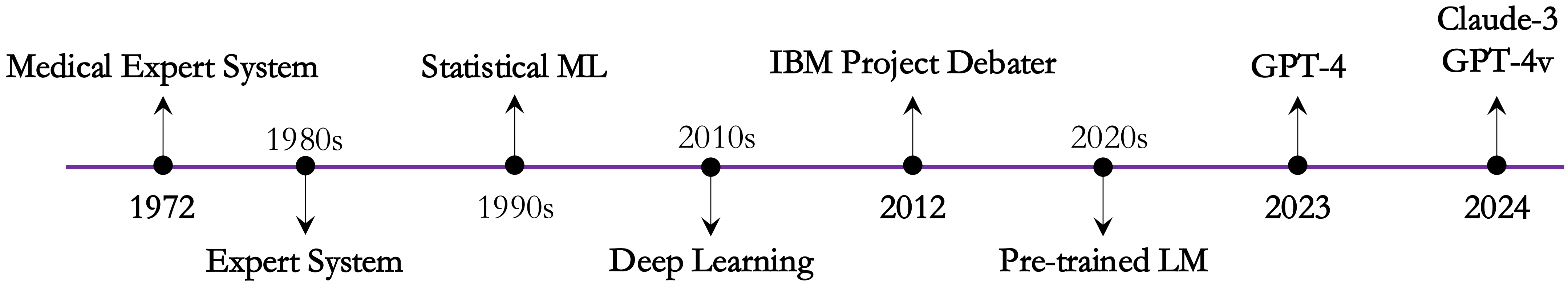
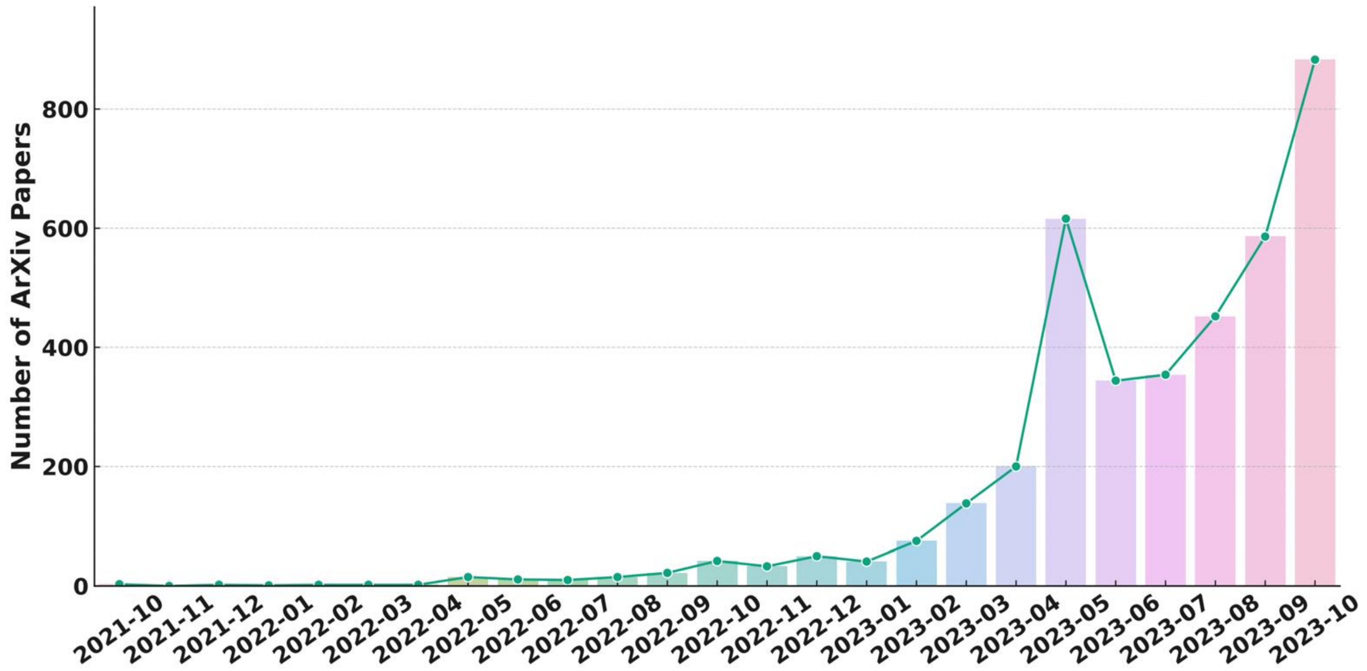
AMOR: A Recipe for Building Adaptable MOdulaR Knowledge Agents Through Process Feedback

Jian Guan, Wei Wu, Zujie Wen, Peng Xu, Hongning Wang, Minlie Huang

CoAI Group, Tsinghua University; Ant Group

Reasoning as the Longstanding Aim of AI



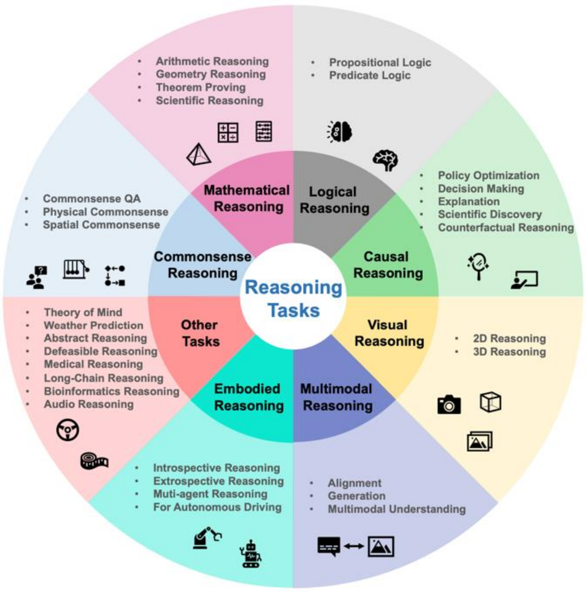


Number of arXiv Papers on "Reasoning with Large Language Models" over the past two years.

Sun et al. A Survey of Reasoning with Foundation Models. 2023.

What is Reasoning in the Context of LLMs?

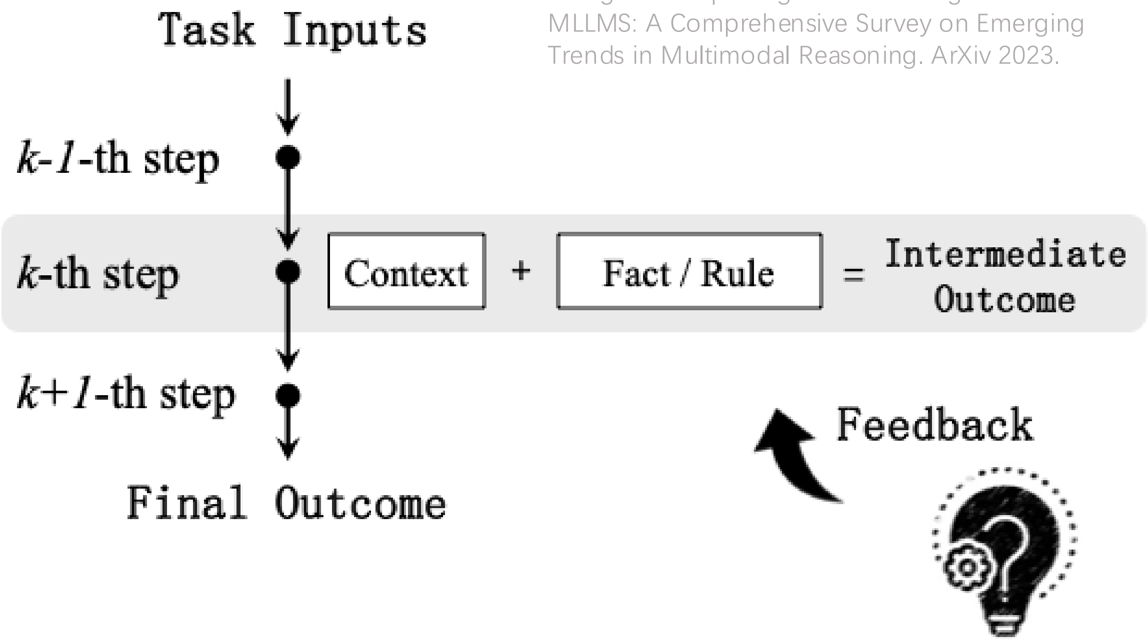
- Natural language reasoning is a **process** of **selecting** and **interpreting information** from given contexts, **making connections**, **verifying**, and finally **drawing conclusions**.



Reasoning tasks span various domains and require broad knowledge.

Sun et al. A Survey of Reasoning with Foundation Models. 2023.

Wang et al. Exploring the Reasoning Abilities of MLLMS: A Comprehensive Survey on Emerging Trends in Multimodal Reasoning. ArXiv 2023.



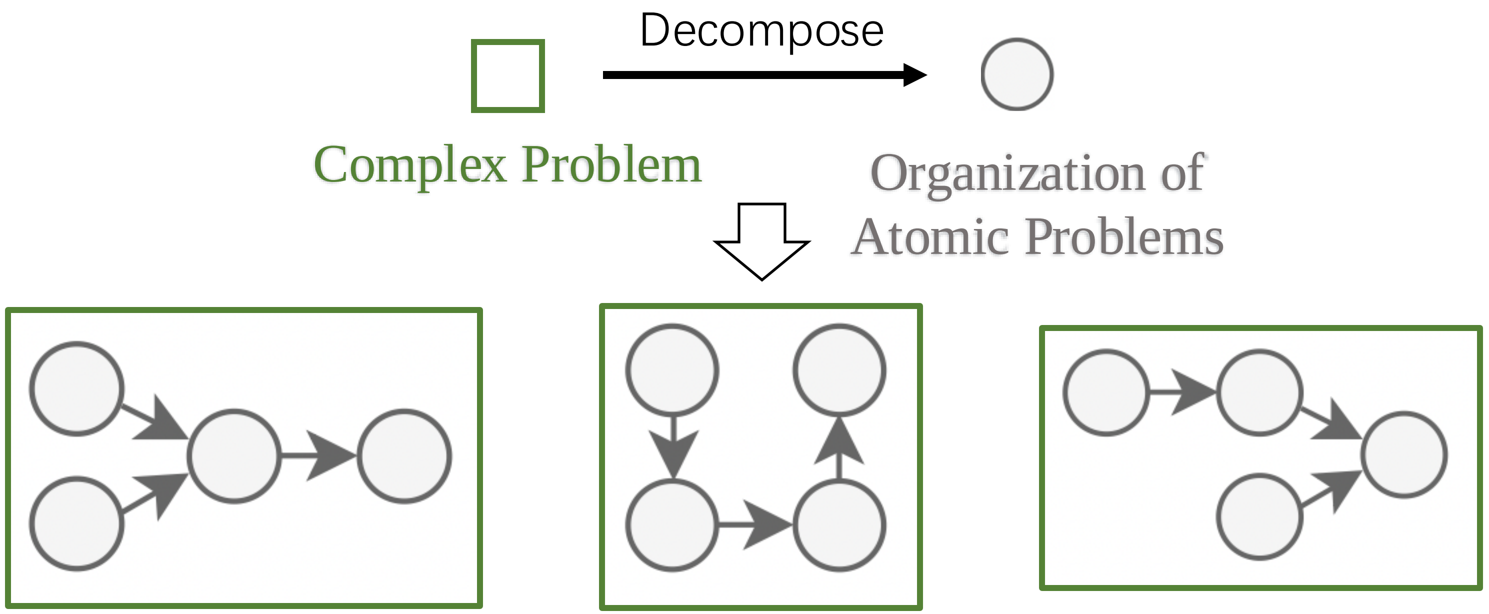
Various reasoning tasks can be formalized as a step-by-step process of solving subtasks.

Principles for Solving Complex Reasoning Tasks

- I'll call "Society of Mind" this scheme in which **each mind is made of many smaller processes.** These we'll call agents. **Each mental agent by itself can only do some simple thing** that needs no mind or thought at all. Yet when we **join these agents … leads to true intelligence.**

- The law of thought depends not only upon the **properties of those brain cells,** but also on **how they are connected.**

—— 《Society of Mind》, Marvin Minsky (1969 Turing Award Winner)



Core Questions

How to define each atomic problem?

How to organize the atomic problems?

Towards Specific and General Approaches

- Math Word Problems
- Coding
- Question-Answering

Model　Data　Algorithm

Expert Knowledge

Specialist

Agent

Core Questions

How to define each atomic problem?

How to organize the atomic problems?

- Deduction
- Induction
- Abduction

General Algorithms/Data/Systems…

Generalist

AGI

Desiderata for a qualifying agent

- ✓ Possess a robust **reasoning logic** to address a specific task
- ✓ Maintain an **adaptive mechanism** to adjust to specific environments
- ✓ Be amenable to human interventions through direct **feedback**

Method	Reasoning Logic		Adaptive Mechanism	Feedback
	Step	Inter-Step Dependency		
WebGPT [27]	Tool Invoking	*Undefined*	Imitation Learning from Humans	Outcome
CoT [43]	Reasoning	*Undefined*	Prompting	*Undefined*
ToT [49]	Reasoning	*Undefined*	Prompting	Process
ReAct [50]	Reasoning&Tool Invoking	*Undefined*	Prompting	*Undefined*
Reflexion [35]	Reasoning&Tool Invoking	*Undefined*	Prompting	Process
AgentLM [52]	Reasoning&Tool Invoking	*Undefined*	Imitation Learning from LLMs	Outcome
MetaGPT [14]	Specialized Module	Sequential Pipeline	Prompting	Process
LUMOS [51]	Specialized Module	Sequential Pipeline	Imitation Learning from Humans	*Undefined*
AMOR	Specialized Module	Finite State Machine	Exploration&Exploitation	Process

No existing agents fulfill all the required criteria due to their *uncontrollable reasoning logic*, *static model capability*, or *sparse/missing feedback signals*.

Our recipe to building agents

- **Part I:** Finite-State Machine (FSM)-based Reasoning Logic
  - Structured Thinking.
  - Skill Disentanglement. (cf. Part II)
  - Intervenable Workflow. (cf. Part III)
- **Part II:** Warming-up open-source LLMs
  - Reasoning steps (modules) of AMOR can be **independently optimized** with separate public datasets.
- **Part III:** Adaptation through process feedback
  - AMOR can adapt to specific knowledge environments through **process-based supervision** to each of the reasoning steps (modules) from users.

Our recipe to building agents

- **Part I:** Finite-State Machine (FSM)-based Reasoning Logic

Driven by Expert Knowledge

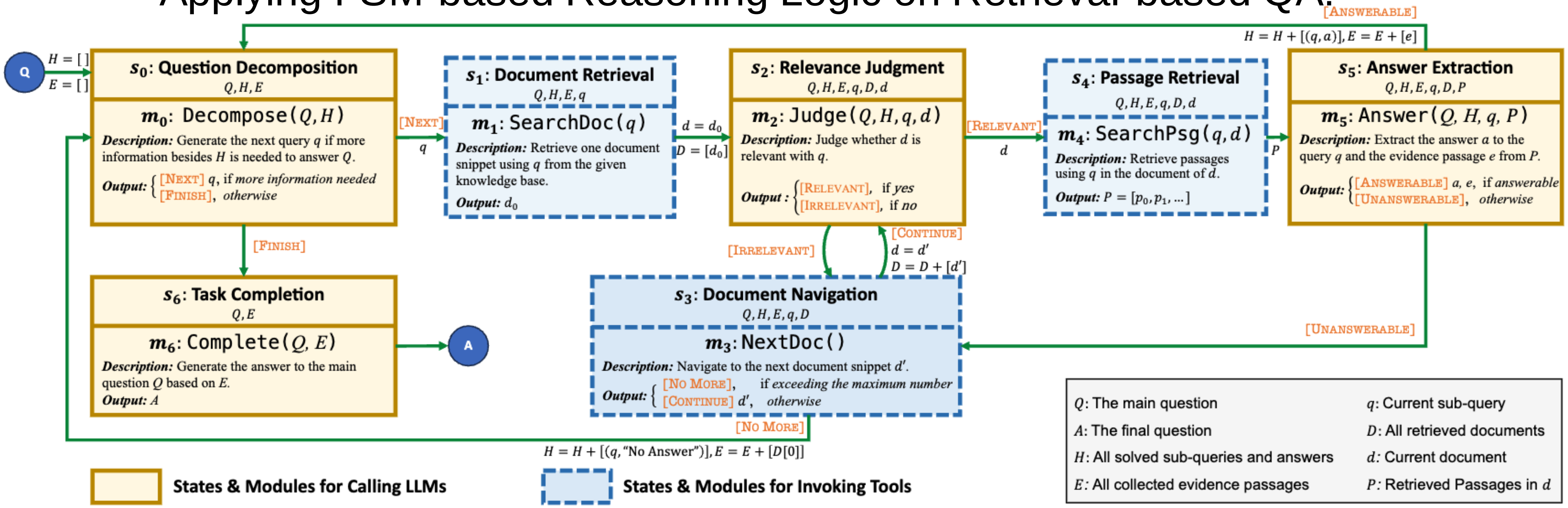
How to define each atomic problem? → Using Specialized Modules

How to organize the atomic problems? → Using FSM

- An FSM can be defined as a quadruple:
  - ➢ $\mathcal{S} = \{s_0, \ldots, s_{N-1}\}$: a set of states ($s_0$: initial state; s_{N-1}: final state)
  - ➢ $\mathcal{M} = \{m_0, \ldots, m_{N-1}\}$: a set of modules, with one-to-one correspondence with $\mathcal{S}$
    - $\mathcal{M}_{\text{TOOL}}$: Tool modules for invoking tools
    - $\mathcal{M}_{\text{LLM}}$: LLM modules for calling LLMs
  - ➢ $\mathcal{E}$: the set of all possible outputs of $\mathcal{M}$
  - ➢ $\mu : \mathcal{S} \times \mathcal{E} \rightarrow \mathcal{S}$: transition function

Our recipe to building agents

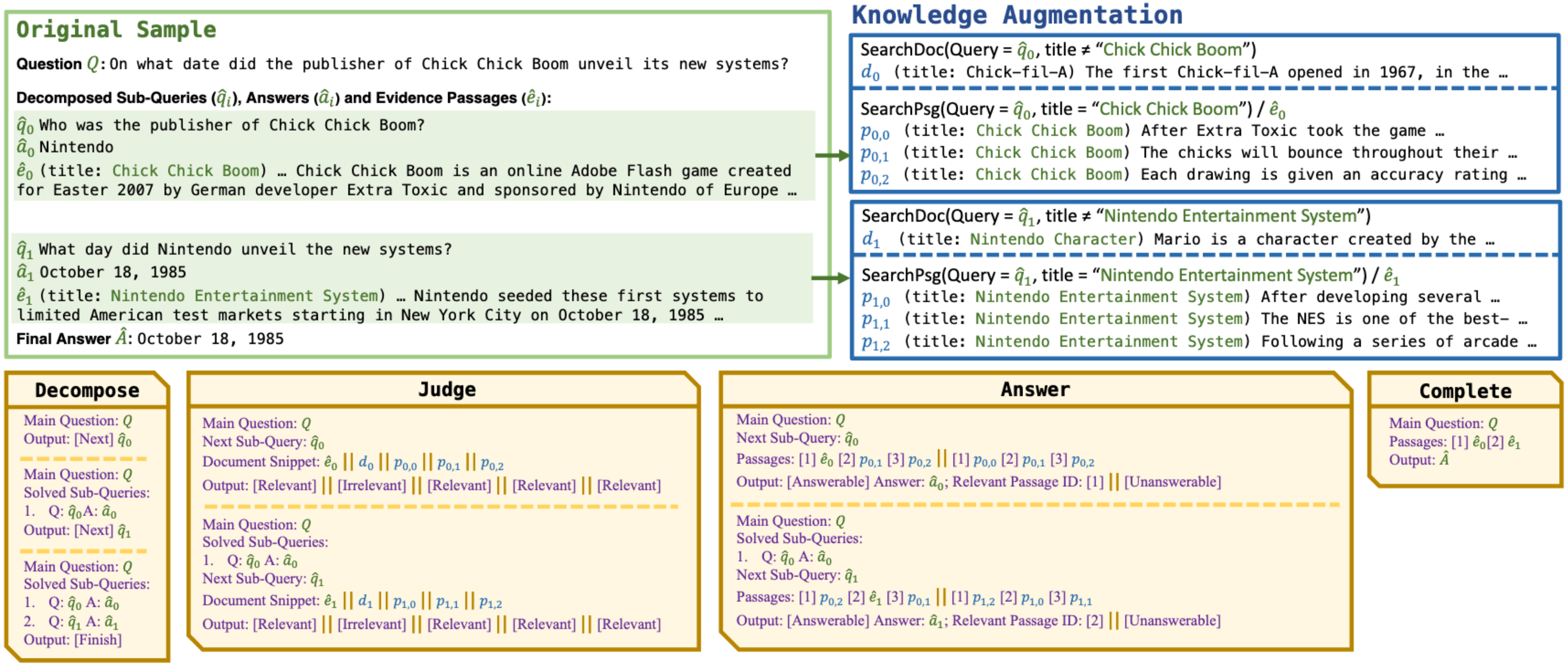
- **Part I:** Finite-State Machine (FSM)-based Reasoning Logic
  - Applying FSM-based Reasoning Logic on Retrieval-based QA:



AMOR's state transition diagram. Each box represents a state and the corresponding module that is executed when entering the state. There may be multiple categories of execution results distinguished by special branch tokens such as "[NEXT]." Then AMOR determines the next state based on the branch tokens.

Our recipe to building agents

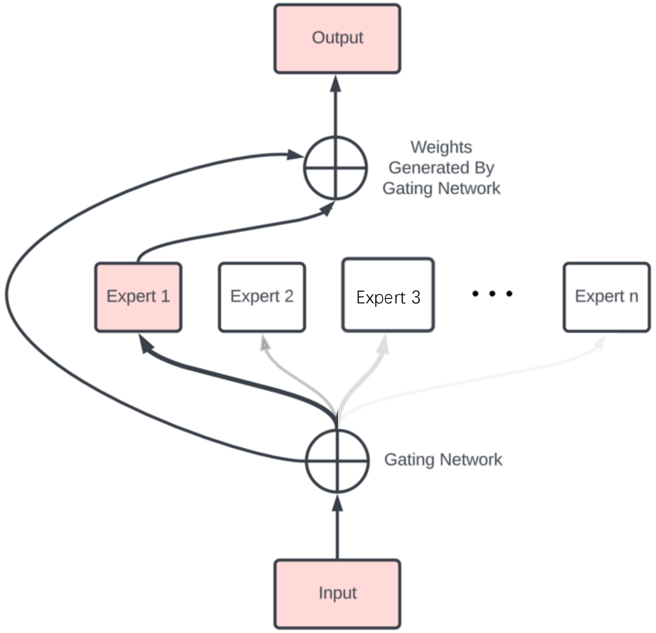
- **Part II:** Warming-up open-source LLMs
  - **Data**



On the top left is a sample question from Musique, providing **ample information** for constructing training examples for four LLM modules of AMOR (bottom). We augment **extra knowledge** for the **Judge** and **Answer** module by invoking the **SearchDoc** and **SearchPsg** tools (top right). In each example, we use "||" to separate different examples for training.

Our recipe to building agents

- **Part II:** Warming-up open-source LLMs
  - **Model: Module-Aware Mixture-of-Experts (MA-MoE).** When AMOR executes a certain module, its module index will be provided to the routers of the model to indicate which expert shoul



  - **Training:**

$$\mathcal{L}_1 = \mathbb{E}_{m \in \mathcal{M}_{\text{LLM}}, (\hat{s}, \hat{y}) \in \mathcal{D}_m} - \lambda_m \log \pi_{\theta_m}(\hat{y} | \hat{s})$$

Our recipe to building agents

- **Part III:** Adaptation through process feedback

Algorithm 2 Adaptation through Process Feedback

Input: $\{\pi_{\theta_m}^{\text{WFT}}\}$: Initial Policy; T: Exploration Steps between Exploitation; I: Number of Iterations.
Output: $\{\pi_{\theta_m}\}$: Adapted Policy.

```
while i ← 1 to I do
    𝓡 = [ ] // Feedback-Refined Reasoning Processes
    while t ← 1 to T do
        // Exploration
        Receive an input question Q.
        Collect AMOR_θ's reasoning process R. // Algorithm 1
        // Feedback Collection for Each LLM Module
        foreach Step r_k ∈ R (k = 0, 1, 2, ···) do
            Extract the state s_k and output y_k from r_k.
            if The corresponding module m_k ∈ 𝓜_LLM then
                Collect feedback f_k for s_k and y_k.
                Determine ỹ_k and o_k based on f_k. // Eq. 2
                𝓡.append([s_k, ỹ_k, o_k])
    // Exploitation
    Optimize {θ_m} to minimize 𝓛_2 on 𝓡. // Eq. 3
return {π_θm}
```

Algorithm 1 FSM-based Reasoning Logic

Input: Agent at the state $s = s_0$; Q: Question.
Output: A: Final Answer; R: Reasoning Process.

```
R = [ ]
while s ≠ s_{N-1} do
    y = m(s) // Obtain the output y given s
                from the corresponding module m.
    R.append({"state": s, "output": y})
A = y
return A, R
```

$$\tilde{y}_k, o_k = \begin{cases} y_k, 1 & \text{if } f_k = \text{“right”}, \\ y_k, 0 & \text{if } f_k = \text{“wrong”}, \\ f_k, 1 & \text{if } f_k \text{ is refinement.} \end{cases}$$

$$\mathcal{L}_2 = \mathbb{E}_{m \in \mathcal{M}_{\text{LLM}}, (s_k, \tilde{y}_k, o_k) \in \mathcal{R}_m} - \lambda_m \left[o_k - \beta \log \left(\pi_{\theta_m}(\tilde{y}_k | s_k) / \pi_{\theta_m}^{\text{WFT}}(\tilde{y}_k | s_k) \right) \right]$$

Empirical Evaluation of AMOR

Dataset	Knowledge Base	Avg. Len	# Train	# Val	# Test
HotpotQA	Wikipedia Articles	138	2,000	100	500
PubMedQA	PubMed Abstracts	303	401	44	445
QASPER	One NLP Paper	102	700	45	382

Datasets for adaptation and evaluation. Avg. Len refers to the average length of passages in the corresponding knowledge base, counted by the GPT tokenizer.

Empirical Evaluation of AMOR

- FSM-based reasoning logic outperforms prior frameworks by 30~40%

Method	Base LLM	HotpotQA		PubMedQA	QASPER	
		EM	F1	ACC	EM	F1
Without Fine-Tuning						
ReAct	L-7B	12.2	16.6	61.8	6.0	19.2
$\text{AMOR}_{\text{w/o FT}}$	L-7B	26.0	34.6	62.9	4.5	21.3
CoT	GPT-3.5	28.0‡	-	*N/A*	*N/A*	*N/A*
OneR	GPT-3.5	33.4	42.1	72.6	6.8	23.3
ReAct	GPT-3.5	30.8	38.8	58.2	5.8	27.0
ReWoo	GPT-3.5	30.4†	40.1†	-	-	-
$\text{AMOR}_{\text{w/o FT}}$	GPT-3.5	39.6	49.3	68.8	10.0	30.8
CoT	GPT-4	45.0‡	-	*N/A*	*N/A*	*N/A*
ReAct	GPT-4	42.0‡	-	62.1	7.1	26.2
$\text{AMOR}_{\text{w/o FT}}$	GPT-4	**55.2**	**65.2**	**80.0**	**11.5**	**37.4**

Experiment results when with two-stage fine-tuning

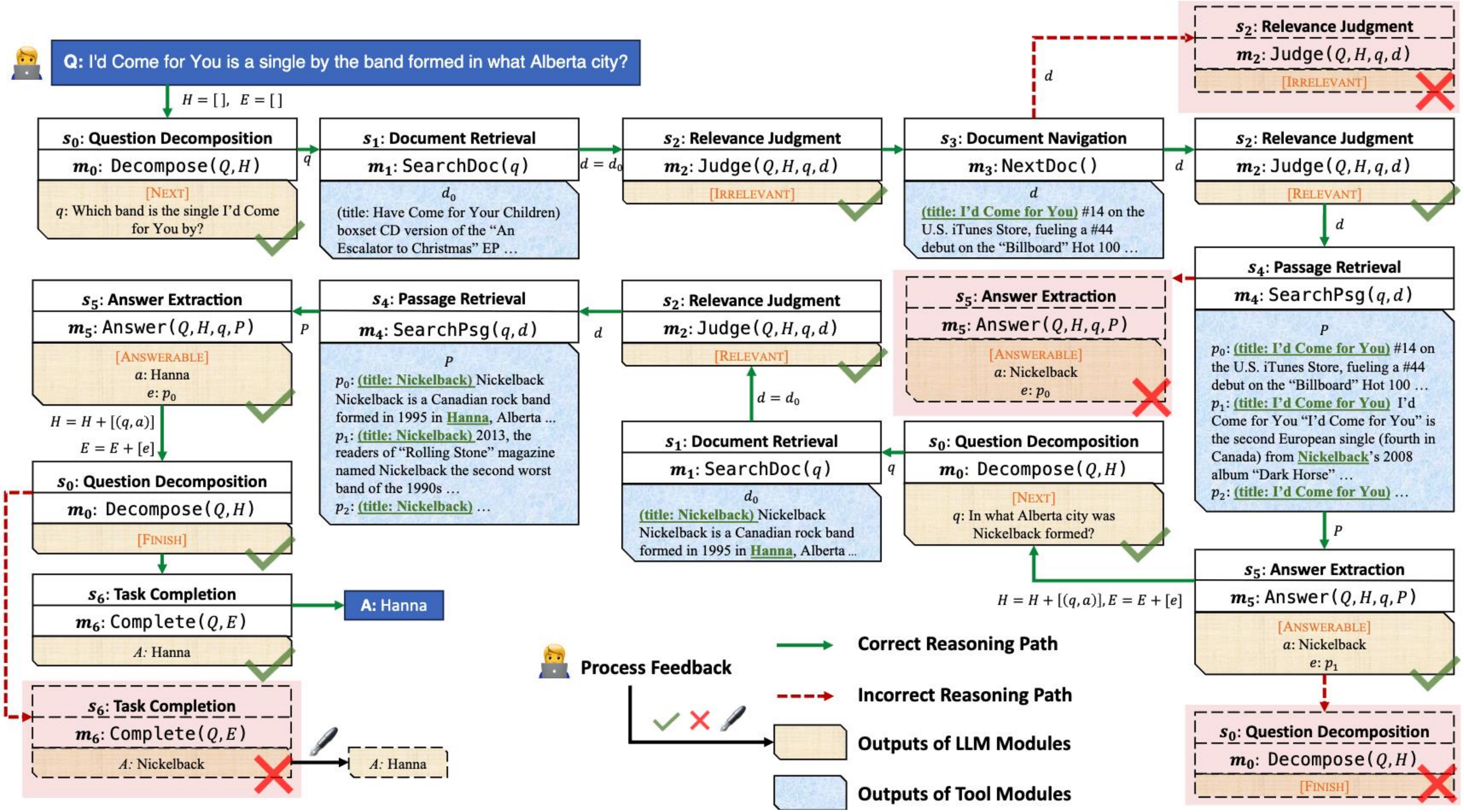
Empirical Evaluation of AMOR

- Process feedback is more effective than outcome feedback.

Method	Base LLM	HotpotQA EM	HotpotQA F1	PubMedQA ACC	QASPER EM	QASPER F1
With Fine-Tuning						
OneR*	**L-7B**	34.8	43.8	75.3	11.0	25.5
Self-RAG	**L-7B**	22.4	32.9	62.6	2.1	17.9
AgentLM	**L-7B**	22.0†	-	64.9	4.2	20.2
FireAct	**L-7B**	26.2†	-	66.1	6.5	18.4
Lumos	**L-7B**	29.4†	-	70.3	7.1	19.5
AMOR$_{\text{Process}}$*	**L-7B**	45.8	54.9	81.1	**19.1**	35.3
AMOR$_{\text{WFT}}$	**L-7B**	33.6	41.9	73.4	11.1	23.6
AMOR$_{\text{Outcome}}$*	**L-7B**	40.8	49.3	77.5	9.4	25.4
AgentLM	**L-13B**	29.6†	-	67.9	7.1	24.4
AMOR$_{\text{Process}}$*	**L-13B**	**48.6**	**55.3**	**82.2**	18.1	**38.0**
AMOR$_{\text{WFT}}$	**L-13B**	36.8	44.1	74.6	15.2	27.3
AMOR$_{\text{Outcome}}$*	**L-13B**	42.4	51.6	80.1	9.9	26.5

Experiment results when without fine-tuning

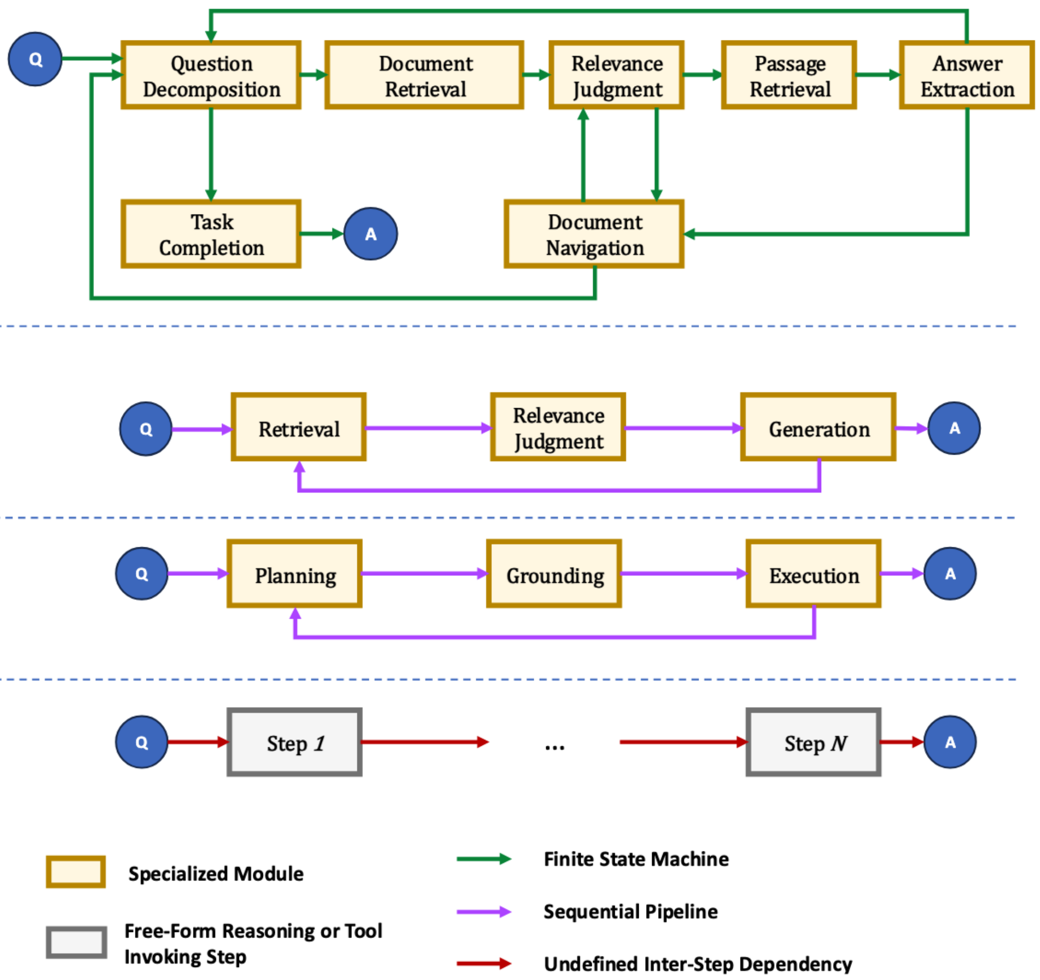
Case Study



Q: I'd Come for You is a single by the band formed in what Alberta city?
s0: Question Decomposition
m0: Decompose(Q,H)
[NEXT]
q: Which band is the single I'd Come for You by?
s1: Document Retrieval
m1: SearchDoc(q)
(title: Have Come for Your Children) boxset CD version of the "An Escalator to Christmas" EP ...
s2: Relevance Judgment
m2: Judge(Q,H,q,d)
[IRRELEVANT]
s3: Document Navigation
m3: NextDoc()
(title: I'd Come for You) #14 on the U.S. iTunes Store, fueling a #44 debut on the "Billboard" Hot 100 ...
[RELEVANT]
s4: Passage Retrieval
m4: SearchPsg(q,d)
s5: Answer Extraction
m5: Answer(Q,H,q,P)
[ANSWERABLE]
a: Nickelback
a: Hanna
[FINISH]
s6: Task Completion
m6: Complete(Q,E)
A: Hanna
A: Nickelback
q: In what Alberta city was Nickelback formed?
(title: Nickelback) Nickelback Nickelback is a Canadian rock band formed in 1995 in Hanna, Alberta ...
(title: Nickelback) 2013, the readers of "Rolling Stone" magazine named Nickelback the second worst band of the 1990s ...
(title: I'd Come for You) I'd Come for You "I'd Come for You" is the second European single (fourth in Canada) from Nickelback's 2008 album "Dark Horse" ...
Process Feedback
Correct Reasoning Path
Incorrect Reasoning Path
Outputs of LLM Modules
Outputs of Tool Modules

Summary

Prior Reasoning Methods	Reasoning Logic	Adaptive Mechanism for New Environments	Human Intervention in the Reasoning Process
AMOR	FSM. **Advantage:** The controllable FSM-based reasoning logic has a stronger capacity for handling complex tasks than simple pipelines	Exploration and exploitation. **Advantage:** It enables AMOR to adapt effectively to specific domains based on human feedback.	Process Feedback. **Advantage:** It enables humans to provide direct feedback on the individual modules within the FSM-based reasoning process.
Retrieval-Augmented Generation (e.g., Self-RAG)	Sequential Pipeline. **Drawback:** It is difficult to handle complex tasks.	*Undefined*	*Undefined*
Agents with Modular Reasoning (e.g., LUMOS)	Sequential Pipeline. **Drawback:** It is difficult to handle complex tasks.	Prompting or Imitation Learning from Humans/LLMs. **Drawbacks:** The former often leads to suboptimal results, while the latter suffers from the scarcity of high-quality data.	*Undefined*
Agents with Free-Form Reasoning (e.g., AgentLM, FireAct)	*Undefined*	Prompting or Imitation Learning from Humans/LLMs. **Drawbacks:** The former often leads to suboptimal results, while the latter suffers from the scarcity of high-quality data.	Outcome Feedback. **Drawbacks:** (1) Outcome feedback alone is often too sparse and insufficient to improve the intermediate reasoning steps effectively; (2) The reasoning steps taken by LLMs can frequently contradict or deviate from the desired outcome.



Elaboration regarding the advantages and drawbacks when comparing AMOR with prior agents

The reasoning processes of AMOR and related works.

Future Work

- Math Word Problems
- Coding
- Question-Answering

Model Data Algorithm

Expert Knowledge

Specialist

Agent

Core Questions

How to define each atomic problem?

How to organize the atomic problems?

- Deduction
- Induction
- Abduction

General Algorithms/Data/Systems…

Generalist

AGI