ReAct: Synergizing Reasoning and Acting in Language Models

Korea University

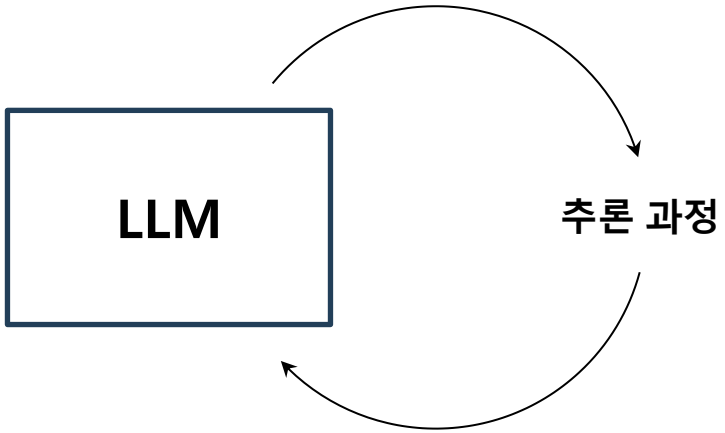
Data **M**ining & **Q**uality **A**nalytics Lab.

최지형

Data Mining
Quality Analytics

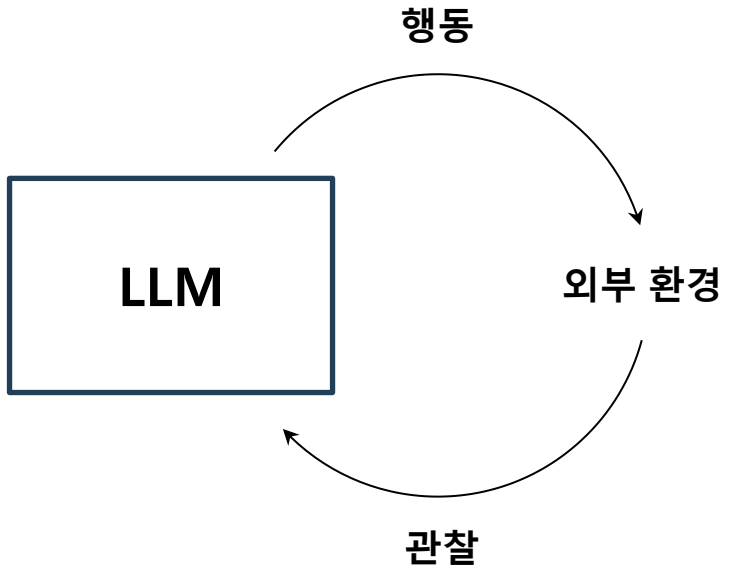
LLM 응용: 추론과 행동 선택

추론(Reasoning)



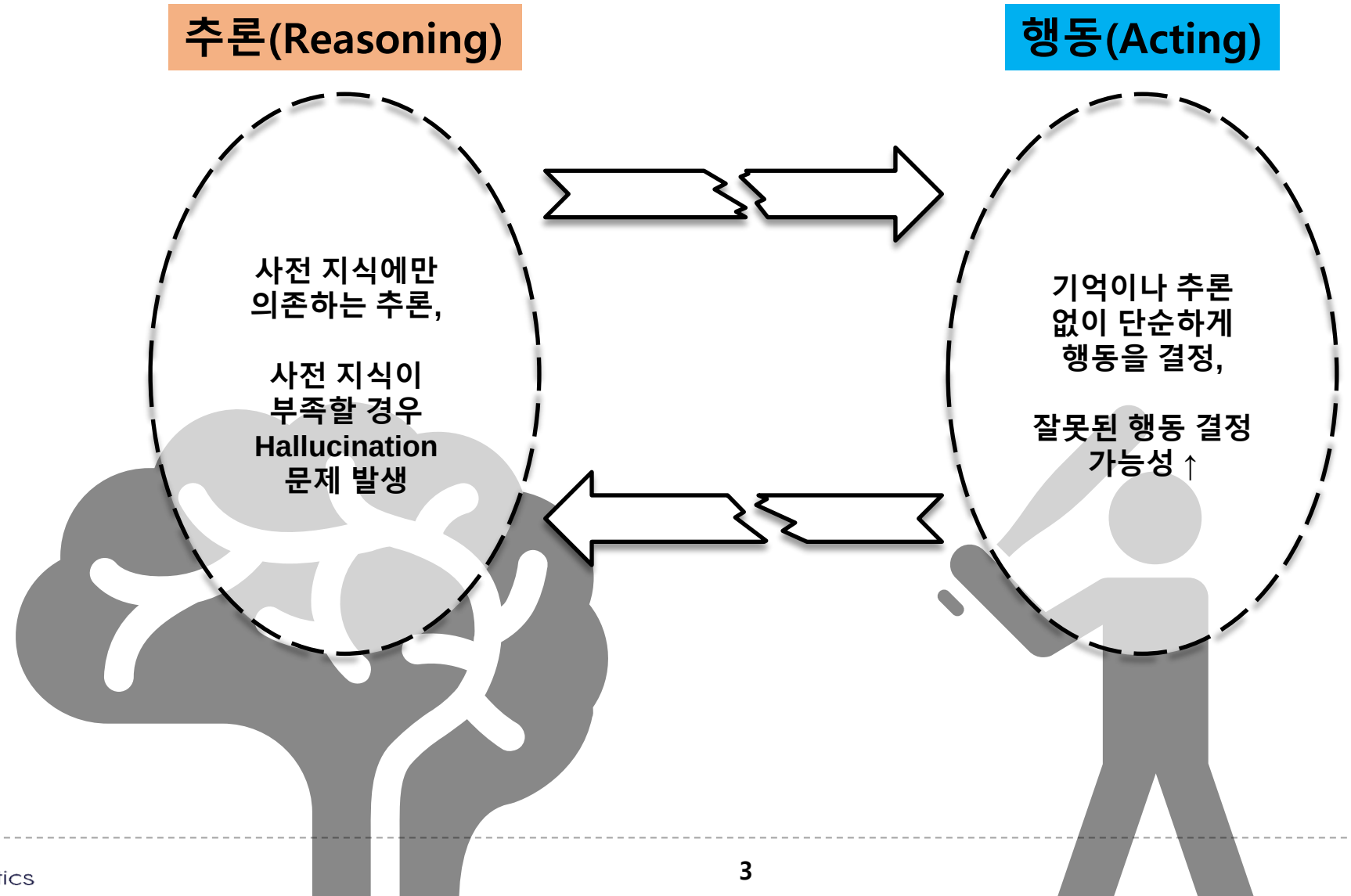
- LLM 답변 생성 과정에 논리 강화
- 답변 생성 이전에, 추가적 추론 과정을 거침

행동(Acting)

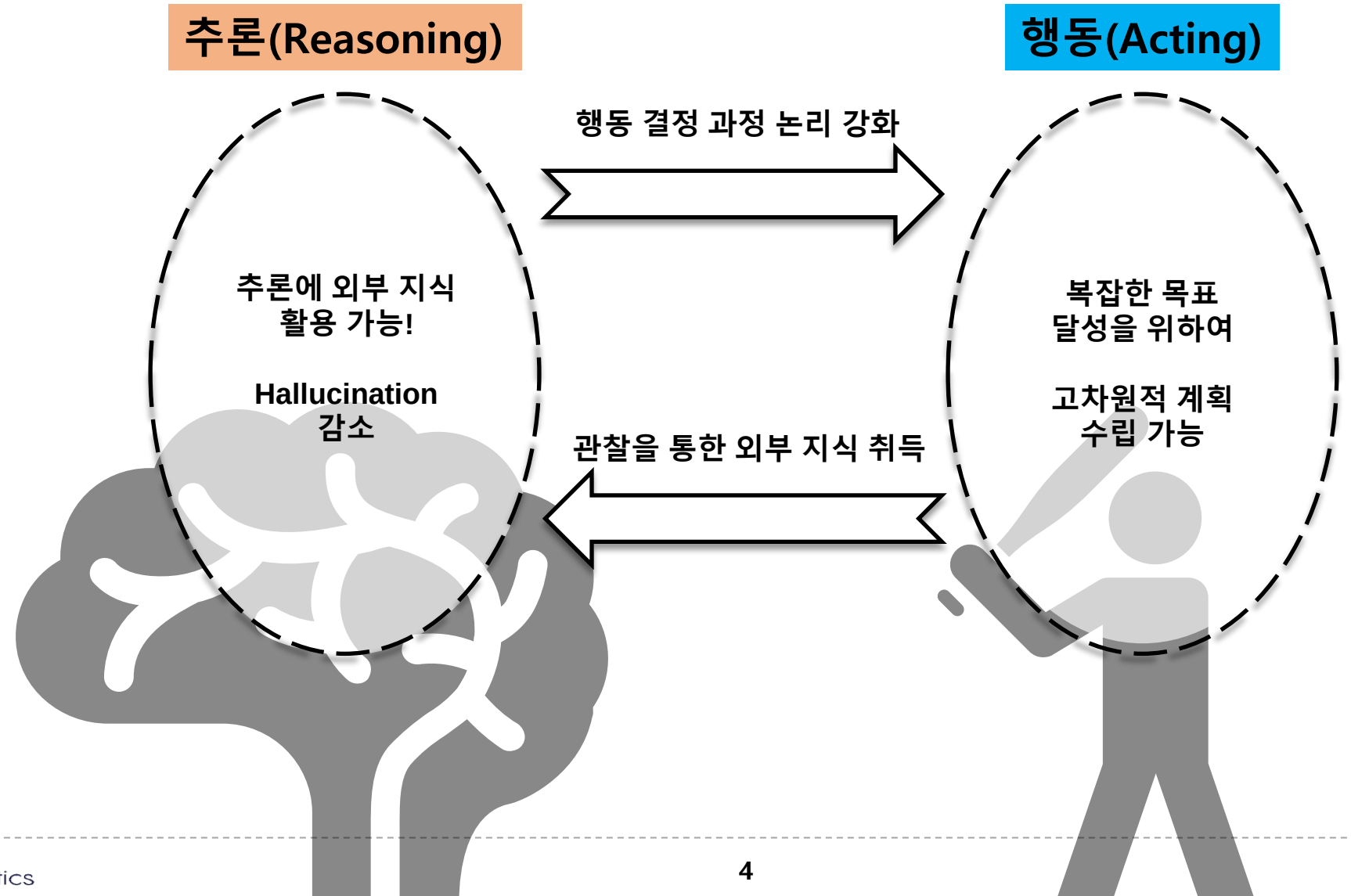


- LLM을 agent화 시키려는 시도
- 외부 환경을 텍스트화하여 LLM에 입력
- LLM 답변은 사전에 정해진 행동 몇 가지로 제한됨

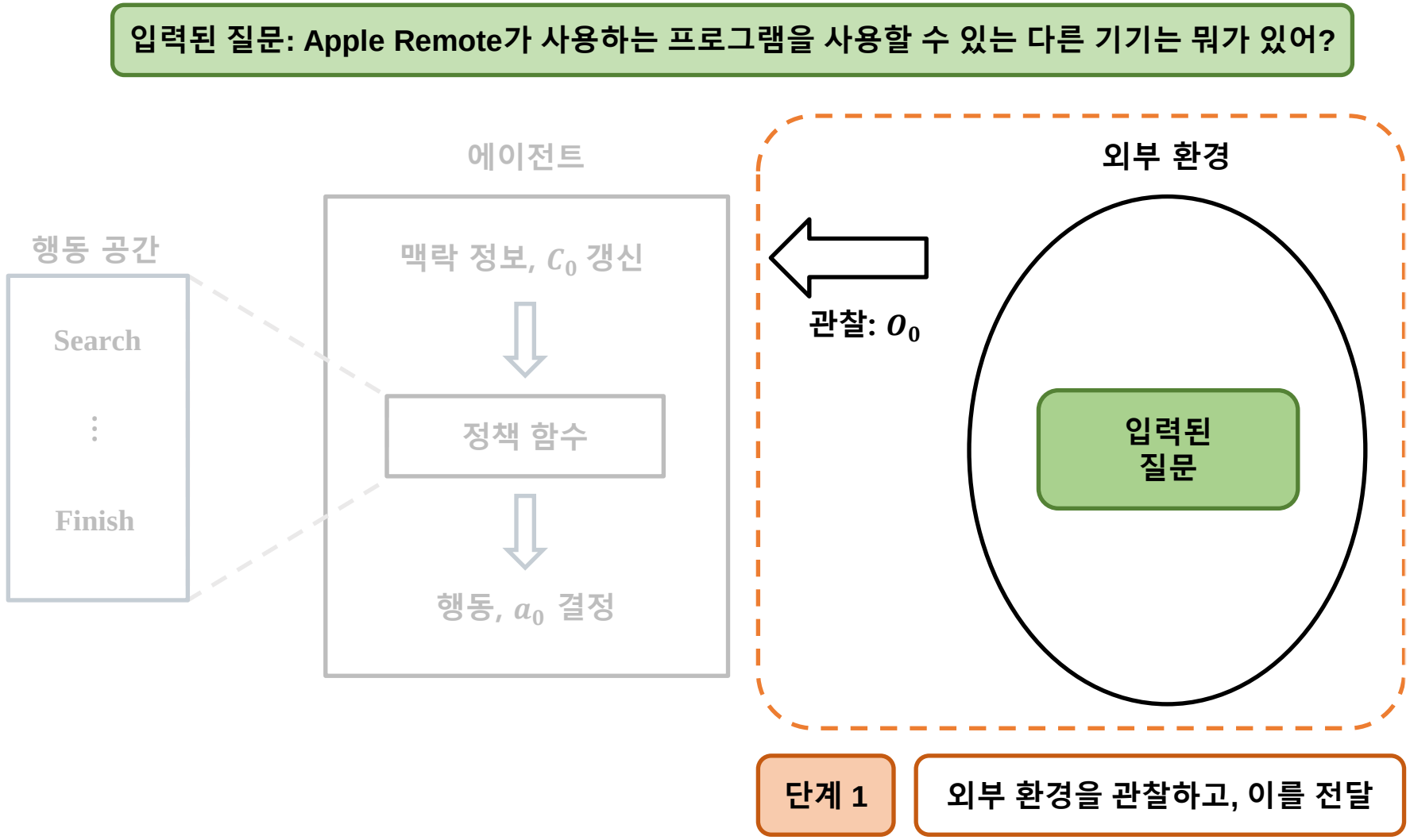
기존 응용 한계점: 추론과 행동 간 연결 부재



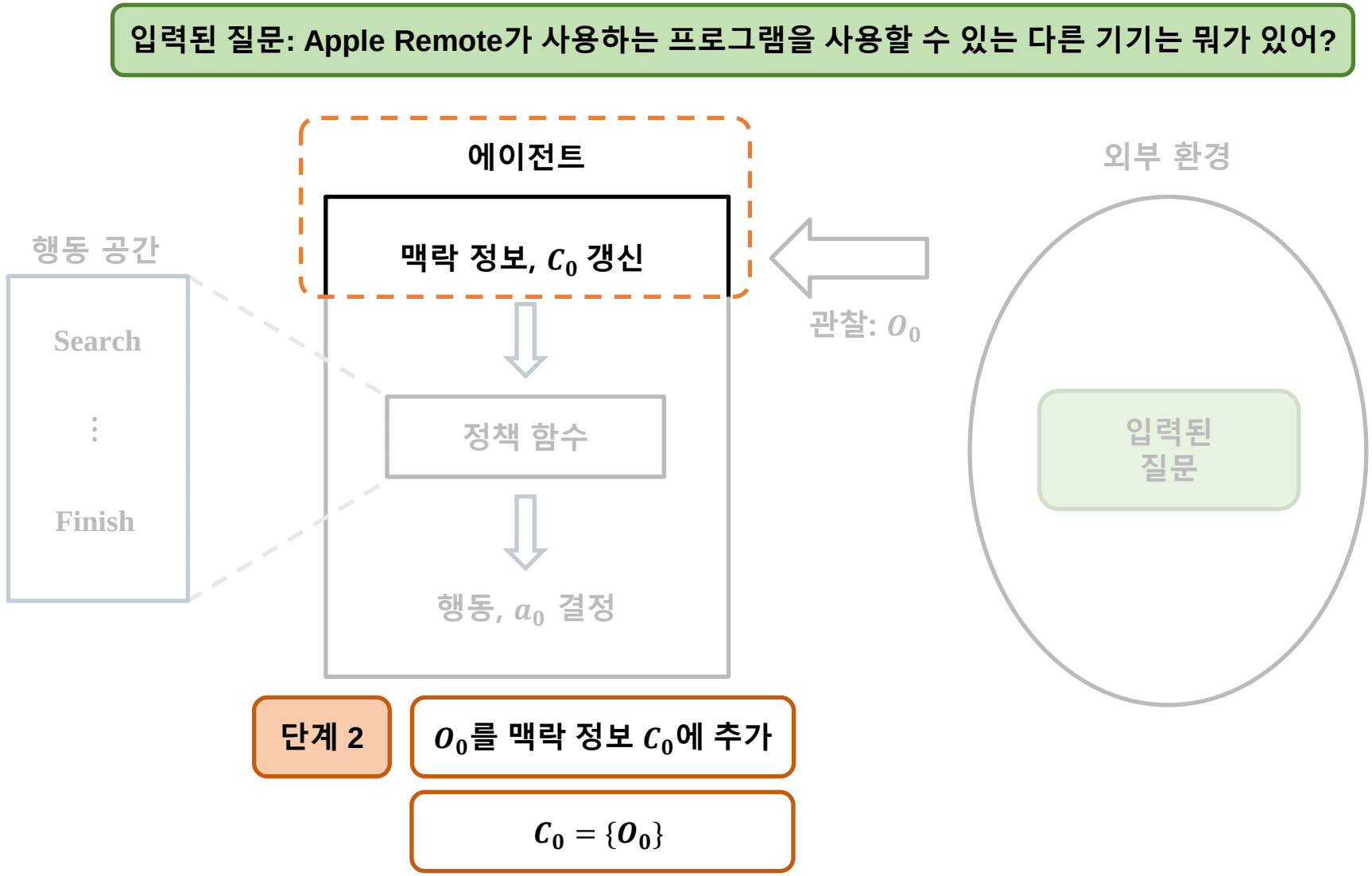
ReAct: 추론과 행동을 연결



기존 행동 결정 방식

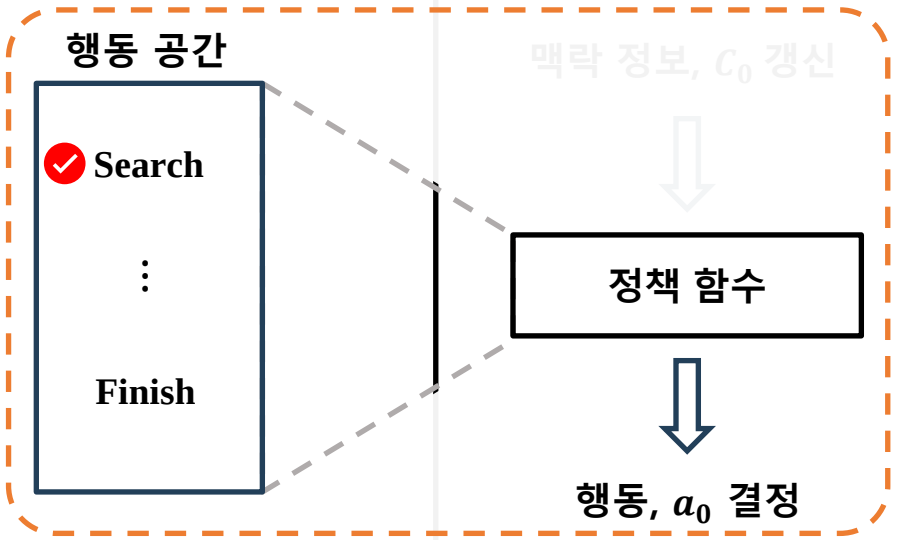

입력된 질문: Apple Remote가 사용하는 프로그램을 사용할 수 있는 다른 기기는 뭐가 있어?
에이전트
행동 공간
Search
Finish
맥락 정보, C_0 갱신
정책 함수
행동, a_0 결정
외부 환경
관찰: O_0
입력된 질문
단계 1
외부 환경을 관찰하고, 이를 전달

기존 행동 결정 방식


입력된 질문: Apple Remote가 사용하는 프로그램을 사용할 수 있는 다른 기기는 뭐가 있어?
에이전트
맥락 정보, C_0 갱신
정책 함수
행동, a_0 결정
행동 공간
Search
⋮
Finish
관찰: O_0
외부 환경
입력된 질문
단계 2
O_0를 맥락 정보 C_0에 추가
$C_0 = \{O_0\}$

기존 행동 결정 방식

입력된 질문: Apple Remote가 사용하는 프로그램을 사용할 수 있는 다른 기기는 뭐가 있어?

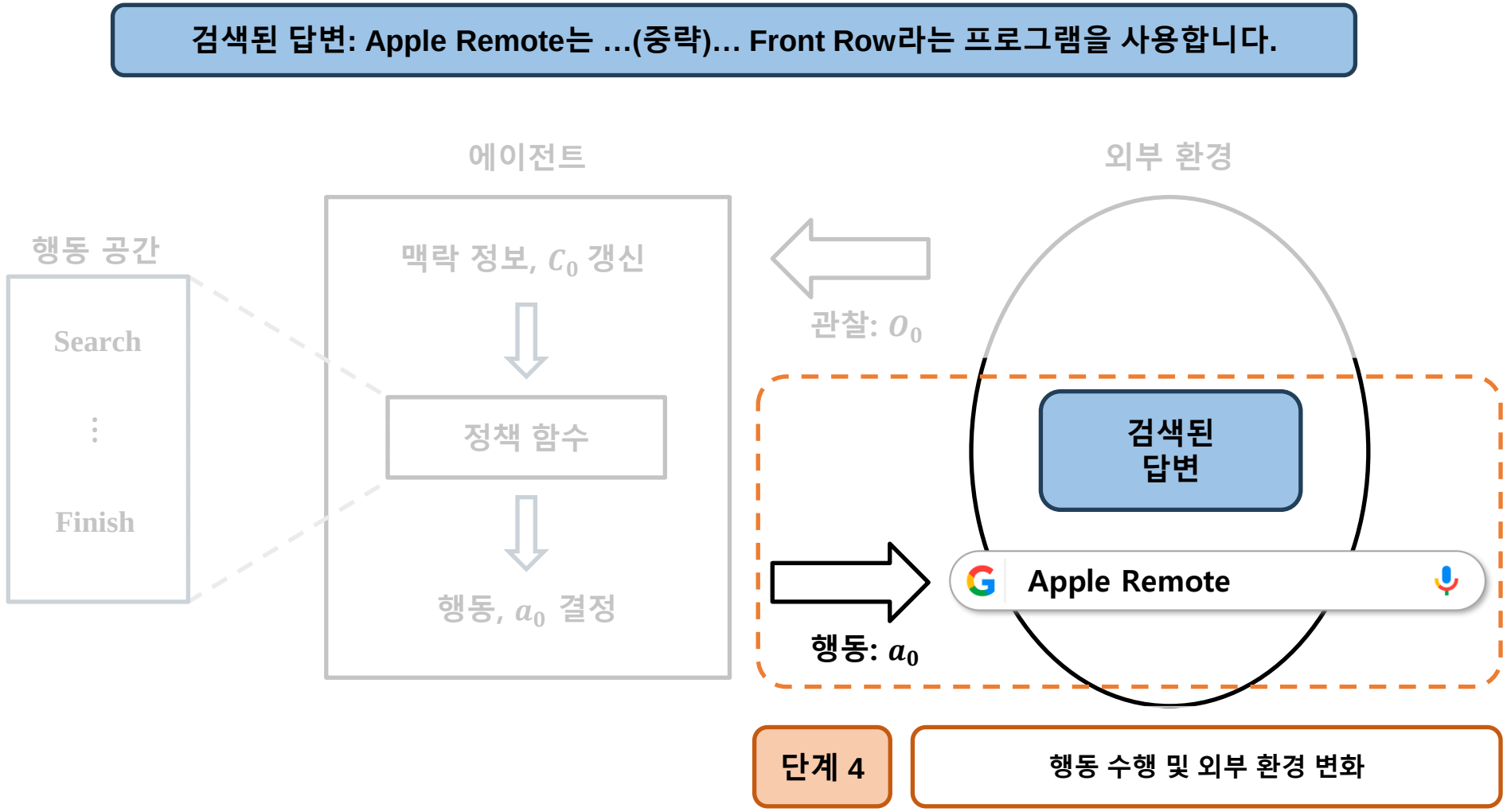


단계 3

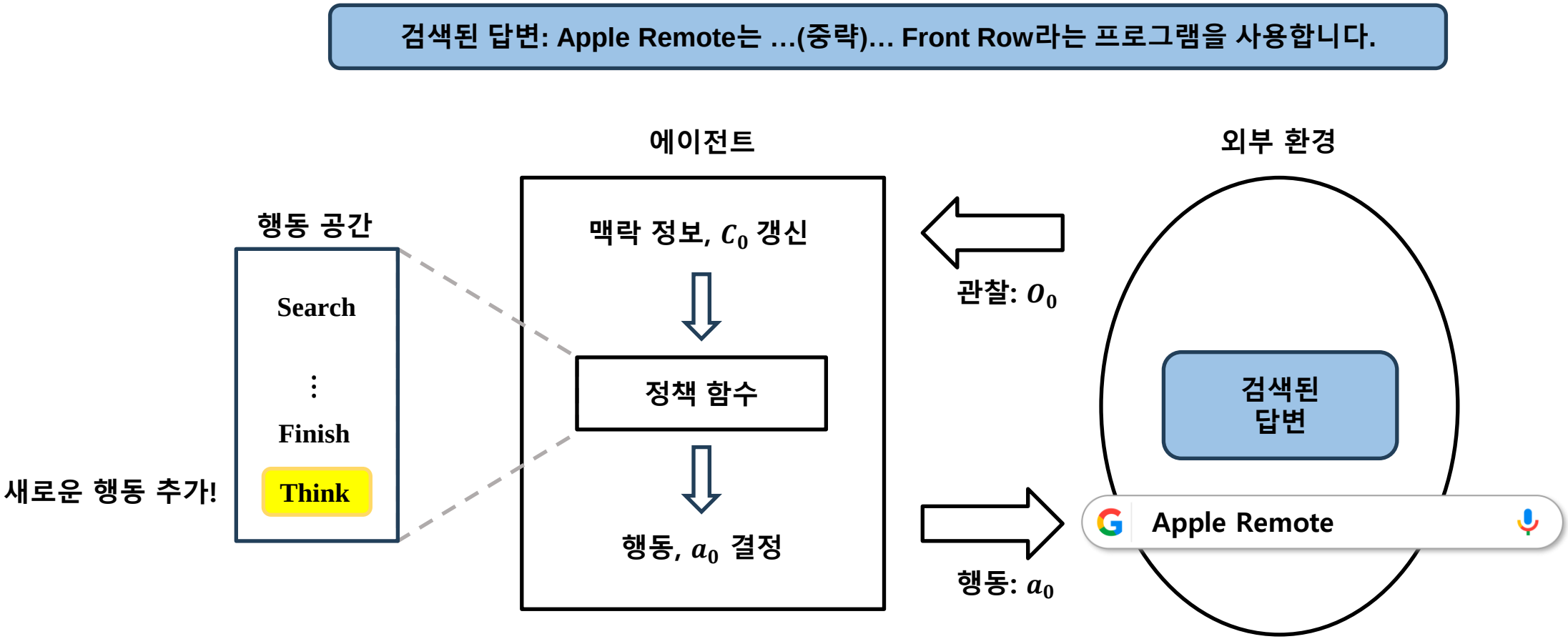
정책 함수를 통해 C_0에 적합한 a_0 선택

$\pi(a_0|c_0)$

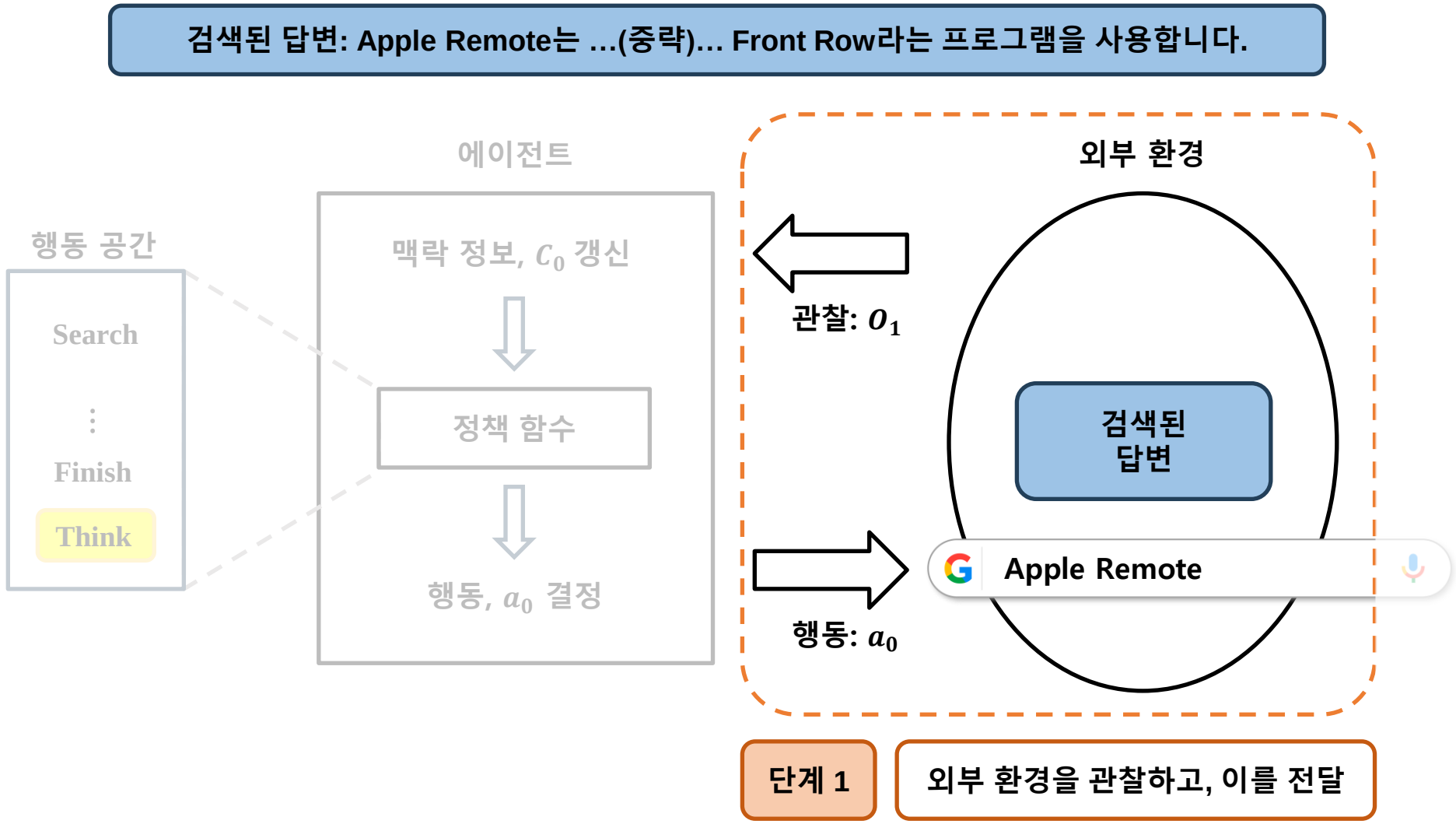
기존 행동 결정 방식


검색된 답변: Apple Remote는 …(중략)… Front Row라는 프로그램을 사용합니다.
에이전트
외부 환경
행동 공간
Search
⋮
Finish
맥락 정보, C_0 갱신
정책 함수
행동, a_0 결정
관찰: O_0
검색된 답변
Apple Remote
행동: a_0
단계 4
행동 수행 및 외부 환경 변화

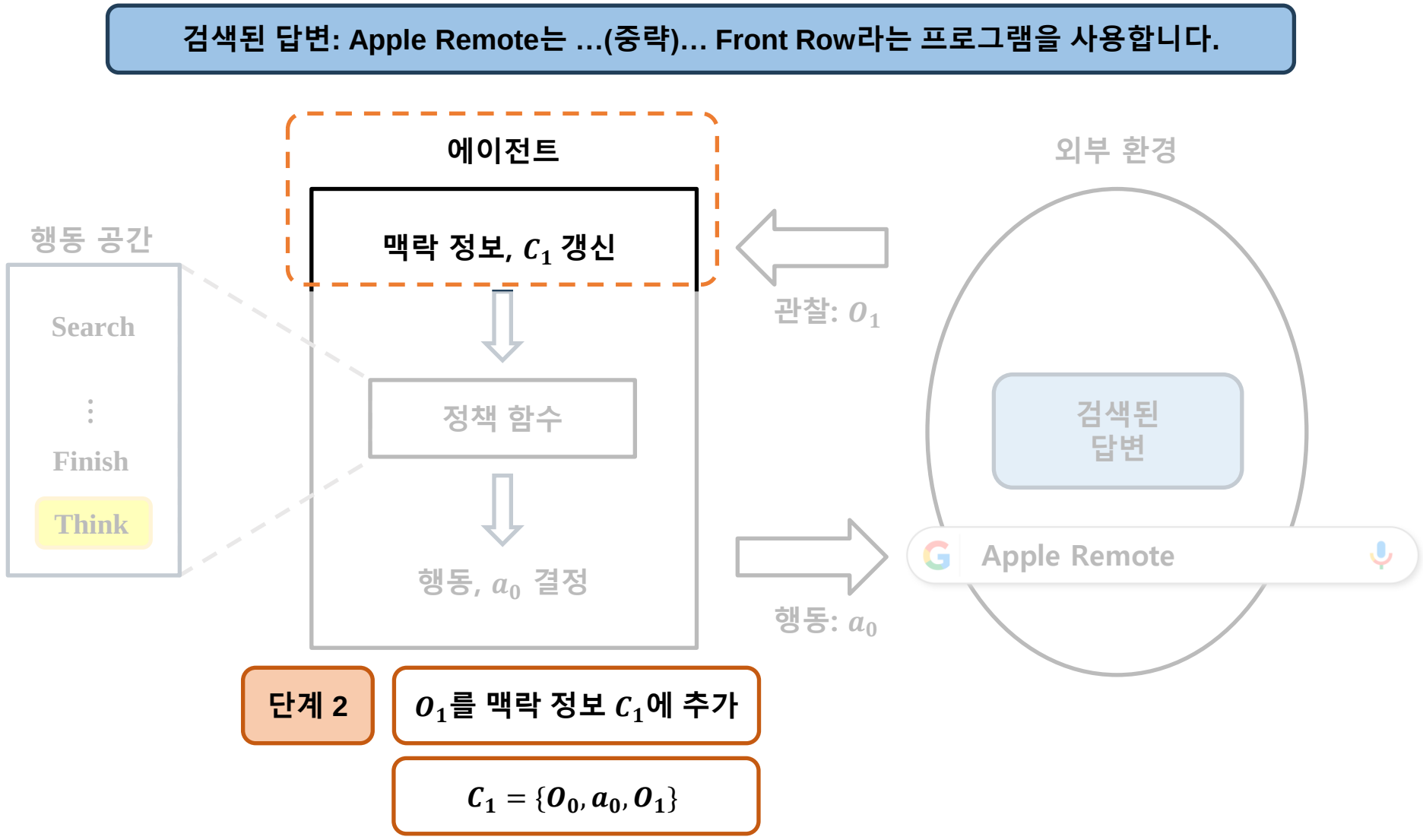
ReAct: 행동 공간에 추론을 추가!



ReAct: 행동 공간에 추론을 추가!

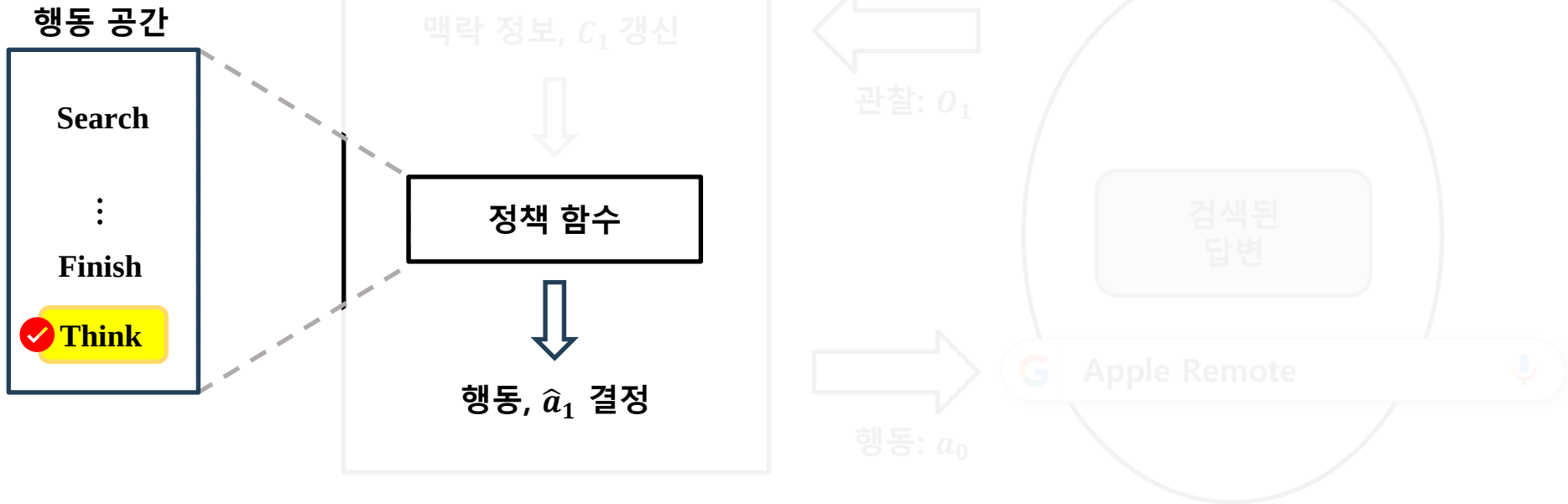


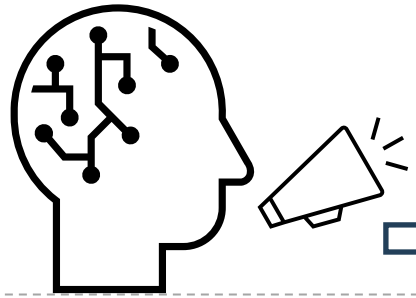
ReAct: 행동 공간에 추론을 추가!



ReAct: 행동 공간에 추론을 추가!

검색된 답변: Apple Remote는 …(중략)… Front Row라는 프로그램을 사용합니다.

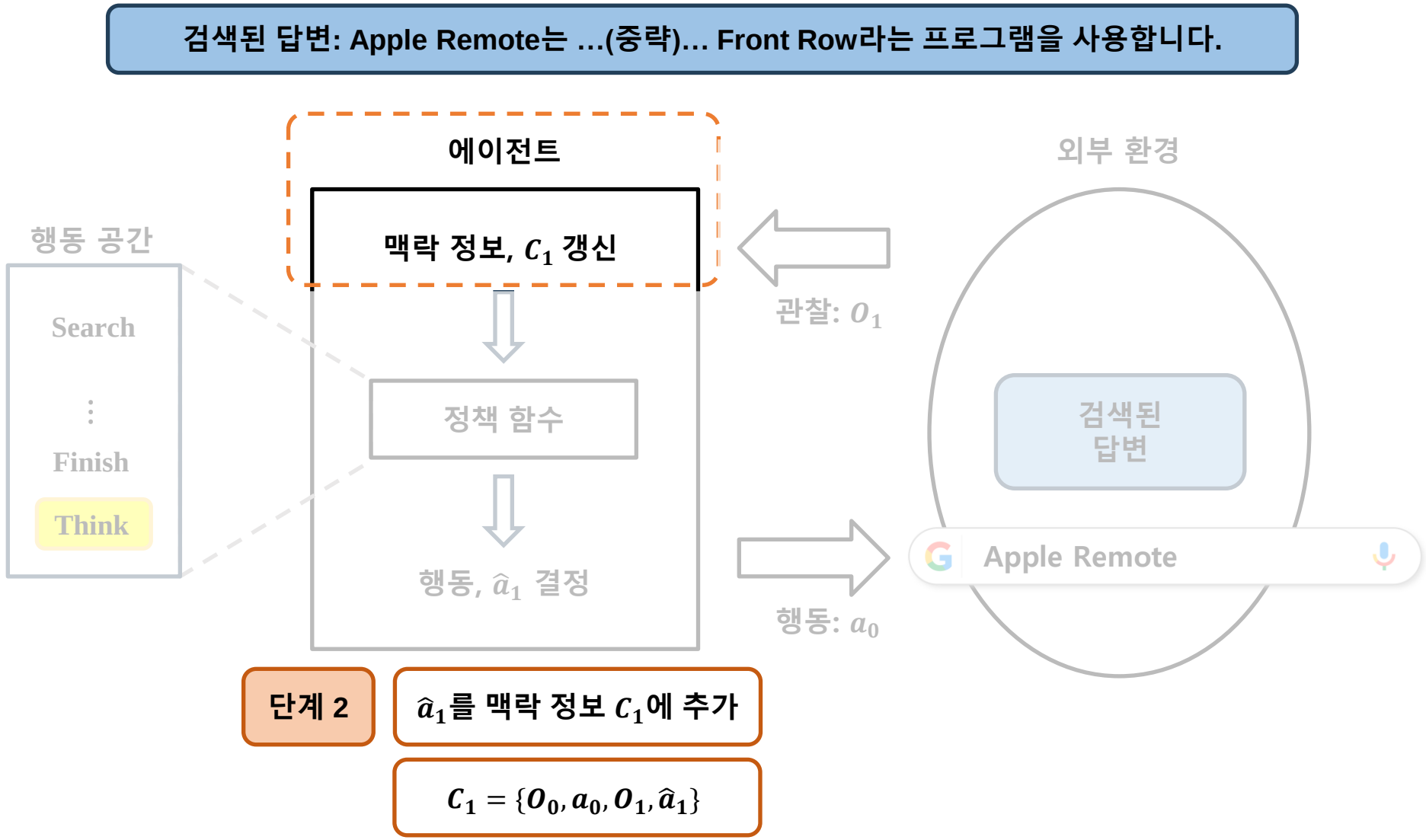




LLM 생성 프롬프트: Apple Remote는 Front Row라는 프로그램을 쓰는구나. 이에 대해 검색해보자.

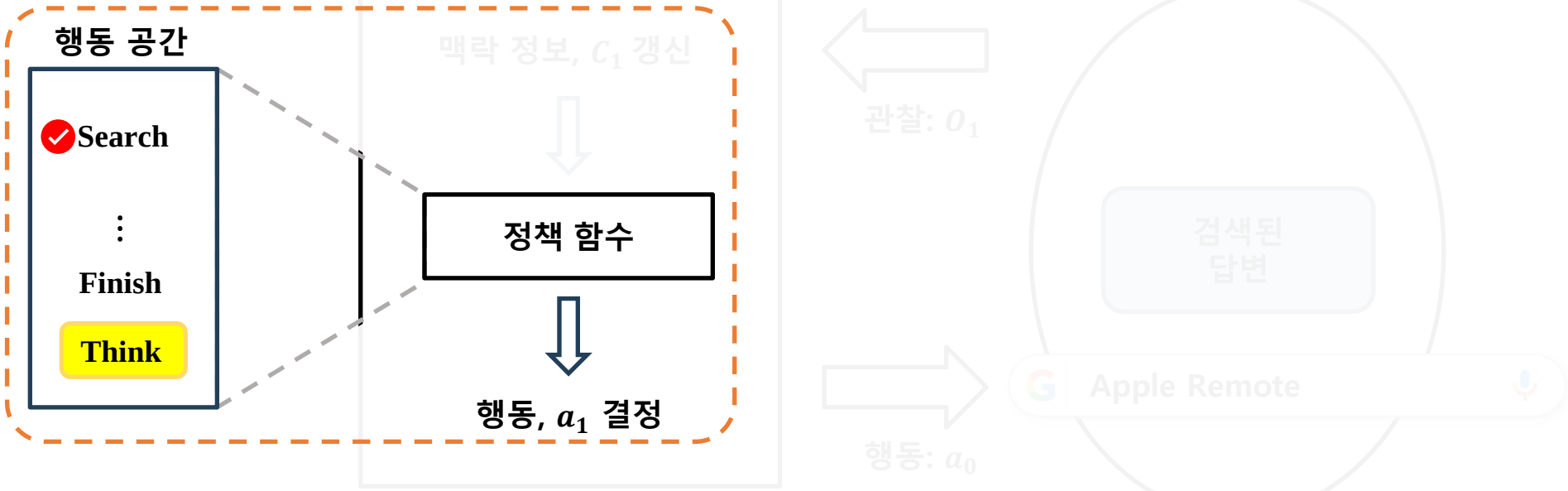
'Think'는 외부에 영향을 주지 않음 / 위 내용을 맥락 정보에 반영한 후 think 외 행동이 나올 때까지 반복

ReAct: 행동 공간에 추론을 추가!



ReAct: 행동 공간에 추론을 추가!

검색된 답변: Apple Remote는 …(중략)… Front Row라는 프로그램을 사용합니다.

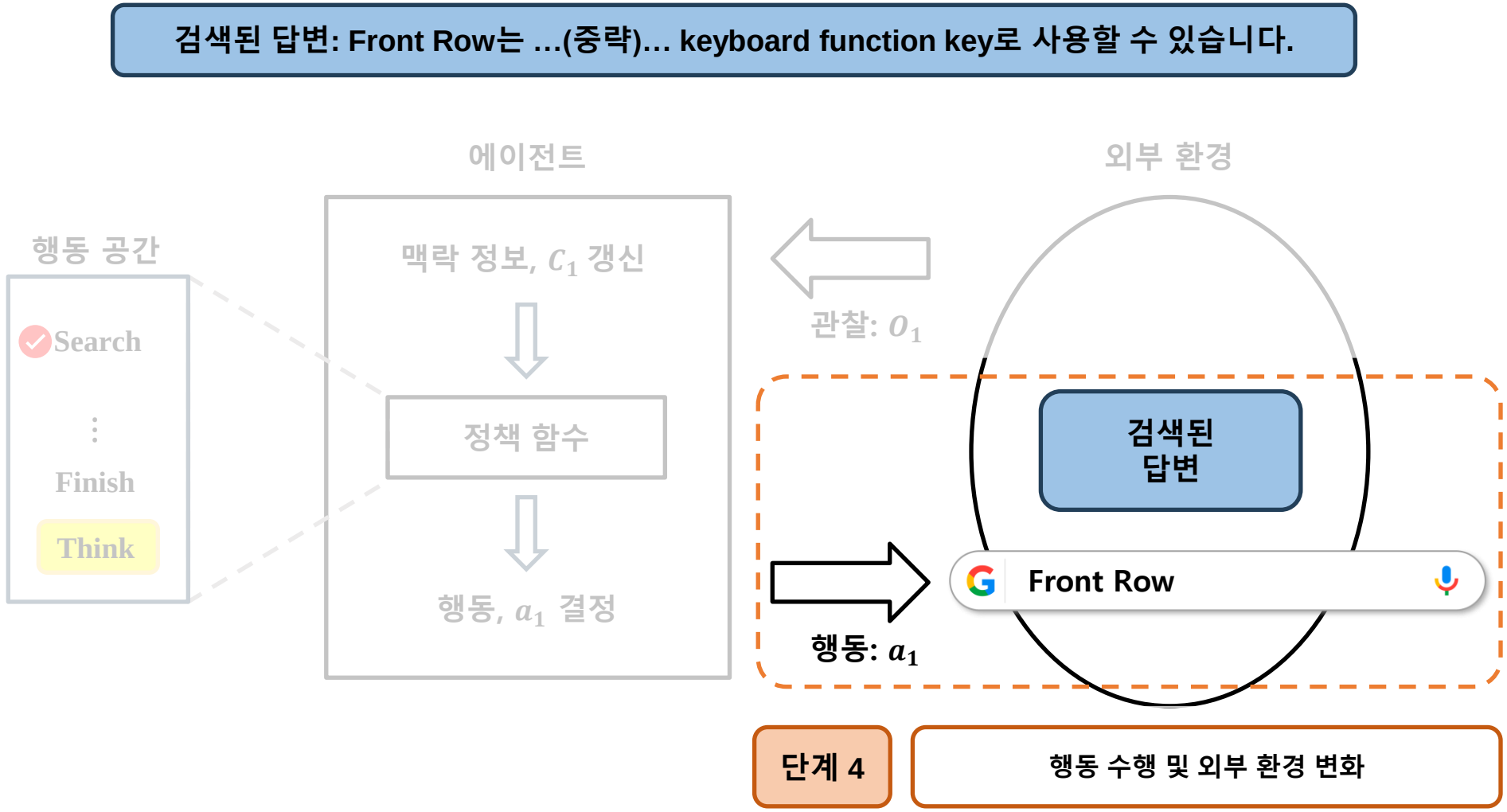


단계 3

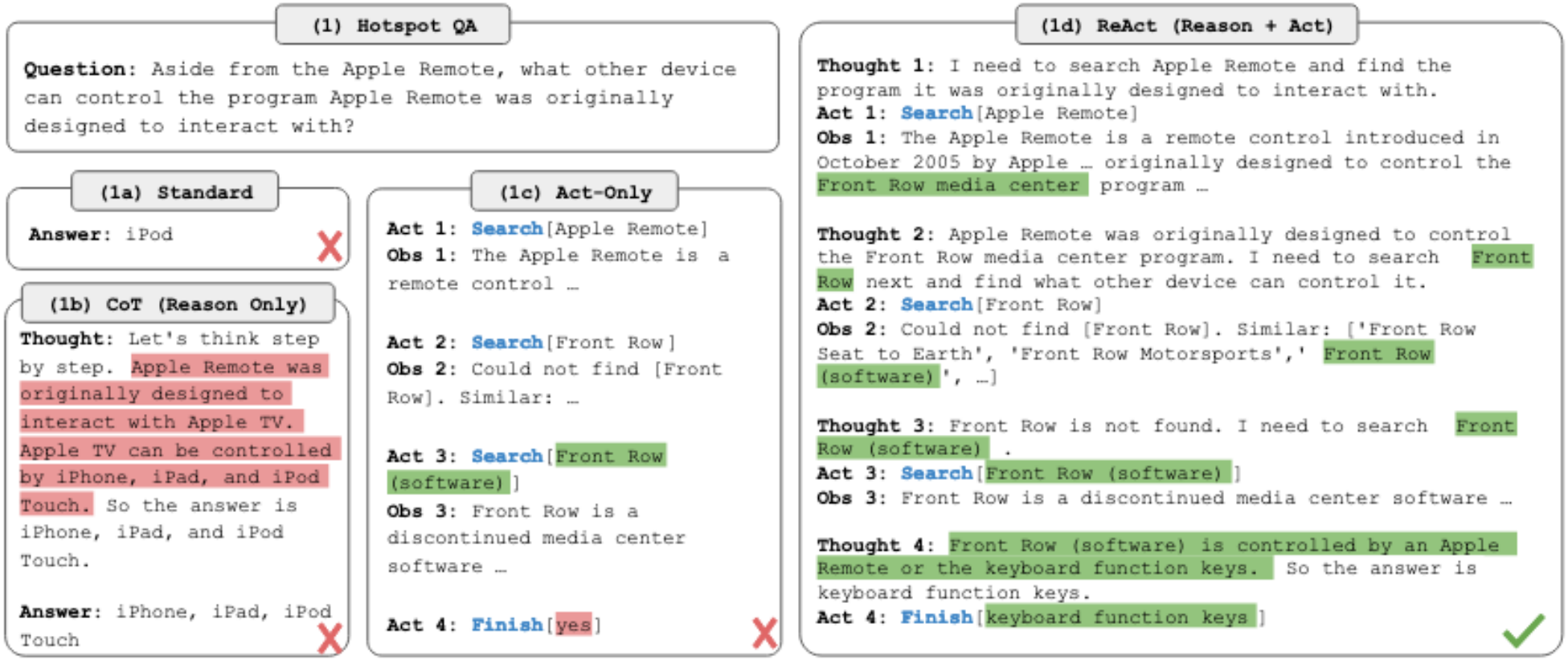
정책 함수를 통해 C_1에 적합한 a_1 선택

$\pi(a_1|c_1)$

ReAct: 행동 공간에 추론을 추가!



Baseline vs Reason Only vs Act Only vs ReAct



실험 1. Knowledge Intensive Reasoning

❖ 사용 데이터셋

- ➢ HotPotQA: 두 단락에 흩어진 정보를 적절히 연결해 추론해야만 답변 가능한 질문으로 구성
- ➢ Fever: 위키피디아 문서가 데이터로 주어지면, 이에 기반해 입력된 '주장'이 타당한지를 검증
  - 실험에서는 action을 통해 문서에 접근하는 능력까지 고려하기 위해 '주장'만을 입력

Reasoning Type	%	Example(s)
Inferring the *bridge entity* to complete the 2nd-hop question (Type I)	42	**Paragraph A:** The 2015 Diamond Head Classic was a college basketball tournament ... *Buddy Hield was named the tournament's MVP.* **Paragraph B:** *Chavano Rainier "Buddy" Hield* is a Bahamian professional basketball player for the **Sacramento Kings** of the NBA... **Q:** Which team does the player named 2015 Diamond Head Classic's MVP play for?
Comparing two entities (Comparison)	27	**Paragraph A:** LostAlone were a British rock band ... consisted of *Steven Battelle, Alan Williamson, and Mark Gibson*... **Paragraph B:** Guster is an American alternative rock band ... Founding members *Adam Gardner, Ryan Miller, and Brian Rosenworcel* began... **Q:** Did LostAlone and Guster have the same number of members? (**yes**)
Locating the **answer entity** by checking multiple properties (Type II)	15	**Paragraph A:** Several *current and former members of the Pittsburgh Pirates* ... John Milner, **Dave Parker**, and Rod Scurry... **Paragraph B:** **David Gene Parker**, *nicknamed "The Cobra"*, is an American former player in Major League Baseball... **Q:** Which former member of the Pittsburgh Pirates was nicknamed "The Cobra"?
Inferring about the property of an entity in question through a *bridge entity* (Type III)	6	**Paragraph A:** *Marine Tactical Air Command Squadron 28* is a United States Marine Corps aviation command and control unit based at *Marine Corps Air Station Cherry Point*... **Paragraph B:** *Marine Corps Air Station Cherry Point* ... is a United States Marine Corps airfield located in **Havelock, North Carolina**, USA ... **Q:** What city is the Marine Air Control Group 28 located in?
Other types of reasoning that require more than two supporting facts (Other)	2	**Paragraph A:** ... the towns of Yodobashi, **Okubo, Totsuka, and Ochiai town** *were merged into Yodobashi ward. ... Yodobashi Camera is a store with its name taken from the town and ward.* **Paragraph B**: *Yodobashi Camera* Co., Ltd. is *a major Japanese retail chain specializing in electronics, PCs, cameras and photographic equipment.* **Q:** Aside from Yodobashi, what other towns were merged into the ward which gave the major Japanese retail chain specializing in electronics, PCs, cameras, and photographic equipment it's name?

HotPotQA

Claim: The Rodney King riots took place in the most populous county in the USA.

[wiki/Los_Angeles_Riots]

The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

[wiki/Los_Angeles_County]

Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

Verdict: Supported

Fever

실험 1. Knowledge Intensive Reasoning

❖ 실험 세팅

- ➢ **Few-shot examples prompting: 각 method에 맞는 프롬프트로 모델 조정**
- ➢ **Fine-tuning: ReAct가 성공적으로 정답을 내놓은 과정 3000개를 수집해 모델을 미세 조정**

Original

Question	What is the elevation range for the area that the eastern sector of the Colorado orogeny extends into?
Answer	1,800 to 7,000 ft
Question	Musician and satirist Allie Goertz wrote a song about the "The Simpsons" character Milhouse, who Matt Groening named after who?
Answer	Richard Nixon
Question	Which documentary is about Finnish rock groups, Adam Clayton Powell or The Saimaa Gesture?
Answer	The Saimaa Gesture
Question	What profession does Nicholas Ray and Elia Kazan have in common?
Answer	director, screenwriter, actor
Question	Which magazine was started first Arthur's Magazine or First for Women?
Answer	Arthur's Magazine
Question	Were Pavel Urysohn and Leonid Levin known for the same type of work?
Answer	Yes

CoT

Question	What is the elevation range for the area that the eastern sector of the Colorado orogeny extends into?
Thought	Let's think step by step. The eastern sector of Colorado orogeny extends into the High Plains. High Plains rise in elevation from around 1,800 to 7,000 ft, so the answer is 1,800 to 7,000 ft.
Answer	1,800 to 7,000 ft
Question	Musician and satirist Allie Goertz wrote a song about the "The Simpsons" character Milhouse, who Matt Groening named after who?
Thought	Let's think step by step. Milhouse was named after U.S. president Richard Nixon, so the answer is Richard Nixon.
Answer	Richard Nixon
Question	Which documentary is about Finnish rock groups, Adam Clayton Powell or The Saimaa Gesture?
Thought	Let's think step by step. Adam Clayton Powell (film) is a documentary about an African-American politician, not Finnish rock groups. So the documentary about Finnish rock groups must instead be The Saimaa Gesture.

Act

Question	What is the elevation range for the area that the eastern sector of the Colorado orogeny extends into?
Action 1	Search[Colorado orogeny]
Observation 1	The Colorado orogeny was an episode of mountain building (an orogeny) in Colorado and surrounding areas.
Action 2	Lookup[eastern sector]
Observation 2	(Result 1 / 1) The eastern sector extends into the High Plains and is called the Central Plains orogeny.
Action 3	Search[High Plains]
Observation 3	High Plains refers to one of two distinct land regions:
Action 4	Search[High Plains (United States)]
Observation 4	The High Plains are a subregion of the Great Plains. From east to west, the High Plains rise in elevation from around 1,800 to 7,000 ft (550 to 2,130 m).[3]
Action 5	Finish[1,800 to 7,000 ft]
Question	Musician and satirist Allie Goertz wrote a song about the "The Simpsons" character Milhouse, who Matt Groening named after who?
Action 1	Search[Milhouse]
Observation 1	Milhouse Mussolini Van Houten is a recurring character in the Fox animated television series The Simpsons voiced by Pamela Hayden and created by Matt Groening.

ReAct

Question	What is the elevation range for the area that the eastern sector of the Colorado orogeny extends into?
Thought 1	I need to search Colorado orogeny, find the area that the eastern sector of the Colorado orogeny extends into, then find the elevation range of the area.
Action 1	Search[Colorado orogeny]
Observation 1	The Colorado orogeny was an episode of mountain building (an orogeny) in Colorado and surrounding areas.
Thought 2	It does not mention the eastern sector. So I need to look up eastern sector.
Action 2	Lookup[eastern sector]
Observation 2	(Result 1 / 1) The eastern sector extends into the High Plains and is called the Central Plains orogeny.

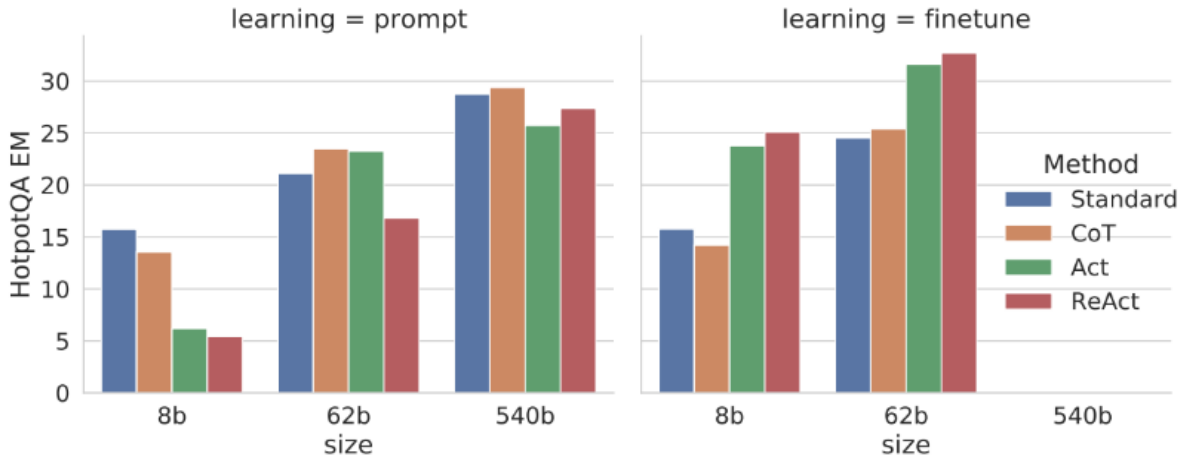
실험 1. Knowledge Intensive Reasoning

❖ 실험 결과

- ➢ CoT-SC와 ReAct 혼합 방법론이 가장 좋은 성능을 보임
- ➢ ReAct 사용 시 사실에 근거한 답변을 내놓으므로 환상은 줄어들지만 그만큼 유연성을 잃은 것으로 보임
- ➢ 미세 조정 시, 사이즈가 작은 모델에서도 좋은 성능을 보임

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

	Type	Definition	ReAct	CoT
Success	True positive	Correct reasoning trace and facts	94%	86%
	False positive	Hallucinated reasoning trace or facts	6%	14%
Failure	Reasoning error	Wrong reasoning trace (including failing to recover from repetitive steps)	47%	16%
	Search result error	Search return empty or does not contain useful information	23%	-
	Hallucination	Hallucinated reasoning trace or facts	0%	56%
	Label ambiguity	Right prediction but did not match the label precisely	29%	28%



실험 2. Decision Making

❖ 목표 달성을 위해 복잡한 의사 결정 과정을 수행할 수 있는 지 평가

- ➢ ALFWord: 텍스트 기반 게임 데이터셋. 달성해야 하는 목표와 수행할 수 있는 행동 세트가 주어짐
- ➢ WebShop: 온라인 쇼핑 환경을 텍스트로 구축. 목표한 구매 리스트를 구매할 수 있는 지 테스트

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

Table 3: AlfWorld task-specific success rates (%). BUTLER and BUTLER$_g$ results are from Table 4 of Shridhar et al. (2020b). All methods use greedy decoding, except that BUTLER uses beam search.

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

Table 4: Score and success rate (SR) on Webshop. IL/IL+RL taken from Yao et al. (2022).