

# AgentLite: A Lightweight Library for Building and Advancing Task-Oriented LLM Agent System

Zhiwei Liu\*, Weiran Yao, Jianguo Zhang, Liangwei Yang, Zuxin Liu, Juntao Tan  
Prafulla K. Choubey, Tian Lan, Jason Wu, Huan Wang, Shelby Heinecke  
Caiming Xiong, Silvio Savarese

Salesforce AI Research, USA

## Abstract

The booming success of LLMs initiates rapid development in LLM agents. Though the foundation of an LLM agent is the generative model, it is critical to devise the optimal reasoning strategies and agent architectures. Accordingly, LLM agent research advances from the simple chain-of-thought prompting to more complex ReAct and Reflection reasoning strategy; agent architecture also evolves from single agent generation to multi-agent conversation, as well as multi-LLM multi-agent group chat. However, with the existing intricate frameworks and libraries, creating and evaluating new reasoning strategies and agent architectures has become a complex challenge, which hinders research investigation into LLM agents. Thus, we open-source a new AI agent library, AgentLite, which simplifies this process by offering a lightweight, user-friendly platform for innovating LLM agent reasoning, architectures, and applications with ease. AgentLite is a task-oriented framework designed to enhance the ability of agents to break down tasks and facilitate the development of multi-agent systems. Furthermore, we introduce multiple practical applications developed with AgentLite to demonstrate its convenience and flexibility. Get started now at: <https://github.com/SalesforceAIResearch/AgentLite>.

## 1 Introduction

The AI agent emerges as a trending research topic along with the wide success of generative AI (OpenAI, 2023; Touvron et al., 2023). Among the outstanding innovations, the LLM agent augments an large language model (LLM) with tool usage (Patil et al., 2023; Schick et al., 2023; Qin et al., 2023) and environment interaction ability (Liu et al., 2023a; Deng et al., 2023a; Zhou et al., 2023). As such, an agent is capable of collecting real-time feedback from environments to close the interaction loop (Liu et al., 2023b; Yao et al., 2023a; Xu et al., 2023), which expands the task-solving capability of LLM from one-time generation to multi-step action execution (Deng et al., 2023b; Zheng et al., 2024; Wu et al., 2023). Therefore, LLM agents exhibit promising performance in various LLM applications, such as BabyAGI (Nakajima, 2023), AutoGPT (Gravitas, 2023), agent world (Park et al., 2023) and etc (Wu et al., 2023; Chase, 2023).

Though the core of the LLM agent is the generative model for information understanding and action generation, the reasoning strategies and agent architectures are also rather crucial for enhancing performance. Regarding the reasoning strategies, ReAct (Yao et al., 2023a) exhibits higher execution correctness when activating one-step of *Think* in agent action execution. Also, enabling reflection (Shinn et al., 2023; Yao et al., 2023b) reasoning enhances the performance of agents. ReWOO (Xu et al., 2023) decouples reasoning and observation in agent execution. Divergent Think (Wang et al., 2023) is also verified as one effective reasoning strategy for agents to explore more action spaces.

\*zhiweiliu@salesforce.com

| Library                    | AgentLite | AutoGen | LangChain | Camel | CrewAI |
|----------------------------|-----------|---------|-----------|-------|--------|
| Task Decomposition         | ✓         | -       | -         | ✓     | ✓      |
| Multi-Agent Orchestration  | ✓         | ✓       | -         | ✓     | ✓      |
| Extendable Reasoning Types | ✓         | -       | ✓         | -     | -      |
| Memory Module              | ✓         | -       | ✓         | ✓     | -      |
| Prompter Module            | ✓         | -       | ✓         | ✓     | -      |
| # (Core) Code Lines        | 959       | 8,966   | 248,650   | 4,987 | 1,504  |

Table 1: A comparison of features supported by AgentLite library vs existing libraries for agent building. ✓ denotes inherently supported in the library while - represents not modulated or hard to refactor.

Besides, the optimal agent architecture is yet not well established. Langchain (Chase, 2023) evolves rapidly as one of the AI agent libraries, simplifying customized tool agent development. BOLAA (Liu et al., 2023b) compares different agent architectures and proposes to use a multi-agent framework to decompose the complexity of tasks. AutoGen (Wu et al., 2023) develops a novel multi-agent conversation mechanism and leverages code execution as actions.  $\alpha$ -UMi (Shen et al., 2024) introduces a multi-LLM agent. It fine-tunes multiple small LLMs in a multi-agent framework, which even outperforms GPT-4. Therefore, more exploration (Liu et al., 2023a; Zeng et al., 2023) in regards to agent architecture designs are essential to advance the developments of LLM agents.

In this sense, speeding up the implementation and validation of new reasoning types and agent architecture designs consolidates the fast development of LLM agents. As such, a lightweight, easy-to-use library for efficiently prototyping LLM agent design is necessary. Nevertheless, the existing libraries (Chase, 2023; Wu et al., 2023; Gravitas, 2023; Li et al., 2023), though showing great progress, do not target at those fundamental designing of agents. For example, Langchain, as one of the representative libraries for developing LLM applications, provides built-in interfaces to initialize different types of agents. However, Langchain is over-complicated for researchers to implement new agent reasoning types and architectures. Moreover, refactoring Langchain library for new research scenarios is rather challenging due to a lot of overhead in its agent creation. Autogen, though achieving success in empowering LLM agent building, has hard-coded reasoning types in its agent interface, which requires a lot of effort for refactoring. Also, its agent architecture is fixed as multi-agent conversation and code execution, which may not be suitable for new research scenarios and benchmarks.

To this end, we introduce a new AI agent building library, *i.e.* AgentLite. AgentLite is a light research-oriented library for developing and implementing advanced LLM-based agent/multi-agent architectures. Building an agent with customized tools in AgentLite is easy and straightforward. AgentLite also enables the easy orchestration of multiple agents via a manager agent. The key features of AgentLite are as follows:

- **Lightweight Code Architecture:** AgentLite is characterized by its lean codebase, necessitating minimal external dependencies, thereby affording researchers unparalleled flexibility in customizing various agent components such as prompts, memory, action sequences, and Large Language Models (LLMs). AgentLite remains both accessible and adaptable to a wide range of research needs.
- **Task-Oriented Design Principle:** At the core of AgentLite is its focus on task-oriented agent development. From the moment of initialization, each agent is endowed with a specific task, setting the stage for intricate task-based interactions. This foundational principle not only simplifies the design and deployment of individual agents but also serves as the basis for the construction of sophisticated multi-agent systems, facilitating seamless communication and collaboration among agents within the AgentLite ecosystem. Therefore, AgentLite has great potential to develop a scalable framework for complex agent interactions.
- **Hierarchical Multi-Agent Orchestration:** AgentLite is setup upon a structured multi-agent systems. This is achieved by assembling a team of agents, and each is governed by a manager agent responsible for orchestrating their operations. Such an arrangement

endows the multi-agent system with a hierarchical structure, wherein the actions of team agents are coordinated to fulfill broader task objectives.

- **Versatile Implementation Potential:** AgentLite distinguishes its capacity for benchmarking evaluations and deployment in real-world scenarios. AgentLite not only demonstrates its practical applicability in a variety of contexts but also underscores its utility in advancing the state-of-the-art in LLM agent applications.

A comparison of existing agent building library is presented in Table 1, including AutoGen (Wu et al., 2023), LangChain (Chase, 2023), Camel (Li et al., 2023) and CrewAI<sup>1</sup>. AgentLite achieves comprehensive abilities of LLM agents with less than 1k lines of code.

## 2 AgentLite Framework

We introduce the AgentLite framework by illustrating the general steps to implement customized Individual Agent and Manager Agent, our two cornerstones in the framework, for building a multi-agent system. The overview of our framework is illustrated in Figure 1.

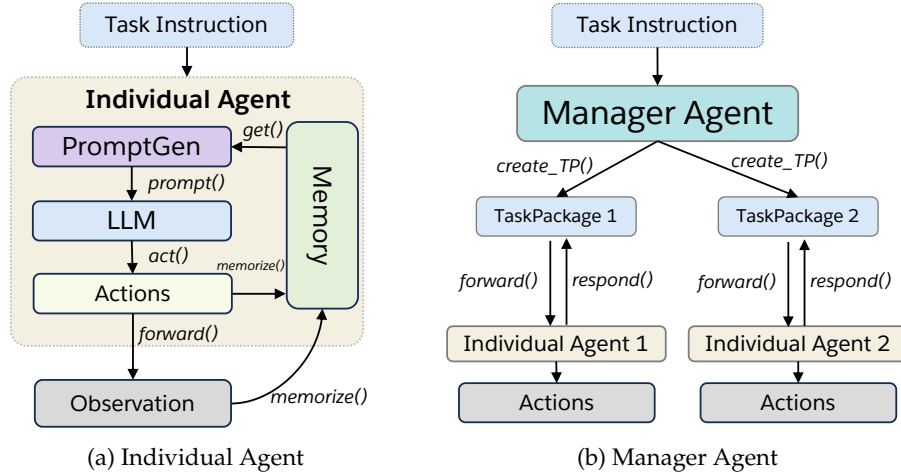


Figure 1: The AgentLite framework for Individual Agent and Manager Agent. (a) Individual Agent is the base agent class in AgentLite. It constitutes four modules: *i.e.* *PromptGen*, *LLM*, *Actions* and *Memory*. The *PromptGen* composes a prompt for LLM from the memory of the agent. LLM generates an executable action, which is forwarded to get observation. Both action and observation are memorized into the memory. (b) Manager Agent is a subclass of an individual agent with more methods in TaskPackage (TP) creation. A manager agent creates sub-tasks and corresponding TPs from the given task instruction. Those TPs are forwarded to associated individual agents sequentially.

### 2.1 Individual Agent

Individual Agent is the base agent class in AgentLite. It is built upon four modules, *i.e.*, *PromptGen*, *Actions*, *LLM*, and *Memory*, which are illustrated in Figure 1a.

**PromptGen** is a module to construct the prompt that will be sent to LLM to generate the action. The agent prompt is composed of multiple components, including agent role description, agent instructions, constraint for generation, agent actions, and examples. AgentLite supports default methods to combine those components. Additionally, AgentLite allows developers to design customized prompts for action generation.

**Actions** are classes that an agent has and can automatically execute. AgentLite provides a *BaseAction* as a wrapper class for developers to subclass their actions, code as follows:

<sup>1</sup><https://github.com/joaomdmoura/crewAI/>

```

class BaseAction:
    def __init__(
        self,
        action_name: str, # the name of this action
        action_desc: str, # the detailed explanation of this action
        params_doc: dict # params documentation for __call__()
    ) -> None:
        self.action_name = action_name
        self.action_desc = action_desc
        self.params_doc = params_doc

    def __call__(self, **kwargs) -> str:
        """
        implement the Action call
        """
        raise NotImplementedError

```

*action\_name*, *action\_desc* and *params\_doc* are the required properties for the agent to understand how to use this action. Specifically, the *params\_doc* contains the key-value pairs to describe all the input parameters for the *\_\_call\_\_()* method. Any customized agent actions should subclass this *BaseAction* and overwrite the *\_\_call\_\_()* method to return a string observation. Note that the *action\_name* could be different from the actual class name. In this way, developers could change it more flexibly for various scenarios. During execution, LLM generates both an action name and its input parameters, which are used to identify the executable actions and its function call input. Different actions of agents also lead to different reasoning abilities. For example, an agent with *Think* action is equivalent to the ReAct reasoning strategy.

LLM is based on API call method. AgentLite has a simple *BaseLLM* wrapper class, which is called with an input string and returns another string generated from LLM.

**Memory** stores the action-observation chains of an agent. During execution, an agent gets its historical actions and corresponding observations, which are concatenated and input to PromptGen. All the generated actions and returned observations are immediately saved into memory.

Initializing an individual agent in AgentLite is then as simple as the following code snippet:

```

from agentlite.agents import BaseAgent

name = "agent_name"
role = "describe the roles of this agent"
actions = [Action1, Action2]
agent = BaseAgent(name=name, role=role, actions=actions)

```

A typical working flow of an individual agent is as follows: firstly, the user gives task instructions to an agent. Then, the agent uses its PromptGen to construct the prompt, which retrieves the historical action-observation chain regarding this task if there is more than one step execution. After getting the full prompt, LLM generates an action for the agent to execute. And the observation from this action is memorized into memory. Note that the *BaseAgent* has default *Finish* action, which is an action to complete the current task and return a response. We also develop a default *Think* action to enable the reasoning step.

## 2.2 Manager Agent

*ManagerAgent* is another core agent class in AgentLite. It subclasses from *BaseAgent* with more ability to communicate with other agents. The manager agent is designed to endow the multi-agent system development in AgentLite. We design this multi-agent system to be hierarchical. Specifically, one manager agent is able to decompose a task into sub-tasks and send them to different individual agents. As such, each individual agent is interpreted as one type of action of this manager agent when adding to the team of this manager agent.

Therefore, this multi-agent could be designed with a deeper hierarchy if setting one manager agent to be a team member of another manager agent.

Initializing a manager agent in AgentLite is simple and straightforward. Developers directly pass in other agents as the *TeamAgents* of this manager agent. For example,

```
from agentlite.agents import ManagerAgent

# assuming we have already designed three agents
# as a team for manager agents to control
team = [agent_1, agent_2]
name = "manager_agent"
role = "controlling multiple agents to complete task"
manager = ManagerAgent(name=name, role=role, TeamAgents=team)
```

A typical working flow of a manager agent is presented in Figure 1b. The user sends a task instruction to a manager agent. Then, each action of the manager agent is to assign a task to one of its team agents to complete. In this step, a manager agent creates a *TaskPackage*, forwards this *TaskPackage* to this team agent and waits for the response from this team agent. *TaskPackage* subclasses a common *pydantic.BaseModel* in AgentLite, which has the following main properties:

```
class TaskPackage(BaseModel):
    instruction: str # task instruction
    completion: str # completion status
    creator: str # agent name that creates this task
    timestamp: str
    answer: str # response for this task
    executor: str # agent name that executes this task
```

A Task Package (TP) serves as a communication basis between a managing agent and its subordinate team agents. The primary role of the managing agent is to concentrate on the generation and delegation of sub-tasks to these team agents. Upon receiving a sub-task, a designated agent is responsible for resolving this sub-task and subsequently communicates the outcome back to manager agent. This modular approach facilitates the assembly of a multi-agent system, where a developer can incorporate multiple specialized agents under the supervision of a managing agent to address complex tasks requiring diverse actions. Harnessing the collective capabilities of its components enhances its efficiency and brings a lot of potentials to a scalable multi-agent system. Note that the generation of TPs follows a sequential order, with the creation of a subsequent TP predicated on the feedback received from the execution of its predecessor. This sequential process ensures that the system dynamically adapts and responds to the evolving requirements of the initial task.

### 3 Agent Development

In this section, we introduce how researchers could use AgentLite as code base to develop their new reasoning type and architecture.

#### 3.1 New Reasoning Type

Adding a new reasoning type in AgentLite is similar to adding a new action into an agent, which is easy to implement. For example, an agent with the reasoning type as direct *Act* has actions for execution. In comparison, an agent with ReAct (Yao et al., 2023a) reasoning is taking one *Think* step before actual execution. In AgentLite implementation, we unify those agents **inner actions** like *Think* and *Plan*, and other executable actions like *Search* as the same *Action* module for the agent to call. A sample implementation of a *Think* action in AgentLite is as follows<sup>2</sup>:

<sup>2</sup>The codes are simplified for legibility.

```

class ThinkAction(BaseAction):
    action_name = "Think"
    action_desc = "Conduct thinking..."
    params_doc = {
        INNER_ACT_KEY: "this is your thinking response.."
    }

    def __call__(self, **kwargs):
        return DEF_INNER_ACT_OBS

```

Enabling ReAct reasoning of agent is then simplified as adding this new *ThinkAction* into the action of an agent as:

```
agent.actions += [ThinkAction]
```

To be more specific, LLM uses the *action\_name* and *action\_desc* to understand this action. Developers are also able to change them to different prompts. We formulate the actual reasoning process of LLM to generate the parameters of *ThinkAction*. In this way, the reasoning process and the tool usage execution are unified. We could also develop *Plan* and *Reflect* actions analogously.

### 3.2 New Agent Architecture

AgentLite also supports new Agent Architectures, which can be easily configured within lines of Python code. It provides sufficient flexibility for users to design agents. This section illustrates two types of new Agent Architectures with AgentLite.

**Copilot Agent.** A Copilot Agent is able to receive human instructions and complete tasks accordingly, which is useful for a wide range of practical applications. It can be easily built by defining a *HumanInstruction* action, and adding the new action to the agent. The new action should inherit from the *BaseAction* class in AgentLite, and the code is given as follows:

```

from agentlite.actions import BaseAction

class HumanInput(BaseAction):
    action_name = "HumanInput"
    action_desc = "Obtain instruction from human."
    params_doc = {"question": "Questions for human."}

    def __call__(self, question):
        instruction = input(question)

        return f"My instruction is: {instruction}:"

```

By adding the newly-defined *HumanInput* Action to the Agent, we can obtain a Copilot Agent that is able to acquire instruction from humans. The code for building a Copilot Agent is as follows:

```

from agentlite.agents import BaseAgent

name = "Copilot_Agent"
role = "Copilot agent that is able to receive human instructions."
actions = [HumanInput(), Action_1, Action_2, ...]
copilot_agent = BaseAgent(name=name, role=role, actions=actions)

```

**Copilot Multi-Agent.** AgentLite also supports an arbitrary number of agents for multi-agent orchestration. For example, a Copilot Multi-Agent is a multi-agent system with a human agent to collect instructions from human input. Designing this agent structure with

AgentLite is realized via adding a *HumanAgent* into a manager agent. A simple *HumanAgent* subclasses from an abstract agent class in AgentLite, i.e. *ABCAgent*.

```
from agentlite.agents import ABCAgent

class HumanAgent(ABCAgent):
    name = "Human_Agent"
    role = "Collect instructions from human."

    def __call__(self, question):
        human_instruction = input(question)

        return f"My instruction is: {instruction}"
```

Then a copilot multi-agent is defined as a manager agent to control other agents with this human agent as a team. A sample code block is as follows:

```
team = [HumanAgent(), agent_1, agent_2, ...]
name = "manager_agent"
role = "controlling a human agent, agent_1, agent_2, ..."
constraints = "should follow human instructions to complete task"
copilot_multiagent = ManagerAgent(name=name, role=role, constraints=
                                constraints, TeamAgents=team)
```

**Multi-LLM Multi-Agent.** AgentLite also supports different LLMs as backbones for different Agents. Before we initialize the agent, we should provide different LLM configs<sup>3</sup>. Then, an agent with a specific LLM is defined as follows:

```
from agentlite.llm import LLMConfig
# Define a specific LLM agent
llm_config_dict = {"llm_name": "LLM_1"}
llm_config = LLMConfig(llm_config_dict)
llm = get_llm_backend(llm_config)
agent_1 = BaseAgent(llm, name=name, role=role, actions=actions)
```

Assuming we have multiple agents with different LLMs, we finally build them as a team into a manager agent as follows:

```
team = [agent_1, agent_2, agent_3, ...]
multi_LLM_multi_agent = ManagerAgent(team=team)
```

Note the LLM for this ManagerAgent is also configurable through this LLMConfig.

## 4 Benchmark

In this section, we implement specific agents with AgentLite to evaluate its performance on different agent benchmarks Liu et al. (2023b); Ma et al. (2024).

### 4.1 Retrieval-Augmented Question Answering

The HotPotQA dataset (Yang et al., 2018) is a benchmark for evaluating the capabilities of question-answering (QA) systems in the context of multi-hop reasoning across multiple documents, which is adopted as a benchmark for evaluating LLM agent Liu et al. (2023b). The performance of models on the HotPotQA dataset is quantitatively assessed using two primary metrics: **F1-Score** and **Accuracy**.

<sup>3</sup>Those LLMs should be running in the backend.

In AgentLite, we devise the action space for this task as three main actions: *WikipediaSearch*, *Think*, and *Finish*. The *WikipediaSearch* action involves querying the Wikipedia dump to retrieve relevant articles or passages. The *Think* action represents the model’s internal processing to integrate the retrieved information, reason about it, and formulate a hypothesis toward the answer. Finally, the *Finish* indicates the model’s decision to conclude the search and reasoning process and output the final answer. The results comparison on different models is reported in Table 2.

Table 2: Performance of AgentLite on HotPotQA Dataset

| LLM                    | Easy         |             | Medium       |             | Hard         |             |
|------------------------|--------------|-------------|--------------|-------------|--------------|-------------|
|                        | F1-Score     | Accuracy    | F1-Score     | Accuracy    | F1-Score     | Accuracy    |
| GPT-3.5-Turbo-16k-0613 | 0.410        | 0.35        | 0.330        | 0.25        | 0.283        | 0.20        |
| GPT-4-0613             | 0.611        | <b>0.47</b> | 0.610        | 0.48        | <b>0.527</b> | <b>0.38</b> |
| GPT-4-32k-0613         | <b>0.625</b> | 0.46        | <b>0.644</b> | <b>0.54</b> | 0.520        | 0.37        |
| xLAM-v0.1              | 0.532        | 0.45        | 0.547        | 0.46        | 0.455        | 0.36        |

Our experiments on the HotPotQA dataset with models including GPT-3.5-Turbo-16k-0613, GPT-4-0613, GPT-4-32k-0613, and our xLAM-v0.1 model. xLAM-v0.1 (Zhang et al., 2024) is our advanced large action model (LAM) trained on agent action trajectories from multiple environments, which is built based on the Mixtral 8x7b Mix-of-Expert LLM (Jiang et al., 2024). The evaluation are reported across easy, medium, and hard difficulty levels, demonstrate a clear trend of performance improvement with the evolution of model architectures and capacities. The GPT-4 variants outperform the GPT-3.5 model across all difficulty levels, with the 32k variant of GPT-4 showing superior performance in the medium difficulty set. Our xLAM-v0.1 model also outperforms GPT-3.5, which indicates the necessity of fine-tuning an LLM on diverse action trajectories for improving agent performance. This also demonstrates AgentLite is adaptable for arbitrary LLMs.

#### 4.2 Web Browser Decision Making

We also implement an agent with AgentLite for Webshop environment. WebShop (Yao et al., preprint) is an simulation online shopping website environment with 1.18M real-world products. The environment also provides human instructions as shopping tasks. Each instruction is associated with one ground-truth product, and contains attribute requirements. Most existing agent evaluation adopts this environment for benchmarking (Liu et al., 2023b;a; Ma et al., 2024). We follow AgentBoard (Ma et al., 2024) environment setup to design the webshop agent in AgentLite. Implementing a webshop agent with AgentLite is simple and straightforward. We only need to wrap up both actions with the *BaseAction* class and insert them into the *BaseAgent* class. The webshop environment also provides a final reward for shopping. We employ the same test tasks as in the AgentBoard, including a total of 251 instances. Each task is classified into one of two categories: easy (182 tasks) or hard (69 tasks). In Table 3, we report the average rewards obtained for the easy, hard, and overall tasks when various models are applied. GPT-4-32k performs much better than other LLMs, especially on hard tasks. The reason is that its longer context length enables more information understanding ability. Our xLAM model performs comparable with GPT-3.5 with AgentLite implementation.

Table 3: Performance of AgentLite on Webshop Environment

| LLM                    | avg. reward  |              |              |
|------------------------|--------------|--------------|--------------|
|                        | easy         | hard         | all          |
| GPT-3.5-Turbo-16k-0613 | 0.528        | 0.506        | 0.522        |
| GPT-4-0613             | <b>0.676</b> | 0.674        | 0.676        |
| GPT-4-32k-0613         | 0.675        | <b>0.696</b> | <b>0.681</b> |
| xLAM-v0.1              | 0.532        | 0.512        | 0.524        |



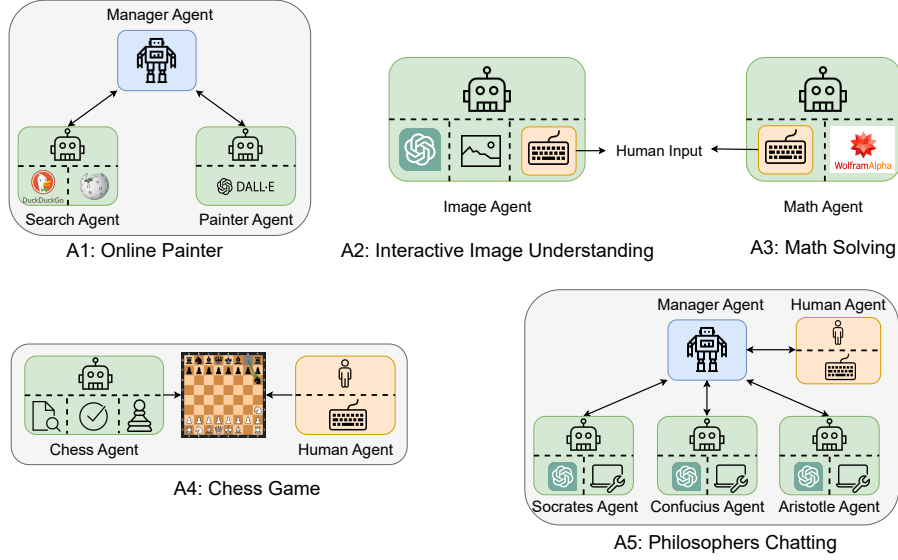


Figure 2: Application illustrations of AgentLite.

## 5 Applications

In this section, we present various applications built via AgentLite to show the simplified design principle and innovative application capacity. These examples showcase varied characteristics including Practical Application (A3, A4), Human-in-the-Loop (A1, A2), Multi-modality (A1, A2), Innovative Potential (A5).

### 5.1 A1: Online Painter

In this subsection, we illustrate the Online Painter application as shown in Fig. 2(A1). Given any object, Online Painter is able to search relevant visual features online first and then paint an image of the object based on the detailed features.

The Online Painter application is completed by three agents (Manager Agent, Search Agent, Painter Agent). For example, the Painter Agent in AgentLite is initialized as follows:

```
name = "Painter_agent"
role = "You are a painter and can draw a picture with Paint action."
actions = [Paint()]
PaintAgent = BaseAgent(name=name, role=role, actions=actions)
```

Search Agent can be defined analogously. Then, we could build one manager agent to control them as follows:

```
manager = ManagerAgent(TeamAgents=[PaintAgent, SearchAgent])
```

The three agents all have the *Think* action for reasoning and *Finish* action to finish the task. Manager Agent receives tasks directly with the Task Package defined in Sec. 2.2 and dismantles the task to different steps for its team members consisting of Search Agent and Painter Agent. The Search Agent has two actions to use for collecting information. *DuckSearch* collect online information via DuckDuckGo<sup>4</sup> API based on the query. And *WikipediaSearch* is able to find relevant information on Wikipedia<sup>5</sup>. These two actions endow

<sup>4</sup><https://duckduckgo.com/>

<sup>5</sup>[https://en.wikipedia.org/wiki/Main\\_Page/](https://en.wikipedia.org/wiki/Main_Page/)

the Search Agent with the ability to obtain information about the painted object. (3) The Painter Agent have one DALLE<sup>6</sup> action to paint the picture given the description.

## 5.2 A2: Interactive Image Understanding

Interactive Image Understanding is a multi-modality application with human-in-the-loop instructions. Given an image, the application can provide answers to the questions from the human based on the image. Humans can give questions for several rounds to the application with natural languages such as *"What is in the image?"*, *"What is the color of the bridge?"*, etc. The application can also be terminated with ending instructions from humans such as *"I good for today, have a great weekend."*, as real-life conversations.

The application is a copilot agent as described in Sec 3.2. The *ImageDisplay* action can show the image to a human user and *ImageQuery* action is able to answer questions based on the image. The *ImageQuery* action is backed by the gpt-4-vision-preview api<sup>7</sup>. Empowered by the *HumanInput* action, the Image Agent is able to acquire and follow instructions from humans. The code for building the Image Agent is as follows:

```
from agentlite.agents import BaseAgent
name = "Image_agent"
role = "You are image agent to view and answer questions from image."
actions = [HumanInput(), ImageQuery(), ImageDisplay()]
ImageAgent = BaseAgent(name=name, role=role, actions=actions)
```

## 5.3 A3: Math Problem Solving

Math Problem Solving is an application built simply with a Math Agent via AgentLite. The Math Agent receives and solves a math problem from the human input. Human inputs can be either direct math equations as *"75\*34+12="* or descriptions of the problem as *"What is the result of 75 multiplied by 34 and then plus 12?"*. We use AgentLite to develop a Math Copilot Agent with *HumanInput* and *WolframAlphaSolver* actions. The Math Agent acquires math questions from humans with *HumanInput* action, and then solve the problem with *WolframAlphaSolver* action. It is integrated with the *WolframAlpha* API<sup>8</sup>, which is able to solve a wide range of math problems from basic operations to equation solving and calculus. The code for building the Math Problem Solving Copilo Agent is as follows:

```
from agentlite.agents import BaseAgent
name = "MathAgent"
role = "You can answer math questions by WolframAlphaSolver action."
actions = [HumanInput(), WolframAlphaSolver()]
MathAgent = BaseAgent(name=name, role=role, actions=actions)
```

This is an example to show how to build applications directly with the individual agent. In AgentLite, the Manager Agent is not required for development. Instead, it depends on the application's complexity and scale. For the simple Math Problem Solving application, an individual agent is sufficient. But for more complex scenarios involving multiple agents to work together as A1, A5, it is better to have a Manager Agent to organize the working flow of individual agents.

## 5.4 A4: Chess Game

Chess Game is a practical application to enable playing the Chess game against Large Language models. It takes turns to ask the Chess Agent and Human to move a step on the Chessboard. As the workflow can be easily managed, we do not need to build a Manager

<sup>6</sup><https://openai.com/dall-e-2>

<sup>7</sup><https://platform.openai.com/docs/guides/vision>

<sup>8</sup><https://www.wolframalpha.com/>

Agent here. It involves one Human Agent collecting input from the keyboard and one Chess Agent that is able to view the board, make analysis, and move the chess.

```
from agentlite.agents import BaseAgent
name = "Chess_Player"
role = "You are a chess player to make move on the chess board."
actions = [BoardMove(), BoardView(), LegalMoves()]
Chess_Agent = BaseAgent(name=name, role=role, actions=actions)
```

Human Agent directly collects the human input from the keyboard and makes the corresponding move on the board. Chess Agent has 5 actions to play chess with human. *Think* and *Finish* action are internally supported for reasoning and finish. Besides, 3 more actions are further added to the Chess Agent. *BoardView* action obtains current board's Forsyth-Edwards Notation (FEN), which is a standard notation to describe the positions of chess, and display the current board. *LegalMoves* action obtains current available moves of the board. Chess Agent finds available moves with this action. *BoardMove* action moves the chess on the board. To be noted that, agent can be on both side of the Chess Board, which further supports Agent-Agent Chess game.

## 5.5 A5: Philosophers Chatting

Philosophers Chatting is an innovative application to use the Large Language Model to mimic great Philosopher's thoughts including Socrates, Confucius, and Aristotle. The whole application is a copilot multi-agent as described in Sec. 3.2, which is a Manager Agent with its team members as Human Agent, Socrates Agent, Confucius Agent, and Aristotle Agent. Each time Human can give one philosophy question to the Manager Agent such as "*What should we pursue during our life?*". After the Manager Agent receives the question, it asks the Socrates/Confucius/Aristotle agents for their opinion one by one and summarizes the results to answer the question.

To accomplish the application, three kinds of agents are involved. Human Agent directly provide philosophy questions for the Manager Agent. Philosopher Agents (Socrates, Confucius, Aristotle) can represent different great philosophers in history and answer the question grounded by their literary work. For example, we could define *Socrates\_Agent* as follows:

```
from agentlite.agents import BaseAgent
name = "Socrates"
role = "You are Socrates. You are very familiar with Socrates's Book and Thought. Tell your opinion on behalf of Socrates"
Socrates_Agent = BaseAgent(name=name, role=role)
```

The Manager Agent is instantiated through the integration of four distinct agents into a team. Manager Agent orchestrates discussion flow, which encompasses identification of relevant questions, forwarding of these inquiries to appropriate agent, and summarization of the discussion. This structured approach facilitates an effective exchange of information.

## 6 Conclusion and Future Work

AgentLite is a lightweight codebase for researchers to develop customized LLM agent systems: **1)** One can easily use it for reproducing the benchmark results and prototyping various applications; **2)** AgentLite enables effortless integration and evaluation of new reasoning strategies and agent architectures. We develop a series of applications with AgentLite to showcase the superiority. In the future, we will introduce additional communication methods among agents and offer a wider range of reasoning types in the agent class.

## References

- Harrison Chase. Langchain. <https://github.com/hwchase17/langchain>, 2023.
- Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun, and Yu Su. Mind2web: Towards a generalist agent for the web. *arXiv preprint arXiv:2306.06070*, 2023a.
- Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun, and Yu Su. Mind2web: Towards a generalist agent for the web. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023b. URL <https://openreview.net/forum?id=kiYqb03wqw>.
- Significant Gravitas. Autogpt. <https://github.com/Significant-Gravitas/Auto-GPT>, 2023.
- Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. Mixtral of experts. *arXiv preprint arXiv:2401.04088*, 2024.
- Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang. Agentbench: Evaluating llms as agents, 2023a.
- Zhiwei Liu, Weiran Yao, Jianguo Zhang, Le Xue, Shelby Heinecke, Rithesh Murthy, Yihao Feng, Zeyuan Chen, Juan Carlos Niebles, Devansh Arpit, et al. Bolaa: Benchmarking and orchestrating llm-augmented autonomous agents. *arXiv preprint arXiv:2308.05960*, 2023b.
- Chang Ma, Junlei Zhang, Zhihao Zhu, Cheng Yang, Yujiu Yang, Yaohui Jin, Zhenzhong Lan, Lingpeng Kong, and Junxian He. Agentboard: An analytical evaluation board of multi-turn llm agents, 2024.
- Yohei Nakajima. Babyagi. <https://github.com/yoheinakajima/babyagi>, 2023.
- OpenAI. Gpt-4 technical report. *ArXiv*, 2023.
- Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *In the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.
- Shishir G Patil, Tianjun Zhang, Xin Wang, and Joseph E Gonzalez. Gorilla: Large language model connected with massive apis. *arXiv preprint arXiv:2305.15334*, 2023.
- Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789*, 2023.
- Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *arXiv preprint arXiv:2302.04761*, 2023.
- Weizhou Shen, Chenliang Li, Hongzhan Chen, Ming Yan, Xiaojun Quan, Hehong Chen, Ji Zhang, and Fei Huang. Small llms are weak tool learners: A multi-llm agent, 2024.
- Noah Shinn, Federico Cassano, Beck Labash, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *arXiv preprint arXiv:2303.11366*, 2023.

- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models, 2023.
- Yu Wang, Zhiwei Liu, Jianguo Zhang, Weiran Yao, Shelby Heinecke, and Philip S Yu. Drdt: Dynamic reflection with divergent thinking for llm-based sequential recommendation. *arXiv preprint arXiv:2312.11336*, 2023.
- Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. 2023.
- Binfeng Xu, Zhiyuan Peng, Bowen Lei, Subhabrata Mukherjee, Yuchen Liu, and Dongkuan Xu. Rewoo: Decoupling reasoning from observations for efficient augmented language models. *arXiv preprint arXiv:2305.18323*, 2023.
- Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2023a.
- Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. Webshop: Towards scalable real-world web interaction with grounded language agents. In *ArXiv*, preprint.
- Weiran Yao, Shelby Heinecke, Juan Carlos Niebles, Zhiwei Liu, Yihao Feng, Le Xue, Rithesh Murthy, Zeyuan Chen, Jianguo Zhang, Devansh Arpit, Ran Xu, Phil Mui, Huan Wang, Caiming Xiong, and Silvio Savarese. Retroformer: Retrospective large language agents with policy gradient optimization, 2023b.
- Aohan Zeng, Mingdao Liu, Rui Lu, Bowen Wang, Xiao Liu, Yuxiao Dong, and Jie Tang. Agenttuning: Enabling generalized agent abilities for llms, 2023.
- Jianguo Zhang, Tian Lan, Murthy Rithesh, Zhiwei Liu, Weiran Yao, Juntao Tan, Yihao Feng, Thai Hoang, Liangwei Yang, Tulika Awalganekar, Zuxin Liu, Juan Carlos Niebles, Silvio Savarese, Shelby Heinecke, Huan Wang, and Caiming Xiong. Agentohana: Design unified data and training pipeline for effective agent learning. 2024.
- Boyuan Zheng, Boyu Gou, Jihyung Kil, Huan Sun, and Yu Su. Gpt-4v(ision) is a generalist web agent, if grounded. *arXiv preprint arXiv:2401.01614*, 2024.
- Shuyan Zhou, Frank F Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Yonatan Bisk, Daniel Fried, Uri Alon, et al. Webarena: A realistic web environment for building autonomous agents. *arXiv preprint arXiv:2307.13854*, 2023. URL <https://webarena.dev>.