

# A Review of Prominent Paradigms for LLM-Based Agents: Tool Use (Including RAG), Planning, and Feedback Learning

Xinzhe Li \*

Independent Researcher  
sergioli212@outlook.com

## Abstract

Tool use, planning, and feedback learning are currently three prominent paradigms for developing Large Language Model (LLM)-based agents across various tasks. Although numerous frameworks have been devised for each paradigm, their intricate workflows and inconsistent taxonomy create challenges in understanding and reviewing the frameworks across different paradigms. This survey introduces a unified taxonomy to systematically review and discuss these frameworks. Specifically, 1) the taxonomy defines environments/tasks, common LLM-profiled roles or LMPRs (policy models, evaluators, and dynamic models), and universally applicable workflows found in prior work, and 2) it enables a comparison of key perspectives on the implementations of LMPRs and workflow designs across different agent paradigms and frameworks. 3) Finally, we identify three limitations in existing workflow designs and systematically discuss the future work. Resources have been made publicly available at in our GitHub repository <https://github.com/xinzhel/LLM-Agent-Survey>.

## 1 Introduction

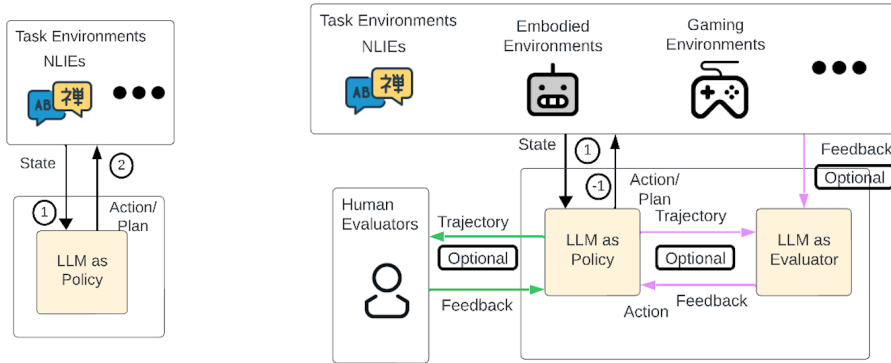
Large Language Models (LLMs) have acquired extensive general knowledge and human-like reasoning capabilities (Santurkar et al., 2023; Wang et al., 2022; Zhong et al., 2022, 2023), positioning them as pivotal in constructing AI agents known as LLM-based agents. In the context of this survey, LLM-based agents are defined by their ability to interact actively with external tools (**tool use**) or environments (Yao et al., 2023b) and are designed to function as integral components of agency for **planning** (Yao et al., 2023a) and **feedback learning** (Shinn et al., 2023).

\*Part of work conducted while affiliated with Deakin University.

**Comparisons with Existing Surveys** Current surveys lack a coherent and unified starting point for discussing the three paradigms due to two main limitations: **1) Focusing on a specific paradigm or domain:** For example, Huang et al. (2024) explore frameworks in the planning paradigm. Hu et al. (2024); Gallotta et al. (2024) examine LLM-based agents in the context of games. In contrast, these paradigms represent general principles and mental models that govern the development and behavior of these agents. Hence, their frameworks should be discussed in a task-agnostic way. To achieve this, we extract universal workflows. Unless otherwise specified, in this survey, the term “framework” refers to a complete workflow for specific low-level tasks, while “workflow” refers to the task-agnostic process derived from one or more frameworks. **2) Lack of a unified basis for comparison:** Although Wang et al. (2024a) cover all three paradigms, they do not analyze the implementation of algorithmic frameworks in a unified basis. In contrast, we summarize task-agnostic LLM-profiled roles as the foundation for the development of algorithmic frameworks across different paradigms. Notably, Wang et al. (2024a) also discuss LLM profiling, but their focus is on personas, which are not relevant to the general roles that underpin these frameworks. We select four surveys according to their citations on Google Scholar for demonstration purposes, as shown in Table 1.

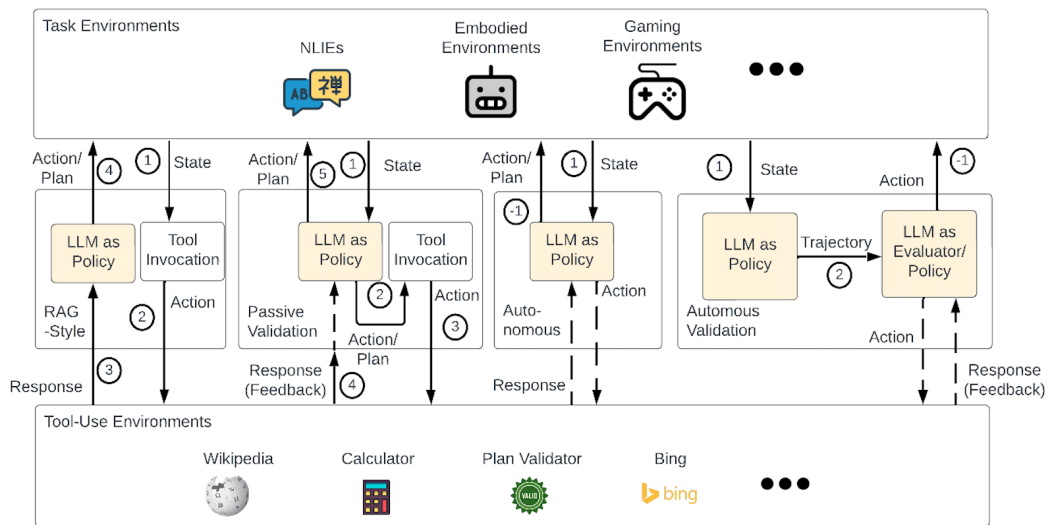
**Contributions** This survey offers the following contributions.

**1) Reviewing the breadth of environments the agents are applicable to (§2):** Before investigating LLM-based agents, it is essential to define the scope of universality by summarizing the environments in which these agents operate. We categorize the evaluated environments and tasks from the original manuscripts of the reviewed frameworks into two main types: **decision-making environments**

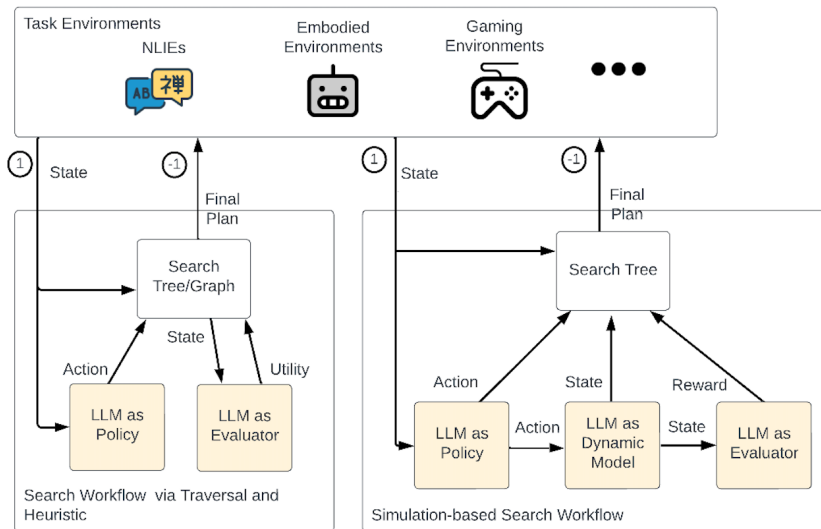


(a) Base Workflows.

(b) Workflows for Feedback Learning.



(c) Workflows for Tool Use, with validation types categorized under both Tool Use and Feedback Learning paradigms.



(d) Search Workflows for Planning.

Figure 1: Common workflows based on three LLM-Profiled Roles (LMPRs): Policy, Evaluator and Dynamic Model. Numbers in circles indicate the sequence of execution for each step. Unnumbered lines represent iterative steps. Tool-use steps autonomously determined by LMPRs are marked by dashed lines. Circles labeled “-1” in the circles indicate the final step. The ‘Tool Invocation’ and ‘Search Tree/Graph’ text boxes indicate that these processes/objects are maintained in agent programs.

	Tool Use		Planning		Feedback Learning		Universal LMPRs	
	Task-Specific Tools	Workflows	Direct Planning	Planning via Search	Workflows	Feedback Sources		Workflows
Guo et al. (2024)	✗	✗	✗	✗	✗	✓(\$3.4)	✗	✗
Xi et al. (2023)	✓(\$3.3)	✗	✓(\$3.1.4)	✗	✗	✓(\$3.1.4)	✗	✗
Huang et al. (2024)	✗	✗	✓(\$3)	✓(\$4)	✗	✓(\$6)	✗	✗
Wang et al. (2024a)	✓(\$2.1.4)	✗	✓(\$2.1.3)	✓(\$2.1.3)	✗	✓(\$2.1.3)	✗	✗
Ours	✗	✓(\$4.2)	✓(\$4.1)	✓(\$4.3)	✓	✓(\$4.4)	✓(\$4.4)	✓(\$3)

Table 1: Coverage of four highly cited surveys and ours. The section numbers of the original manuscripts for the covered perspectives are demonstrated for further reference. We acknowledge the coverage with at least one paragraph focus on that topic.

and **Natural Language Interaction (NLI) environments**, the latter framed from traditional Natural Language Processing (NLP) tasks for agent-based setups.

**2) Summarizing universal LLM-profiled roles and workflows (§3 and §4):** Workflow designs under planning and feedback-learning paradigms are often based on complex algorithms, such as depth- or breadth-first search (Yao et al., 2023a) and Monte Carlo Tree Search (Hao et al., 2023) for planning, as well as Reinforcement Learning (RL) for feedback learning (Shinn et al., 2023). The complexity increases when these algorithms are adapted to different environments. We aim to transcend the boundaries of these complex workflows by exploring those that can be universal across all types of environments. Additionally, we also investigate the workflows for tool use. Such exploration is based on three universal LLM-profile roles.

**3) Highlighting essential perspectives on LMPRs, workflows, and frameworks (§5 and §6):** We compare the implementations of LMPRs across various workflow types, emphasize workflows across different paradigms, and present alternative viewpoints to our formalization to prevent confusion. Additionally, we critically discuss three limitations of current frameworks and systematically identify potential workflow designs in future research.

## 2 Task Universality

This section explores task environments where various workflow types are applicable.

### 2.1 Feedback-Based, Decision-Making Environments

In this setting, actions yield clear environmental dynamics and rewards for agents to learn from. They are typical environments to evaluate planning and RL agents (Russell and Norvig, 2010; Sutton and Barto, 2018).

**Rule-Based Game Environments** These environments, which are deterministic and fully observable, include a variety of abstract strategy games like Chess and Go, and logic puzzles such as the Game of 24 (Yao et al., 2023a) and Blocksworld (Hao et al., 2023). They demand deep logical reasoning and strategic planning to navigate and solve.

**Embodied Environments** These settings simulate physical interactions and spatial relationships in the real world. They require agents to engage in navigation, object manipulation, and other complex physical tasks (Shridhar et al., 2021; Puig et al., 2018; Fan et al., 2022), reflecting changes in the physical environment.

### 2.2 Information Processing Environments

**Web Environments** The focus in web-based AI applications is more on processing information and user interaction. Webshop (Yao et al., 2022) necessitates a sequence of actions to interact with the environment, such as searching for a product, choosing a color, and clicking “buy.” AppWorld (Trivedi et al., 2024) demands a more complex control flow, based on rich programs.

Env Types	Entities Interacted With by Agent	Action Properties	Examples of Action Instances	Examples of Env
Game Environments	Virtual game elements (objects, avatars, other characters), and possibly other players or game narratives	Discrete, Executable, Deterministic	<code>Move(Right)</code>	BlocksWorld (Valmeekam et al., 2022), CrossWords (Yao et al., 2023a)
Embodied Environments	Physical world (through sensors and actuators)	Discrete, Executable, Deterministic	<code>Pick_Up[Object]</code>	AlfWorld (Shridhar et al., 2021), VirtualHome (Puig et al., 2018), Minecraft (Fan et al., 2022)
Web Environments	Virtual web elements	Discrete, Executable, Deterministic	<code>search(3 ounce bright citrus), click(Buy Now)</code>	Webshop (Yao et al., 2022), WebArena (Zhou et al., 2024b), AppWorld (Trivedi et al., 2024)
NLIEs	Humans (through conversation or text)	Free-form, Stochastic	Discrete, <code>The answer is Answer, Finish[Answer]</code>	GSM8K Cobbe et al. (2021), HotpotQA (Yang et al., 2018)

Table 2: Common Task Environments. An action instance is commonly formalized by **action predicates** and **action arguments**. NLIEs refer to Natural Language Interaction Environments.

### Natural Language Interaction Environments

Typical NLP tasks are often recontextualized as agentic environments in the study of LLM-based agents (Yao et al., 2023b; Hao et al., 2023; Yao et al., 2023a). In this survey, we refer to this setting as Natural Language Interaction Environments (NLIEs) for brevity. In NLIEs, the environment remains static until the agent acts. Unlike typical task environments where natural language serves as an intermediary, in NLIEs, both the states and actions are defined linguistically, making the states conceptual and the actions often ambiguous and broadly defined. Common setups include: **1) Single-Step NLIEs:** Some works (Yao et al., 2023b; Shinn et al., 2023) frame the traditional question-answering (QA) setup as a single-step decision-making process, where the agent generates an answer in response to a question. The process starts with the question as the initial state and concludes when the answer is provided as an action. Since each trial consists of just one step, there are no environmental dynamics or external feedback. Here, QA encompasses not only typical tasks (Cobbe et al., 2021) but also those often modeled as QA tasks, such as code optimization (Shypula et al., 2024). **2) Deliberate Multi-step NLIEs:** For tasks without naturally defined intermediate steps, several studies have transformed NLP tasks into a Markov Decision Process to facilitate agentic workflows, e.g., search and planning (Yao et al., 2023a; Hao et al., 2023). For example, Hao et al. (2023) reformulate subquestions in QA tasks as actions, enabling responses to user queries through

a multi-step process. This approach allows the initial question to serve as the beginning of a series of state transitions. Actions may vary from providing direct, free-form answers in single-step QA to strategically proposing subquestions that navigate the agent through sequential updates toward a comprehensive solution. Additionally, Wan et al. (2024) suggest that “splitting an output sequence into tokens might be a good choice” for defining multi-step NLIEs methodically. Furthermore, Yao et al. (2023a) formulate two-step NLIEs for creative writing by segmenting the problem-solving process into distinct planning and execution phases.

**Remark 1.** *In NLP, tasks represent the highest level of abstraction for modeling, whereas in decision-making processes, environments serve this role. It is common to describe an agent as being developed for a specific environment, such as an embodied environment, or for a particular task, such as QA. We refer to it as an NLIE-QA.*

### 3 LLM-Profiled Roles (LMPRs)

In this section, we demonstrate three common types of LLM-profiled Roles (LMPRs): policy models, evaluators, and dynamic models. They are task-agnostic and commonly used across various workflows.

**LLM-Profiled Policy  $glm_{\text{policy}}$**   $glm_{\text{policy}}$  is designed to generate decisions, which could be an action or a series of actions (plans) for execution in external environments or planning. In contrast to typical RL policy models, which learn to max-

imize cumulative rewards through trial and error, LLM-profiled policy models, denoted as  $\text{glm}_{\text{policy}}$ , utilize pre-trained knowledge and commonsense derived from extensive textual data. We distinguish between two types of  $\text{glm}_{\text{policy}}$ : an actor  $\text{glm}_{\text{actor}}$  directly maps a state to an action, whereas a planner  $\text{glm}_{\text{planner}}$  generates a sequence of actions from a given state.

**LLM-Profiled Evaluator  $\text{glm}_{\text{eval}}$**   $\text{glm}_{\text{eval}}$  provides feedback crucial for different workflows. During planning, it evaluates each step of actions or the resulting states (Hao et al., 2023; Yao et al., 2023a), and during feedback learning, they revise the entire decisions (Shinn et al., 2023; Wang et al., 2023b). Further details are provided in the next subsection.

**LLM-Profiled Dynamic Models  $\text{glm}_{\text{dynamic}}$**  They predict or describe changes to the environment. Generally, dynamic models form part of a comprehensive world model by predicting the next state  $s'$  from the current state  $s$  and action  $a$ . While typical RL uses the probability distribution  $p(s' | s, a)$  to model potential next states, LLM-based dynamic models directly predict the next state  $s' = \text{glm}_{\text{dynamic}}(s, a)$ .

## 4 LMPR-Based Workflows

We explore four types of workflows based on the three types of LMPRs, as illustrated in Figure 1. Table 3 summarizes the use of these workflows for different paradigms in the prior works.

### 4.1 Base Workflows

The base workflow is as simple as the interaction between  $\text{glm}_{\text{policy}}$  and the environment. These workflows can be categorized based on LLM profiling into two types: planners and actors. **1) Planners:** Many existing frameworks, such as those designed for embodied environments (e.g., Huang et al. (2022)), fall under this category. While some frameworks (Dasgupta et al., 2022; Wang et al., 2023b) involve complex interactions with task-specific components and low-level, non-LLM-based actors, their universal workflow remains as simple as the base workflow. For NLIEs, the complete workflow often follows this base model without additional interaction, as seen in Wang et al. (2023a). **2) Actors:** Early prompting frameworks for language generation tasks (classified as single-step NLIEs), such as Chain-of-Thought (Wei et al., 2022; Kojima et al., 2022), fit into this category.

For embodied tasks, ReAct (Yao et al., 2023b) employs  $\text{glm}_{\text{actor}}$ .

### 4.2 Tool-Use Workflows

We categorize two types of passive workflows and identify two types of autonomous workflows described in previous studies.

**RAG-Style Tool Use** A common example of passive tool use is Retrieval-Augmented Generation (RAG) (Lewis et al., 2020), commonly used in NLIE-QA tasks. In this setup, given a query, a retrieval mechanism collects relevant information to assist  $\text{glm}_{\text{policy}}$  in generating a response.

**Passive Validation** Guan et al. (2023) adopt an inverse approach for plan generation. Here,  $\text{glm}_{\text{policy}}$  first generates a plan, which is then validated by a separate tool. Depending on the validation outcome, the information may or may not be used to revise the initial plan generated by  $\text{glm}_{\text{policy}}$ .

**Autonomous Tool Use** In this paradigm, LLMs must be aware of the available tools, which requires including tool information during LLM profiling. The workflow must also handle signals from LMPR generation to invoke tools. Different methods can be applied to enable  $\text{glm}_{\text{policy}}$  to autonomously trigger tool usage. **1) In-Generation Triggers:** Tools could be invoked during the reasoning process (Inaba et al., 2023; Gou et al., 2024). The agent program monitors token generation and pauses when a tool trigger is detected. This pause allows the tool to be invoked, its output processed, and the results integrated into the reasoning process. Triggers are defined through tool descriptions, few-shot demonstrations<sup>1</sup>, or a combination of both<sup>2</sup>. **2) Reasoning-Acting Strategy:** Introduced by Yao et al. (2023b), each reasoning or acting step completes a full inference cycle, ending with the generation of a stop token. Hence, token-level monitoring is unnecessary. The workflow prompts explicitly for each acting step. **3) Confidence-Based Invocation:** Firstly,  $\text{glm}_{\text{policy}}$  generates an initial action, and the decision to invoke a tool is based on the confidence level of the generated tokens. Jiang et al. (2023) use this method for retrieval invocation, although it is not suitable for general tool use since it cannot specify which tool to invoke.

<sup>1</sup>See an example prompt in Table 14

<sup>2</sup>See an example prompt in Table 8



Types	Subtypes	Universal LMPRs	Used For	Related Frameworks
Base	$glm_{actor}$	$glm_{actor}$	/	ReAct (Yao et al., 2023b), CoT (Wei et al., 2022)
	$glm_{planner}$	$glm_{planner}$	Planning	Huang et al. (2022), DEPS (Wang et al., 2023b), Planner-Actor-Reporter (Dasgupta et al., 2022), Plan-and-solve (Wang et al., 2023a), OPEX (Shi et al., 2024a)
Tool-Use	RAG-Style (Passive)	$glm_{policy}$	Tool Use	RAG (Lewis et al., 2020; Shi et al., 2024b; Wang et al., 2024b; Zhang et al., 2024)
	Passive Validation	Tool Use, Feedback Learning	$glm_{policy}$	Guan et al. (2023)
	Autonomous	$glm_{policy}$	Tool-Use	MultiTool-CoT (Inaba et al., 2023), ReAct (Yao et al., 2023b), Active RAG Jiang et al. (2023)
	Autonomous Validation	$glm_{policy}$ , $glm_{eval}$	Tool Use, Feedback Learning	CRITIC (Gou et al., 2024)
Search	Traversal & Heuristic	$glm_{policy}$ , $glm_{eval}$	Planning	Tree-of-Thoughts (ToT) (Yao et al., 2023a), Tree-BeamSearch (Xie et al., 2023), Boost-of-Thoughts (Chen et al., 2024a), Graph-of-Thoughts (Besta et al., 2024), Tree-of-Traversal (Markowitz et al., 2024)
	Simulation-based (MCTS)	$glm_{policy}$ , $glm_{eval}$ , $glm_{dynamic}$	Planning	RAP (Hao et al., 2023), LLM-MCTS (Zhao et al., 2023), Wan et al. (2024), AgentQ (Putta et al., 2024), Chen et al. (2024b), Yu et al. (2023), Sprueill et al. (2023)
Feedback Learning	from $glm_{eval}$ only	$glm_{policy}$ , $glm_{eval}$	Feedback Learning	Reflexion (Shinn et al., 2023), Self-refine (Madaan et al., 2023), TextGrad (Yuksekgonul et al., 2024)
	from $glm_{eval}$ & Task Env	$glm_{policy}$ , $glm_{eval}$	Feedback Learning	Reflexion (Shinn et al., 2023)
	from Humans	$glm_{policy}$	Feedback Learning	CRITIC (Gou et al., 2024)

Table 3: Universal Workflows of LLM-Based Agents.

**Autonomous Validation** Gou et al. (2024) utilize  $glm_{policy}$  to generate an initial response. The resulting action(s) and the state(s) (i.e., a trajectory) are then passed to  $glm_{evaluator}$ , which autonomously determines whether tools should be invoked for validation.

**Remark 2.** *Tool-use workflows for validation can be viewed as a form of feedback learning, where  $glm_{policy}$  receives feedback from the tools.*

### 4.3 Search Workflows

**Traversal and Heuristic-Based Search** Generations from  $glm_{policy}$ , instead of direct execution in environments, are used to expand nodes for exploration, stored in a tree or graph structure, such as Tree-Of-Thoughts (ToT) (Yao et al., 2023a) and its variants (Chen et al., 2024a; Besta et al., 2024).  $glm_{eval}$  provides a fixed value estimate to select a node for further expansion. To expand a tree, ToT applies depth-/breadth-first search (DFS and BFS), while Xie et al. (2023) apply beam search. Notably, the BFS here is functionally equivalent to beam

search with  $N$  beams as the utility model  $glm_{eval}$  is used to maintain the  $N$  most promising nodes.<sup>3</sup>

**Simulation-Based Search** Simulation-based search for LLM-based planning agents is often carried out using the classic Monte Carlo Tree Search (MCTS) algorithm (Hao et al., 2023; Zhao et al., 2023; Putta et al., 2024). Similar to ToT (Yao et al., 2023a), a tree is built through search and is expanded with  $glm_{policy}$  and  $glm_{eval}$ . However, there are two key differences: 1) Node Selection: The nodes chosen for expansion are determined not only by the static outputs from  $glm_{eval}$  or other heuristics (whether they indicate goal attainment), but also by the cumulative statistics accrued over multiple simulations. Specifically, nodes that lead to better average rewards for subsequent nodes across all simulations (or trajectories) are indeed more likely to be expanded further. 2) Simulation : Following the selection and expansion phases.

<sup>3</sup>Typically, BFS does not rely on a utility model to decide which nodes to expand, since it systematically explores all possible nodes at each level until a terminal state.

A simulation phrase is required where  $\text{glm}_{\text{policy}}$ ,  $\text{glm}_{\text{dynamic}}$  and  $\text{glm}_{\text{eval}}$  are intimately collaborated, functioning as the roll-out policy. Specifically,  $\text{glm}_{\text{policy}}$  samples an action  $a_t$  given the current state  $s_t$ , which in turn, is assessed by  $\text{glm}_{\text{eval}}$ . The top-scoring action is selected, with  $\text{glm}_{\text{dynamic}}$  using it to derive  $s_{t+1}$ , iteratively simulating the trajectory.

#### 4.4 Feedback-Learning Workflows

Within feedback-learning workflows, feedback is fed into  $\text{glm}_{\text{policy}}$  for learning. One common source of feedback is  $\text{glm}_{\text{eval}}$ , as in Self-Refine (Madaan et al., 2023). Other common feedback sources include task environments (e.g., Reflexion (Shinn et al., 2023)), tools (Gou et al., 2024; Guan et al., 2023) and humans Guan et al. (2023). Among these,  $\text{glm}_{\text{eval}}$  can optionally be used to revise feedback with more contextualized information. In the original manuscript of Reflexion (Shinn et al., 2023),  $\text{glm}_{\text{eval}}$  corresponds to “self-reflection,” whereas the term “evaluator” refers to either heuristics or an LLM-profiled classifier that generates sparse feedback. However, this evaluator could be disregarded as a universal LMPR for two reasons: 1) Heuristics are mostly used and lead to better performance, and 2) the evaluator’s outputs eventually are fed to “self-reflection” for verbal feedback. When tools are employed to provide feedback (Gou et al., 2024; Guan et al., 2023), the workflow is the same as the tool-use workflow. In this setup, the necessity of invoking tools for feedback is either autonomously determined by  $\text{glm}_{\text{eval}}$  (Gou et al., 2024) (See Table 14 for an example) or hardcoded (Guan et al., 2023).

**Remark 3. Unrealistic Feedback Source:** So far, external feedback from NLIE-QA cannot be realistically generated. Some frameworks use the feedback-learning loop where the feedback source is unrealistic. Particularly, Shinn et al. (2023); Zhou et al. (2024a) use ground truths to generate feedback. However, we acknowledge that ground-truth feedback is appropriate when the framework serves as an evaluation baseline, as used by Yao et al. (2023a).

## 5 Discussions

**Prompting Methods for Profiling** In the previous subsection, we categorize the original works proposing Chain-of-Thought (CoT) prompting, including zero-shot CoT (Wei et al., 2022) and CoT

with few-shot demonstrations (Kojima et al., 2022), under the base workflow since the original work solves task directly via the base workflow. However, these methods themselves can be generalized to different types of LMPRs and workflows, as shown in Table 4. Some points of the specific use in agents should be highlighted: 1) For planner profiling, zero-shot CoT implementations often fail to produce long-horizon plans (Wang et al., 2023b). 2) To actor profiling under autonomous tool-use workflows, particularly with reasoning-acting strategies, the tool definitions are required to be included in the prompt; and few-shot demonstrations are important to give a clue of generation formats for LLMs including the indication of when to stop.

#### Workflow Comparisons for Plan Generation

Both the base workflow using  $\text{glm}_{\text{planner}}$  and search workflows generate a sequence of actions (i.e., a plan). However, they differ fundamentally in how the plans are generated and used. 1) **Greedy Generation vs. Exploration:** The base workflow leverages  $\text{glm}_{\text{planner}}$  to greedily generate a static plan in a single inference step. This approach often struggles with long-horizon plans for complex tasks (Sun et al., 2023). In contrast, search workflows explore multiple potential solutions and support backtracking, allowing for more robust exploration of options. 2) **Plan Execution:** The plans generated by  $\text{glm}_{\text{planner}}$ , BFS, DFS, and Beam Search are intended for full execution. However, this can lead to unexecutable actions, especially in the stochastic decision-making environments, if prior actions  $(a_1, \dots, a_{t-1})$  leads to a state where the next action  $a_t$  is invalid (e.g., ‘standing in front of the fridge’ but needing to ‘open the microwave’). This may cause abrupt interruption in the agent performance. In contrast, in the simulated-absed search via MCTS, only the action at the root node is executed in the actual environment. Subsequent simulated states and actions are discarded, though some implementations may partially retain these states to avoid recomputation. This search process repeats after every action taken, continuously recalculating the best action for the updated state.

**Workflow Comparisons for Using  $\text{glm}_{\text{actor}}$**  The  $\text{glm}_{\text{actor}}$  performs one of three types of actions, depending on the workflow: 1) In base and feedback-learning workflows,  $\text{glm}_{\text{actor}}$  consistently generates task-relevant actions to be executed within the task environment (**executable task actions**). 2) In a search workflow, actions generated by  $\text{glm}_{\text{actor}}$  are

	Prompting	Example Works	Example Prompts (in Appendix)
$glm_{actor}$	Few-shot	ReAct (Yao et al., 2023b), Reflexion (Shinn et al., 2023), RAP (Hao et al., 2023), MultiTool-CoT (Inaba et al., 2023)	Table 8, 9
$glm_{planner}$	Zero-shot	Plan-and-Solve (Wang et al., 2023a), LLM Planner (Huang et al., 2022)	Table 6
	Few-shot	DEPS (Wang et al., 2023b), Planner-Actor-Reporter (Dasgupta et al., 2022)	
$glm_{evaluator}$	Few-shot	RAP (Hao et al., 2023), Tree-BeamSearch (Xie et al., 2023), Reflexion (Shinn et al., 2023), CRITIC (Gou et al., 2024)	Table 11, 12
$glm_{dynamic}$	Few-shot	RAP (Hao et al., 2023)	Table 15

Table 4: CoT Prompting Methods for LLM-Profiled Roles. Note that the standalone CoT method can be categorized into the base workflow.

used for planning algorithms rather than direct execution (**planning task actions**). 3) In autonomous tool-use workflows,  $glm_{actor}$  is responsible for generating tool-related actions (**tool actions**).

**Workflow Comparisons for Using  $glm_{eval}$**  In feedback-learning workflows, the generation is passed to  $glm_{policy}$  for learning, while in search workflows, it is used for planning. This fundamental difference (learning vs. planning) leads to the following distinctions: **1) Feedback Use:** As shown in Figure 1, for learning, the output is generated for  $glm_{policy}$  to revise and regenerate the entire decision. In search workflows, the output is used to construct a search tree/graph for action selection and further expansion. **2) Feedback Forms:** These components— $glm_{policy}$  and search trees/graphs—require different forms of feedback.  $glm_{policy}$  processes free-form text, while MCTS utilizes continuous values as rewards, and trees/graphs in BFS, DFS, or beam search rely on discrete values for node selection. To meet these varying needs, distinct task formulations are employed, as summarized in Table 5. **a) Generating free-form text:**  $glm_{eval}$  is prompted to reflect on previous states and actions, generating reflective text as part of the  $glm_{policy}$  prompt in feedback-learning workflows (Shinn et al., 2023; Gou et al., 2024). **b) Binary/multiclass classification:**  $glm_{eval}$  is prompted with specific constraints to generate discrete output tokens, typically “no” or “yes.” The tokens are converted into scalar values (e.g., 0/1) for use as rewards in MCTS simulations (Hao et al., 2023), or to guide decision-making during tree traversal (Yao et al., 2023a). **c) Binary classification with scalar values:** This approach differs from the previous one by employing token logits generated by LLMs as scalar feedback values.

For example, the probability of a “yes” response is computed as:

$$v = \text{probs}_{\text{“yes”}} = \frac{e^{l_{\text{“yes”}}}}{e^{l_{\text{“yes”}}} + e^{l_{\text{“no”}}}},$$

where  $l_{\text{“yes”}}$  and  $l_{\text{“no”}}$  are the logits for “yes” and “no” tokens, respectively. <sup>4</sup> These scalar values can then be used as rewards in MCTS. **d) Multi-choice QA:** This formulation is used in scenarios that require selecting from multiple choices, such as choosing from top-N possible actions in traversal-based search workflows (Yao et al., 2023a).

## 6 Limitations and Future Works

### Limitation 1: A Unified Solution for Base Workflows and Autonomous Tool-Use Workflows

ReAct (Yao et al., 2023b) tries to unify the autonomous tool-use workflow and the base workflow via the reasoning-acting strategy. In this scenario, they implicitly unify tools as a part of task environments, and tool actions and task actions are unified to similar formats. However, the two workflows are still distinct: 1) specifying and comparing different tool-use workflows in a fine-grained manner, 2) distinguishing the two can avoid an illusion that it is universally implementable workflow. Specifically, for the latter, the sequence in which reasoning and action outputs alternate is task-dependent. For QA tasks, the generations of reasoning steps and tool actions are fixed, with alternating prompts for thinking and acting. <sup>5</sup> In contrast, for embodied tasks, the decision whether to proceed with thinking or

<sup>4</sup>Note that such implementations of  $glm_{eval}$  are less common due to their inaccessibility of state-of-the-art black-box LLMs.

<sup>5</sup>Reasoning-Acting Strategy (QA tasks): See an example prompt in Table 9



Task Formulation	Feedback Types	Applicable Workflows	Example Works
Text Generation	Free-form reflection	Feedback-learning workflows	Self-Refine (Madaan et al., 2023), Reflexion (Shinn et al., 2023), CRITIC (Gou et al., 2024)
Binary/Multi-class Classification	Discrete values	Search workflows	RAP (Hao et al., 2023), Tree-BeamSearch (Xie et al., 2023), ToT (Yao et al., 2023a), Koh et al. (2024)
Binary Classification	Continuous values (logits)	Search workflow via MCTS	RAP (Hao et al., 2023)
Multi-choice QA	Choices of top-N actions	Search workflows via traversal and heuristic	ToT (Yao et al., 2023a)

Table 5: Types of LLM-Profiled Evaluators According to Task Formulation and Feedback Types

acting in the next step is autonomously determined by  $\text{glm}_{\text{policy}}$ <sup>6</sup>.

**Limitation 2: Lack of Universal Tool-Use Workflow Design** Another future direction is the development of universal tool use. Although tool use is a universally applicable paradigm, current research tends to focus on using tools for specific tasks such as NLIE-QA or specialized purposes—either enabling  $\text{glm}_{\text{policy}}$  to retrieve new information or enabling  $\text{glm}_{\text{eval}}$  for validation.

**Limitation 3: Formulating Language Tasks as Agentic Tasks** The formulation can help develop a more universally used workflow, e.g., Reflexion (Shinn et al., 2023). However, some deliberate design should be reconsidered for future work, such as the use of ground truths as external feedback in Remark 3. Another questionable setting is deliberate multi-step NLIEs, e.g., what if a question is simple enough to be answered without the need to be decomposed into subquestions.

**Devising New Workflows** Our systematic review reveals potential directions for devising new workflows by intertwining existing ones. These workflows can be combined within a single paradigm—for instance, integrating different feedback sources or blending validation-based tool use with autonomous tool use. There is even greater potential for combining workflows across paradigms, such as incorporating optional feedback sources into non-validation tool-use workflows. In fact, validation-style workflows (Gou et al., 2024; Jiang et al., 2023) inherently merge elements of both the feedback-learning and tool-use paradigms.

<sup>6</sup>Reasoning-Acting Strategy (embodied tasks): See an example prompt in Table 7

## 7 Conclusion

This survey formalizes three common types of LLM-Profiled Roles (LMPRs) and investigates universal workflows for tool use, planning, and feedback learning. Since they are not task-specific, we hope that this will enlighten future research on employing LLM-based workflow designs across different tasks. Also, under the shuttle of these LM-PRs and workflows, we discuss some nuances of LLM profiling, workflow pros & cons, and framework implementations.

## Limitations

Due to the page limit, we can only focus on representative and pioneering work (primarily in \*ACL, ICML, ICLR, and NeurIPS) and the perspectives that align with our focus. That means some recent work may be missed, and task-specific components and complete workflow designs are mostly ignored, especially for those frameworks designed for embodied environments. For example, a visual model is normally required to translate pixel observations into textual inputs for  $\text{glm}_{\text{policy}}$  (Wang et al., 2023a). Again, this review aims to summarize the task-agnostic workflows to facilitate a coherent understanding and identify their limitations for future research.

## References

- Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. 2024. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17682–17690.
- Sijia Chen, Baochun Li, and Di Niu. 2024a. Boosting of thoughts: Trial-and-error problem solving with

- large language models. In *The Twelfth International Conference on Learning Representations*.
- Ziru Chen, Michael White, Ray Mooney, Ali Payani, Yu Su, and Huan Sun. 2024b. [When is tree search useful for LLM planning? it depends on the discriminator](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13659–13678, Bangkok, Thailand. Association for Computational Linguistics.
- Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*.
- Ishita Dasgupta, Christine Kaeser-Chen, Kenneth Marino, Arun Ahuja, Sheila Babayan, Felix Hill, and Rob Fergus. 2022. [Collaborating with language models for embodied reasoning](#). In *Second Workshop on Language and Reinforcement Learning*.
- Linxi Fan, Guanzhi Wang, Yunfan Jiang, Ajay Mandlekar, Yuncong Yang, Haoyi Zhu, Andrew Tang, De-An Huang, Yuke Zhu, and Anima Anandkumar. 2022. [Minedojo: Building open-ended embodied agents with internet-scale knowledge](#). In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.
- Roberto Gallotta, Graham Todd, Marvin Zammit, Sam Earle, Antonios Liapis, Julian Togelius, and Georgios N Yannakakis. 2024. Large language models and games: A survey and roadmap. *arXiv preprint arXiv:2402.18659*.
- Zhibin Gou, Zhihong Shao, Yeyun Gong, yelong shen, Yujia Yang, Nan Duan, and Weizhu Chen. 2024. [CRITIC: Large language models can self-correct with tool-interactive critiquing](#). In *The Twelfth International Conference on Learning Representations*.
- Lin Guan, Karthik Valmeekam, Sarath Sreedharan, and Subbarao Kambhampati. 2023. [Leveraging pre-trained large language models to construct and utilize world models for model-based task planning](#). In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xi-angliang Zhang. 2024. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*.
- Shibo Hao, Yi Gu, Haodi Ma, Joshua Hong, Zhen Wang, Daisy Wang, and Zhiting Hu. 2023. [Reasoning with language model is planning with world model](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8154–8173, Singapore. Association for Computational Linguistics.
- Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. 2024. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039*.
- Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. 2022. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *International Conference on Machine Learning*, pages 9118–9147. PMLR.
- Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. 2024. Understanding the planning of llm agents: A survey. *arXiv preprint arXiv:2402.02716*.
- Tatsuro Inaba, Hirokazu Kiyomaru, Fei Cheng, and Sadao Kurohashi. 2023. [MultiTool-CoT: GPT-3 can use multiple external tools with chain of thought prompting](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1522–1532, Toronto, Canada. Association for Computational Linguistics.
- Zhengbao Jiang, Frank Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yiming Yang, Jamie Callan, and Graham Neubig. 2023. [Active retrieval augmented generation](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7969–7992, Singapore. Association for Computational Linguistics.
- Jing Yu Koh, Stephen McAleer, Daniel Fried, and Ruslan Salakhutdinov. 2024. Tree search for language model agents. *arXiv preprint arXiv:2407.01476*.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. [Large language models are zero-shot reasoners](#). In *Advances in Neural Information Processing Systems*.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. [Retrieval-augmented generation for knowledge-intensive nlp tasks](#). In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc.
- Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. 2023. Self-refine: Iterative refinement with self-feedback. *arXiv preprint arXiv:2303.17651*.
- Potsawee Manakul, Adian Liusie, and Mark JF Gales. 2023. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. *arXiv preprint arXiv:2303.08896*.
- Elan Markowitz, Anil Ramakrishna, Jwala Dhamala, Ninareh Mehrabi, Charith Peris, Rahul Gupta, Kai-Wei Chang, and Aram Galstyan. 2024. [Tree-of-traversals: A zero-shot reasoning algorithm for augmenting black-box language models with knowledge](#)

- graphs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12302–12319, Bangkok, Thailand. Association for Computational Linguistics.
- Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. 2018. Virtualhome: Simulating household activities via programs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8494–8502.
- Pranav Putta, Edmund Mills, Naman Garg, Sumeet Ramesh Motwani, Chelsea Finn, Divyansh Garg, and Rafael Rafailov. 2024. [Agent q: Advanced reasoning and learning for autonomous ai agents](#).
- Stuart J Russell and Peter Norvig. 2010. *Artificial intelligence a modern approach*. London.
- Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548*.
- Haochen Shi, Zhiyuan Sun, Xingdi Yuan, Marc-Alexandre Côté, and Bang Liu. 2024a. [OPEX: A component-wise analysis of LLM-centric agents in embodied instruction following](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 622–636, Bangkok, Thailand. Association for Computational Linguistics.
- Weijia Shi, Sewon Min, Michihiro Yasunaga, Minjoon Seo, Richard James, Mike Lewis, Luke Zettlemoyer, and Wen-tau Yih. 2024b. [REPLUG: Retrieval-augmented black-box language models](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8371–8384, Mexico City, Mexico. Association for Computational Linguistics.
- Noah Shinn, Federico Cassano, Edward Berman, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. [Reflexion: Language agents with verbal reinforcement learning](#). *Preprint*, arXiv:2303.11366.
- Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Cote, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. 2021. [{ALFW}orld: Aligning text and embodied environments for interactive learning](#). In *International Conference on Learning Representations*.
- Alexander G Shypula, Aman Madaan, Yimeng Zeng, Uri Alon, Jacob R. Gardner, Yiming Yang, Milad Hashemi, Graham Neubig, Parthasarathy Ranganathan, Osbert Bastani, and Amir Yazdanbakhsh. 2024. [Learning performance-improving code edits](#). In *The Twelfth International Conference on Learning Representations*.
- Henry Sprueill, Carl Edwards, Mariefel Olarte, Udishnu Sanyal, Heng Ji, and Sutanay Choudhury. 2023. [Monte Carlo thought search: Large language model querying for complex scientific reasoning in catalyst design](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8348–8365, Singapore. Association for Computational Linguistics.
- Haotian Sun, Yuchen Zhuang, Lingkai Kong, Bo Dai, and Chao Zhang. 2023. [Adaplanner: Adaptive planning from feedback with language models](#). In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- Harsh Trivedi, Tushar Khot, Mareike Hartmann, Ruskin Manku, Vinty Dong, Edward Li, Shashank Gupta, Ashish Sabharwal, and Niranjana Balasubramanian. 2024. [AppWorld: A controllable world of apps and people for benchmarking interactive coding agents](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16022–16076, Bangkok, Thailand. Association for Computational Linguistics.
- Karthik Valmeekam, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. 2022. [Large language models still can’t plan \(a benchmark for LLMs on planning and reasoning about change\)](#). In *NeurIPS 2022 Foundation Models for Decision Making Workshop*.
- Ziyu Wan, Xidong Feng, Muning Wen, Ying Wen, Weinan Zhang, and Jun Wang. 2024. [Alphazero-like tree-search can guide large language model decoding and training](#).
- Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. 2024a. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):1–26.
- Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng Lim. 2023a. [Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2609–2634, Toronto, Canada. Association for Computational Linguistics.
- Siyuan Wang, Zhongkun Liu, Wanjun Zhong, Ming Zhou, Zhongyu Wei, Zhumin Chen, and Nan Duan. 2022. [From lsat: The progress and challenges of complex reasoning](#). *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 30:2201–2216.
- Xiaohua Wang, Zhenghua Wang, Xuan Gao, Feiran Zhang, Yixin Wu, Zhibo Xu, Tianyuan Shi, Zhengyuan Wang, Shizheng Li, Qi Qian, et al. 2024b. Searching for best practices in retrieval-augmented generation. *arXiv preprint arXiv:2407.01219*.



- Zihao Wang, Shaofei Cai, Guanzhou Chen, Anji Liu, Xiaojian Ma, and Yitao Liang. 2023b. [Describe, explain, plan and select: Interactive planning with LLMs enables open-world multi-task agents](#). In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed H. Chi, Quoc V Le, and Denny Zhou. 2022. [Chain of thought prompting elicits reasoning in large language models](#). In *Advances in Neural Information Processing Systems*.
- Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.
- Yuxi Xie, Kenji Kawaguchi, Yiran Zhao, Xu Zhao, Min-Yen Kan, Junxian He, and Qizhe Xie. 2023. [Self-evaluation guided beam search for reasoning](#). In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. [HotpotQA: A dataset for diverse, explainable multi-hop question answering](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2369–2380, Brussels, Belgium. Association for Computational Linguistics.
- Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. 2022. [Webshop: Towards scalable real-world web interaction with grounded language agents](#). *Advances in Neural Information Processing Systems*, 35:20744–20757.
- Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023a. [Tree of thoughts: Deliberate problem solving with large language models](#). *Preprint*, arXiv:2305.10601.
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. 2023b. [React: Synergizing reasoning and acting in language models](#). In *The Eleventh International Conference on Learning Representations*.
- Xiao Yu, Maximillian Chen, and Zhou Yu. 2023. [Prompt-based Monte-Carlo tree search for goal-oriented dialogue policy planning](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7101–7125, Singapore. Association for Computational Linguistics.
- Mert Yuksekgonul, Federico Bianchi, Joseph Boen, Sheng Liu, Zhi Huang, Carlos Guestrin, and James Zou. 2024. [Textgrad: Automatic "differentiation" via text](#). *arXiv preprint arXiv:2406.07496*.
- Peitian Zhang, Zheng Liu, Shitao Xiao, Zhicheng Dou, and Jian-Yun Nie. 2024. [A multi-task embedder for retrieval augmented LLMs](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3537–3553, Bangkok, Thailand. Association for Computational Linguistics.
- Zirui Zhao, Wee Sun Lee, and David Hsu. 2023. [Large language models as commonsense knowledge for large-scale task planning](#). In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Wanjun Zhong, Ruixiang Cui, Yiduo Guo, Yaobo Liang, Shuai Lu, Yanlin Wang, Amin Saied, Weizhu Chen, and Nan Duan. 2023. [Agieval: A human-centric benchmark for evaluating foundation models](#). *arXiv preprint arXiv:2304.06364*.
- Wanjun Zhong, Siyuan Wang, Duyu Tang, Zenan Xu, Daya Guo, Yining Chen, Jiahai Wang, Jian Yin, Ming Zhou, and Nan Duan. 2022. [Analytical reasoning of text](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2306–2319, Seattle, United States. Association for Computational Linguistics.
- Andy Zhou, Kai Yan, Michal Shlapentokh-Rothman, Haohan Wang, and Yu-Xiong Wang. 2024a. [Language agent tree search unifies reasoning acting and planning in language models](#).
- Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. 2024b. [Webarena: A realistic web environment for building autonomous agents](#). In *The Twelfth International Conference on Learning Representations*.

## A Examples Prompts

### A.1 LLM-Profiled Policy Models

**1) Base Workflow:** Prompts for a planner and an actor are demonstrated in Table 6 and 7, respectively. The content in the shadow box is generated by LLM. **2) Tool-Use Workflow:** Actors under NLIE-QA is demonstrated in Table 8 and 9. The former shows the implementation of in-generation triggers, while the last one demonstrates the ReAct implementation. **3) Search Workflow:** An actor is required during the expansion stage of MCTS in the RAP workflow. The prompts and expected generations are shown in Table 10.

### A.2 Prompts for LLM Evaluators

Table 11 and 12 show LLM evaluators that are profiled as classification tasks, while Table 14 demonstrates another to generate free-form text.

### A.3 Prompts as Dynamic Models

Table 15 is profiled as a dynamic model.



Q: In a dance class of 20 students, 20% enrolled in contemporary dance, 25% of the remaining enrolled in jazz dance, and the rest enrolled in hip-hop dance. What percentage of the entire students enrolled in hip-hop dance?

A: Let's first understand the problem and devise a plan to solve the problem. Then, let's carry out the plan and solve the problem step by step.

Plan:

Step 1: Calculate the total number of students who enrolled in contemporary and jazz dance.

Step 2: Calculate the total number of students who enrolled in hip-hop dance.

Step 3: Calculate the percentage of students who enrolled in hip-hop dance.

Solution:

Step 1: 20% of 20 students is 4 students. 25% of the remaining 16 students is 4 students. So, a total of 8 students enrolled in contemporary and jazz dance.

Step 2: The remaining 12 students enrolled in hip-hop dance.

Step 3: The percentage of students who enrolled in hip-hop dance is  $12/20 = 60\%$ .

Therefore, the answer (arabic numerals) is

60%

Table 6: An LLM planner under NLIE-QA, implemented by zero-shot CoT. The CoT trigger is highlighted in red font. The content in the shadow box is generated by LLM. Source from Wang et al. (2023a)

Interact with a household to solve a task. Here are two examples.

{Demo 1}

{Demo 2}

Here is the task: You are in the middle of a room. Looking quickly around you, you see a cabinet 6, a cabinet 5, a cabinet 4, a cabinet 3, a cabinet 2, a cabinet 1, a coffeemachine 1, a countertop 3, a countertop 2, a countertop 1, a drawer 3, a drawer 2, a drawer 1, a fridge 1, a garbagecan 1, a microwave 1, a shelf 3, a shelf 2, a shelf 1, a sinkbasin 1, a stoveburner 4, a stoveburner 3, a stoveburner 2, a stoveburner 1, and a toaster 1.

Your task is to: put a cool tomato in microwave.

>

think: To solve the task, I need to find a tomato, then cool it with the fridge, and finally put it in the microwave.

OK.

>

think: First, I need to find a tomato. A tomato is more likely to be found on the countertop (1-3), cabinet (1-6), drawer (1-3), shelf (1-3), or the sinkbasin. I can start by checking the countertop 1.

OK. >

go to countertop 1

On the countertop 1, you see a dish sponge 1, and a mug 2. ...

Table 7: An LLM actor within base workflow under the Alfred World benchmark, implemented by few-shot CoT. Source from (Yao et al., 2023b).

Use the following tools at your disposal and answer some questions:

- «Calculator»
- «Chemical reaction predictor»
- «Molar mass list»

Q: Find the amount of H<sub>2</sub>CO<sub>3</sub> formed on combining 2 moles of CO<sub>2</sub> and 2 moles of H<sub>2</sub>O.

A: First, use the chemical reaction predictor to determine the chemical equation for the reaction.

Reactants: 2CO<sub>2</sub>, 2H<sub>2</sub>O

Products: ?H<sub>2</sub>CO<sub>3</sub>

«Chemical reaction predictor»



Second, use the molar mass list to determine the molar mass of H<sub>2</sub>CO<sub>3</sub>.

H<sub>2</sub>CO<sub>3</sub> «Molar mass list» 62 g/mol

Third, use the calculator to calculate the amount of 2 moles of H<sub>2</sub>CO<sub>3</sub>.  $2 \times 62 =$  «Calculator» 124

Therefore, 124 g of H<sub>2</sub>CO<sub>3</sub> is formed.

Q: Find the amount of Calcium hydroxide that is required to react with 2 moles of Carbon dioxide to form 2 moles of Calcium carbonate along with 2 moles of Water.

A:

Table 8: An LLM actor within tool-use workflow. In-generation triggers are highlighted in red font. Source from MultiTool-CoT (Inaba et al., 2023)

Solve a question answering task by having a Thought, then Finish with your answer. Thought can reason about the current situation. Finish[answer] returns the answer and finishes the task.

Here are some examples:

**Question 1:** 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.

**Action:** Finish[1,800 to 7,000 ft]

...

**Question 6:** Were Pavel Urysohn and Leonid Levin known for the same type of work?

*Thought:* Let's think step by step. Pavel Urysohn is a mathematician. Leonid Levin is a mathematician and computer scientist. So Pavel Urysohn and Leonid Levin have the same type of work.

**Action:** Finish[Yes]

END OF EXAMPLES)

**Question:** VIVA Media AG changed its name in 2004. What does their new acronym stand for?

**Thought:**

Let's think step by step. VIVA Media AG changed its name in 2004. The new acronym must stand for the new name they adopted. To find the answer, we need to research the new name of VIVA Media AG after 2004.**Action:** Finish[Further research is needed to determine the new acronym.]

**Action:**

Finish[Further research is needed to determine the new acronym.]

Table 9: An LLM actor within tool-use workflow for the HotPot-QA benchmark, implemented by few-Shot CoT. Source from ReAct (Yao et al., 2023b).



Given a question, please decompose it into sub-questions. For each sub-question, please answer it in a complete sentence, ending with "The answer is". When the original question is answerable, please start the subquestion with "Now we can answer the question:"

**Question 1:** James writes a 3-page letter to 2 different friends twice a week. How many pages does he write a year?

**Question 1.1:** How many pages does he write every week?

*Answer 1.1:* James writes a 3-page letter to 2 different friends twice a week, so he writes  $3 * 2 * 2 = 12$  pages every week. The answer is 12.

**Question 1.2:** How many weeks are there in a year?

*Answer 1.2:* There are 52 weeks in a year. The answer is 52.

**Question 1.3:** Now we can answer the question: How many pages does he write a year?

*Answer 1.3:* James writes 12 pages every week, so he writes  $12 * 52 = 624$  pages a year. The answer is 624.

...

**Question 5:** Janet's ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder at the farmers' market daily for \$2 per fresh duck egg. How much in dollars does she make every day at the farmers' market?

**Question 5.1:**

How many eggs does Janet have left after eating three for breakfast and baking muffins with four?

Table 10: An LLM actor for the GSM8K benchmark. Source from [Hao et al. \(2023\)](#).

Given a question and some sub-questions, determine whether the last sub-question is useful to answer the question. Output 'Yes' or 'No', and a reason.

**Question 1:** Four years ago, Kody was only half as old as Mohamed. If Mohamed is currently twice as 30 years old, how old is Kody?

**Question 1.1:** How old is Mohamed?

**Question 1.2:** How old was Mohamed four years ago?

**New question 1.3:** How old was Kody four years ago?

*Is the new question useful? Yes. We need the answer to calculate how old is Kody now.*

...

**Question 5:** Janet's ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder at the farmers' market daily for \$2 per fresh duck egg. How much in dollars does she make every day at the farmers' market?

**New question 5.1:** Now we can answer the question: How much in dollars does she make every day at the farmers' market?

*Is the new question useful?*

Table 11: An LLM evaluator within simulation-based search workflow for NLIE-QA, implemented by few-shot CoT. It assesses the usefulness of new sub-questions in solving the original question. Source from [Hao et al. \(2023\)](#)

Evaluate if given numbers can reach 24 (sure/likely/impossible)

10 14

$10 + 14 = 24$

sure

...

1 3 3

$1 * 3 * 3 = 9$

$(1 + 3) * 3 = 12$

1 3 3 are all too small

impossible

11, 12

Table 12: An LLM evaluator within Tree-of-Thought Workflow under Game 24, implemented by few-Shot CoT prompting. The LLM is profiled for multi-class classification. Source from ToT (Yao et al., 2023a).

Context: ...

Sentence: ...

Is the sentence supported by the context above?

Answer Yes or No:

Table 13: An LLM evaluator. The LLM is profiled for multi-class classification. Source from Manakul et al. (2023).

{Few-shot Demonstrations Omitted for Brevity}

Question: Serianna is a band of what genre that combines elements of heavy metal and hardcore punk?

Proposed Answer: Let's think step by step. Serianna is a band of metalcore genre. Metalcore is a subgenre of heavy metal and hardcore punk. So Serianna is a band of heavy metal and hardcore punk. So the answer is: heavy metal and hardcore punk.

**1. Plausibility:**

The question asks for the genre that combines elements of heavy metal and hardcore punk, and the answer is "heavy metal and hardcore punk", simply repeat the question. So it's not plausible.

**2. Truthfulness:** Let's search the question in google:

> **Search Query:** Serianna is a band of what genre that combines elements of heavy metal and hardcore punk?

> Evidence:

[Metalcore - Wikipedia] Metalcore is a fusion music genre that combines elements of extreme metal and hardcore punk.

The evidence suggests that metalcore is a genre that combines elements of extreme metal and hardcore punk, as stated in the proposed answer.

Let's search the proposed answer in google:

> **Search Query:** Serianna is a band of metalcore genre.

> Evidence:

[Serianna - Wikipedia] Serianna was a metalcore band from Madison, Wisconsin. The band formed in 2006..

The evidence suggests Serianna is a metalcore band, the proposed answer is correct about this.

Above all, the proposed answer correctly identifies that Serianna is a band of the metalcore genre, which combines elements of heavy metal and hardcore punk. However, the final answer is not plausible since it just lists the genres that metalcore combines.

Table 14: An LLM evaluator within the Feedback-Learning workflow (feedback from tools). In-generation triggers are highlighted in **red font**, and tool-generated content is highlighted in **green font**. Source from Gou et al. (2024).

Given a question, please decompose it into sub-questions. For each sub-question, please answer it in a complete sentence, ending with "The answer is". When the original question is answerable, please start the subquestion with "Now we can answer the question: ".

**Question 1:** Weng earns \$12 an hour for babysitting. Yesterday, she just did 50 minutes of babysitting. How much did she earn?

**Question 1.1:** How much does Weng earn per minute?

**Answer 1.1:** Since Weng earns \$12 an hour for babysitting, she earns  $\$12 / 60 = \$0.2$  per minute. The answer is 0.2.

**Question 1.2:** Now we can answer the question: How much did she earn?

**Answer 1.2:** Working 50 minutes, she earned  $\$0.2 \times 50 = \$10$ . The answer is 10.

...

**Question 5:** Janet's ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder at the farmers' market daily for \$2 per fresh duck egg. How much in dollars does she make every day at the farmers' market?

**Question 5.1:** How many eggs does Janet have left after eating three for breakfast and using four for muffins?

**Answer 5.1:**

Table 15: An LLM-Profiled Dynamic Model.