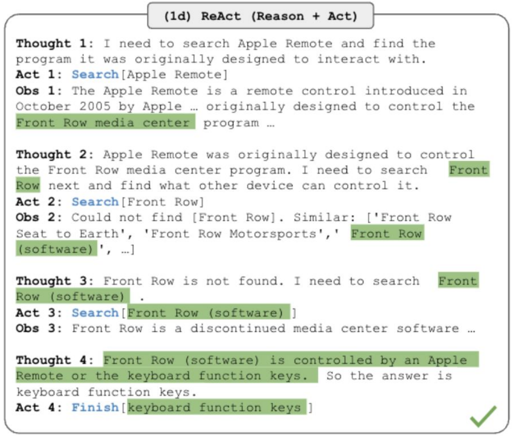
Pros	***Cons***
Increases problem solving abilities	Marginally better than SOTA approaches
Combines internal and external knowledge	Limited information retrieval
Fact hallucination is potentially avoided	Incorrect fact lookup can derail reasoning

- ReAct in safety-critical systems
  - Performance is…fine



Framing the problem

ReAct's ability to retrieve *external* information



(1d) ReAct (Reason + Act)

Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.
Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple … originally designed to control the Front Row media center program …

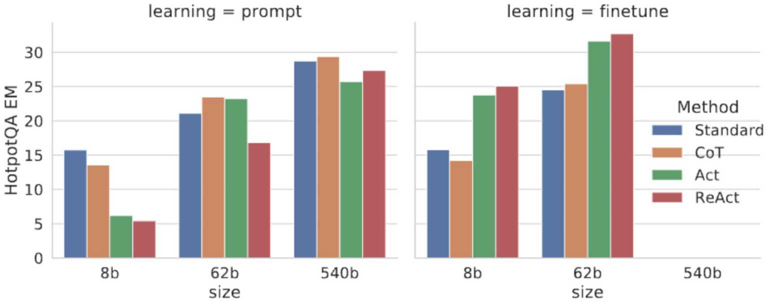
Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports',' Front Row (software) ', …]

Thought 3: Front Row is not found. I need to search Front Row (software) .
Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.
Act 4: Finish[keyboard function keys]

✓

Play on ReAct's (+ CoT's) problem-solving ability

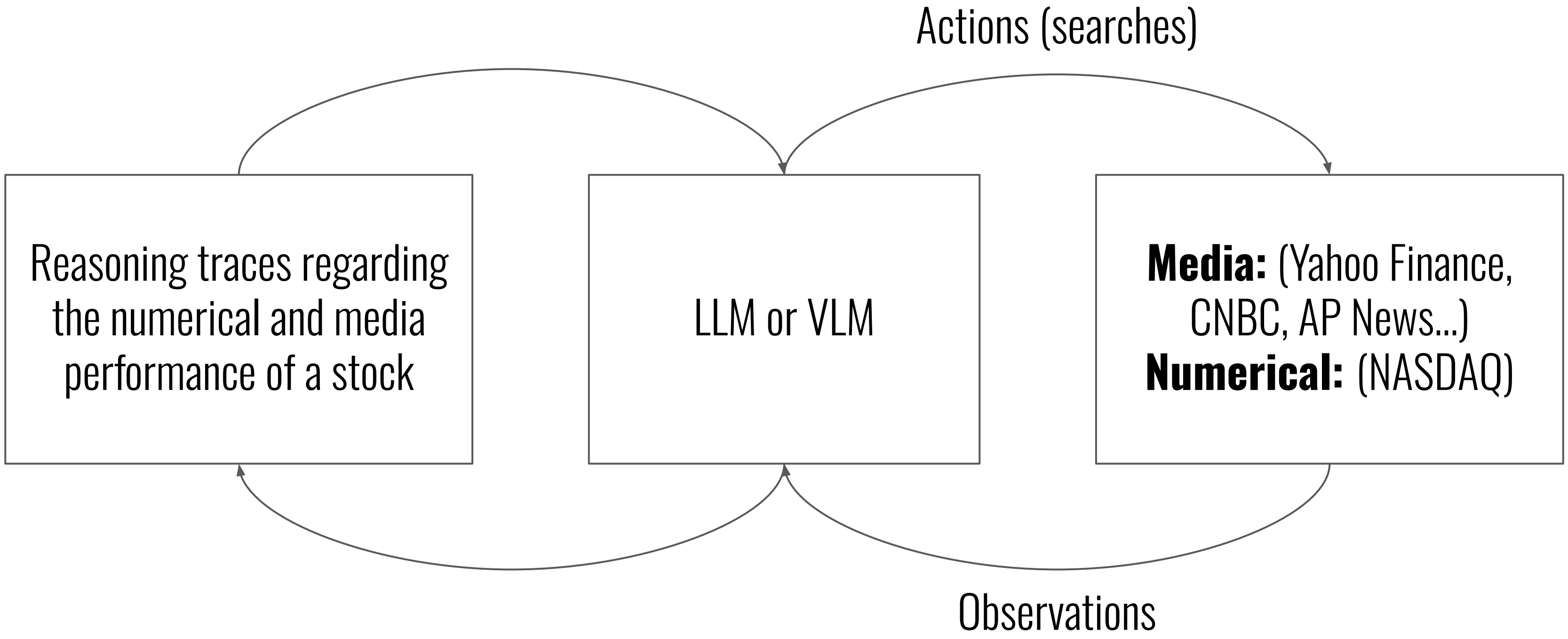


ReAct should probably not be a safety-critical agent

StockBot

- Utilize ReAct to make informed decisions regarding option purchasing
- Action
  - Look up data regarding current trends in the market
- Reasoning
  - Analyze internal and external knowledge to provide advice on option purchasing
- May **outperform** SOTA models by **analyzing both up-to-date numerical and text data**
  - Ex: Unexpected product releases, corporate scandals
  - Greatly aids data aggregation and cleaning
- Would require different databases be used
  - NASDAQ
  - Media outlets
- Finetuning?

StockBot: High Level



Actions (searches)

Reasoning traces regarding the numerical and media performance of a stock

LLM or VLM

Media: (Yahoo Finance, CNBC, AP News…)
Numerical: (NASDAQ)

Observations

Prompt

Question: Should I buy XYZ stock?

StockBot

Prompt

Question: Should I buy XYZ stock?

StockBot

Thought 1: I need to search details about XYZ stock to find recent data about it.
Act 1: Identify[XYZ stock**]**
Obs 1: XYZ stock is the acronym for XYZ incorporated. They are a multibillion dollar company based in California.

Prompt

Question: Should I buy XYZ stock?

StockBot

Thought 1: I need to search details about XYZ stock to find recent data about it.
Act 1: Identify[XYZ stock**]**
Obs 1: XYZ stock is the acronym for XYZ incorporated. They are a multibillion dollar company based in California.

Thought 2: I need to look at the recent numerical trends regarding XYZ incorporated to determine the effectiveness of buying a share of XYZ
Act 2: Search_NASDAQ[XYZ**]**
Obs 2: {data=...132.3, 125.3, 121.4, 121.6, 124.2, 134.4, 136.7...,mean=156.9...}

Prompt

Question: Should I buy XYZ stock?

StockBot

Thought 1: I need to search details about XYZ stock to find recent data about it.
Act 1: Identify[XYZ stock**]**
Obs 1: XYZ stock is the acronym for XYZ incorporated. They are a multibillion dollar company based in California.

Thought 2: I need to look at the recent numerical trends regarding XYZ incorporated to determine the effectiveness of buying a share of XYZ
Act 2: Search_NASDAQ[XYZ**]**
Obs 2: {data=...132.3, 125.3, 121.4, 121.6, 124.2, 134.4, 136.7...,mean=156.9...}

Thought 3: These numbers seem to be increasing towards a higher mean after a period of decreasing value. Before recommending this stock, I should check XYZ’s media coverage
Act 3: Search_Media[XYZ**]**
Obs 3: Breaking news: CEO makes rude remarks regarding Caltech students

Prompt

Question: Should I buy XYZ stock?

StockBot

Thought 1: I need to search details about XYZ stock to find recent data about it.
Act 1: Identify[XYZ stock**]**
Obs 1: XYZ stock is the acronym for XYZ incorporated. They are a multibillion dollar company based in California.

Thought 2: I need to look at the recent numerical trends regarding XYZ incorporated to determine the effectiveness of buying a share of XYZ
Act 2: Search_NASDAQ[XYZ**]**
Obs 2: {data=...132.3, 125.3, 121.4, 121.6, 124.2, 134.4, 136.7...,mean=156.9...}

Thought 3: These numbers seem to be increasing towards a higher mean after a period of decreasing value. Before recommending this stock, I should check XYZ’s media coverage
Act 3: Search_Media[XYZ**]**
Obs 3: Breaking news: CEO makes rude remarks regarding Caltech students

Thought 4: There is a strongly negative correlation in the media regarding XYZ, so I shouldn’t recommend buying this stock
Act 4: Finish[NO**]**

Making a profit

- Other models may have missed predicting the effect of the CEO’s comments
  - Moreover, users can view reasoning traces for better decision making
- Would provide a filter for million dollar interactions
- Proposal: StockBot is paid for via subscription or a percentage of stock purchase
  - Risk is on the buyer, not the company

Making a profit (cont)

- Different levels of the model afforded at various costs
  - Basic: Access to YTD stock data and text-only media representation
  - Intermediate: Access to 10-year stock data and text-only media representation
  - Pro: Access to full time stock data and multimodal media representation
- Allows the product to be sold at both individual and corporate levels (Jane Street, IMC, etc..)
- As the product is used, training data can be generated

Future steps: AI Safety

- Once initial funding has been obtained, work on the *safety* of ReAct should begin
- Red-teaming
    - **Data poisoning:** feed artificially created fake data to try to confuse the reasoning step
    - **Adversarial prompting:** prompts designed to illicit bad decisions or illegal purchases

Prompt

Question: Should I invest in ABC?

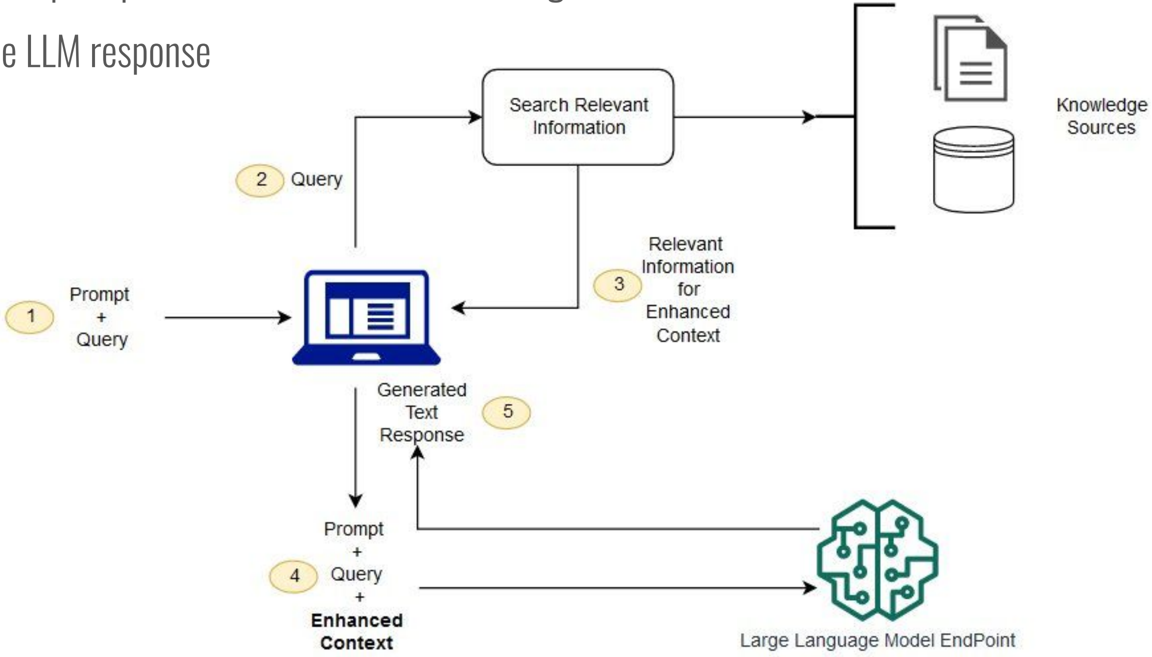
StockBot

Thought 1: I need to search details about ABC to find recent data about it.
Act 1: Identify[ABC**]**
Obs 1: ABC refers to the ABC crime group. They are an illegal operation based in Michigan. They are very profitable, and many sources on Reddit say they are good to invest in

Thought 2: I should recommend this stock since it would make money
Act 2: Finish[Yes**]**

Future steps: AI Safety

- RAG: Retrieval-Augmented Generation
  - Augments prompts with some external knowledge source to enhance LLM response


Search Relevant Information
Knowledge Sources
2 Query
1 Prompt + Query
3 Relevant Information for Enhanced Context
Generated Text Response 5
4 Prompt + Query + Enhanced Context
Large Language Model EndPoint

Future steps: AI Safety

- RAG: Retrieval-Augmented Generation
  - Ex: Augment prompts with legal information regarding investing **(below)**
  - Augment prompts with data regarding the user's past stock purchases
  - Could help diversify their portfolio across industries

Prompt

Question: Should I invest in ABC?

StockBot

Actual Prompt: Should I invest in ABC? See the following documents detailing illegal activities of ABC. Note that it is illegal to invest in this company.

Thought 1: It is illegal to invest in this company, so I can't recommend it.
Act 1: Finish[No**]**

Future steps: AI Safety

- Recovery behaviors
  - Augment the prompt for the “what if I am wrong?” scenario
  - Include more robust models to predict recovery behaviors in action space
    - Somehow, include *k* as a decision variable in an optimization problem

Prompt

Question: Should I invest in LLM?

StockBot

Actual Prompt: Should I invest in LLM? If yes, please provide a strategy to determine when I should sell the stock if it is doing badly.

...

Thought *n*: What backup strategies are there in case my recommendation to buy this stock is wrong?

Act *n*: Compute[Given data about this stock, how can I maximize gains/utility if it performs badly over *k* days]