Outline

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

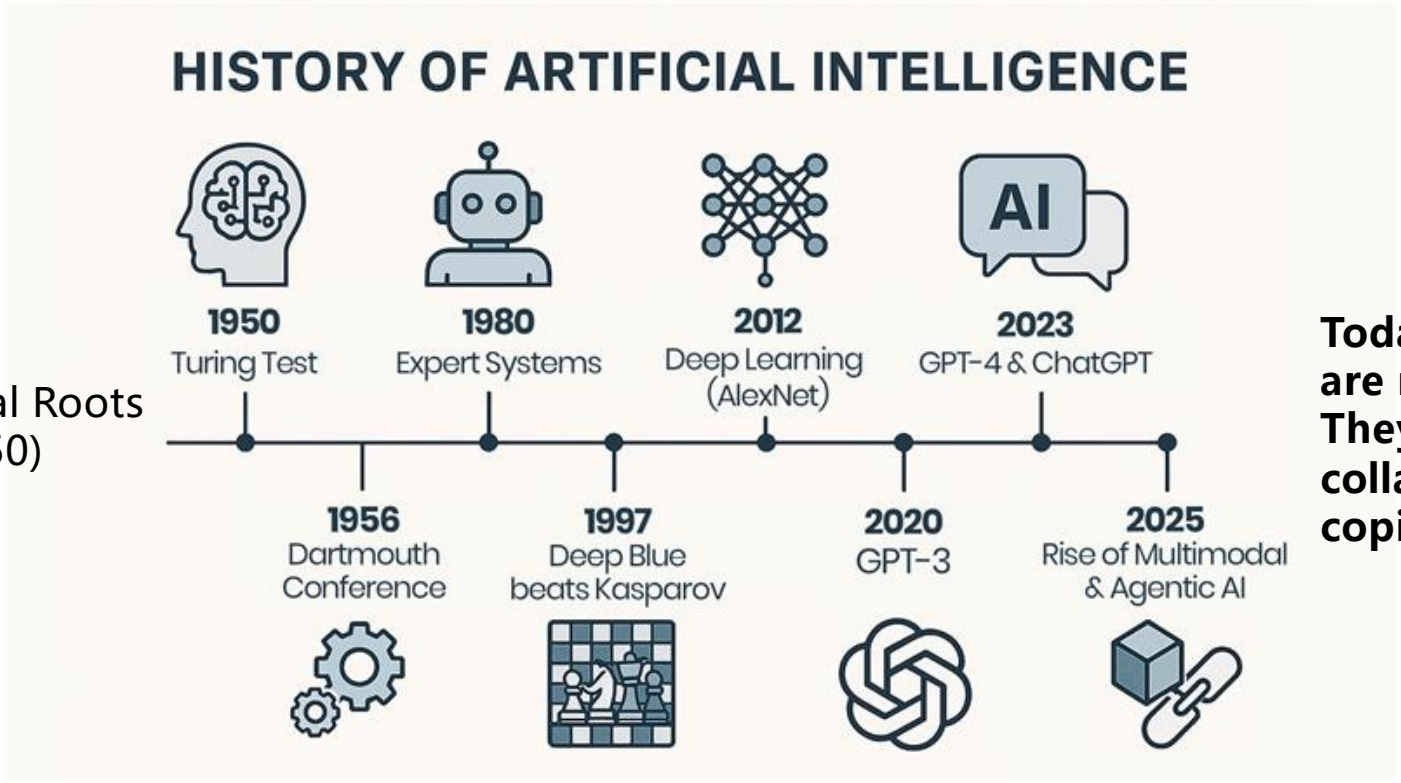
1. Introduction
2. Definition
   - Traditional AI VS Agentic AI
3. Architectures
   - Single-Agent System VS Multi-Agent System
4. Industrial Frameworks
   - AutoGen by Microsoft and ADK by Google
5. Applications
   - Magentic-One by Microsoft
   - SciToolAgent by ZJU
   - Paper-to-Video Agentic System by PKU and DAMO (Alibaba)
6. Summary

Introduction to Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

The Evolution of AI

The Philosophical Roots of AI (Before 1950)



Today's AI systems are not just tools. They're becoming collaborators and copilots.

Neeraj Bansal. "The History of Artificial Intelligence: A Timeline of Innovation from Concept to Code." May 21, 2025.

Introduction to Agentic System

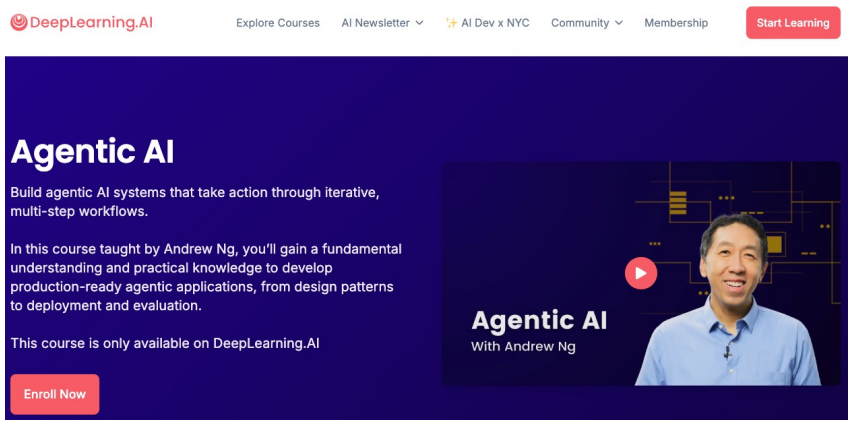
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

How Popular the Agentic System is?

- Adrew Ng mentioned the rise of agentic reasoning in BUILD 2024 keynote. Recently, he started to teach Agentic (AI) system on DeepLearning.AI.

- It's predicted that by 2028, 33% of enterprise software applications will include agentic AI.
  - According to call for proposals, industry is already bracing agents (broadly defined)

Andrew Ng Explores The Rise Of AI Agents And Agentic Reasoning | BUILD 2024 Keynote



Jonathan D. Gough. "Top 10 Agentic AI Examples and Use Cases" Pellera. July 12, 2025.

Introduction to Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Comparison of Agentic AI and Traditional AI:

Feature	Agentic AI	Traditional AI
Primary Purpose	Goal-Oriented Autonomy	Task-Specific Automation
Human Intervention	Low	High
Adaptability	High	Limited
Environment Interaction	Dynamic and Context-Aware	Static or Limited Context
Learning Type	Reinforcement and Self-Supervised	Primarily Supervised
Decision-Making	Autonomous, Contextual Reasoning	Data-Driven, Static Rules
Architectures	Transformers, multi-agent	Shallow models, single-agent

Acharya, Deepak Bhaskar, Karthigeyan Kuppan, and B. Divya. "Agentic ai: Autonomous intelligence for complex goals–a comprehensive survey." *IEEE Access* (2025).

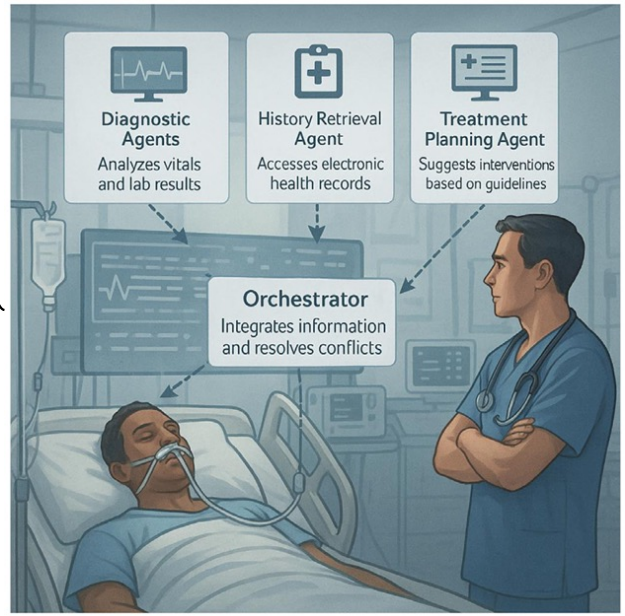
Introduction to Agentic System

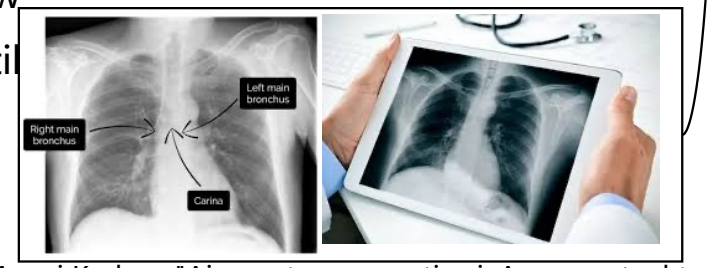
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Detailed Comparisons in Clinical Applications

Agentic AI System

- **Goal-Oriented Autonomy:** Clinical decision support in hospital ICUs
- **Autonomous Reasoning:** through synchronized agents diagnostics, treatment planning, and analysis, enhancing safety and workflow efficiency.
- **Low intervention**: doctors have low process intervention (end results still need human intervention in the medical field).





Traditional AI

- **Task-Specific Automation:** input the data and output the result
- **Data-driven Prediction:** only predict the results without reasoning
- **High intervention:** doctors need further lab experimental results for diagnosis.

Sapkota, Ranjan, Konstantinos I. Roumeliotis, and Manoj Karkee. "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." *arXiv:2505.10468*.

Introduction to Agentic System

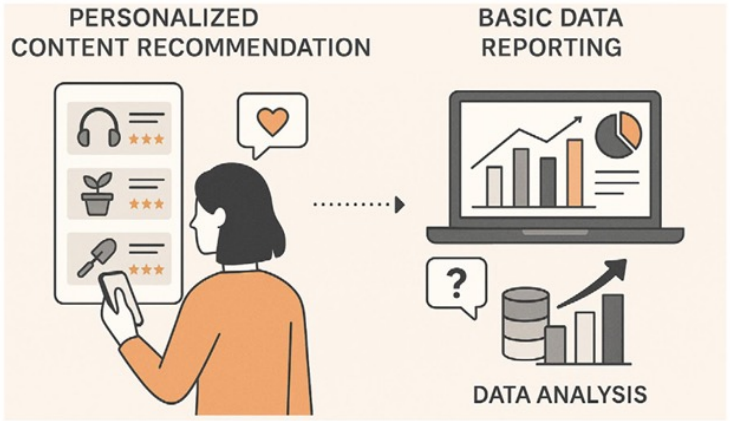
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Detailed Comparisons in Production Applications

Agentic AI System

High Adaptability: by analyzing data from diverse sources like sales, inventory, and shipping, agentic AI system can optimize supply chains, predict demand, and automate logistics.

Dynamic Environment Interaction: once a number or figure has been changed in the pipeline, the agentic system can quickly update for the following steps.



Traditional AI

Low Adaptability: one model for one task.

Static Environment Interaction: cannot respond to dynamic changes without human intervention.

Sapkota, Ranjan, Konstantinos I. Roumeliotis, and Manoj Karkee. "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." *arXiv:2505.10468*.

Definition of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

- **Overall Definition:** An agentic AI system, or simply agentic system, is a collection of agents interacting with humans and the environment with the objective of fulfilling specified goals.
- **Components:**
  - **Agent:** an LLM or large multimodal model (LMM) with access to tools — specialized components/functionalities like APIs, external services, computational resources, or domain-specific software — that allow it to perform specific operations in the environment.
  - **Tool:** both the capabilities (actions) of the agent and the information (via observations/signals) that can be obtained from the environment.

Miehling, Erik, et al. "Agentic ai needs a systems theory." arXiv preprint arXiv:2503.00237 (2025).

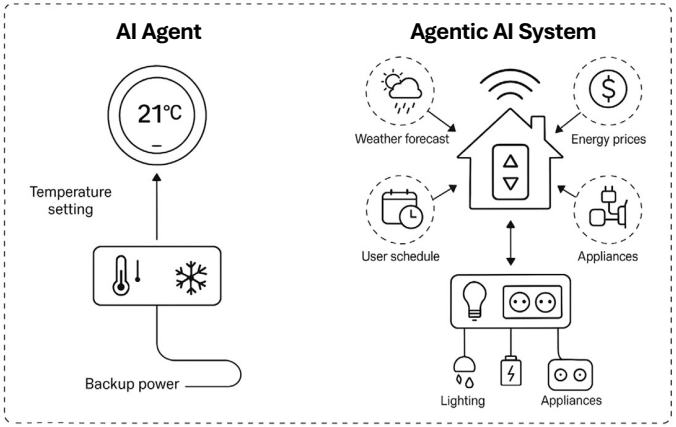
Definition of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

From AI Agents to Agentic AI Systems:

- **AI Agent:** autonomous technologies, designed to perform tasks, make decisions, and interact with their environment to achieve specific predefined goals, with minimal human intervention.
- **Agentic AI System:** integrate one or more LLM agents to automate complex tasks and streamline processes across various domains.

A basic AI Agent **managing a single task** — such as maintaining a thermostat's temperature setting based on user input.



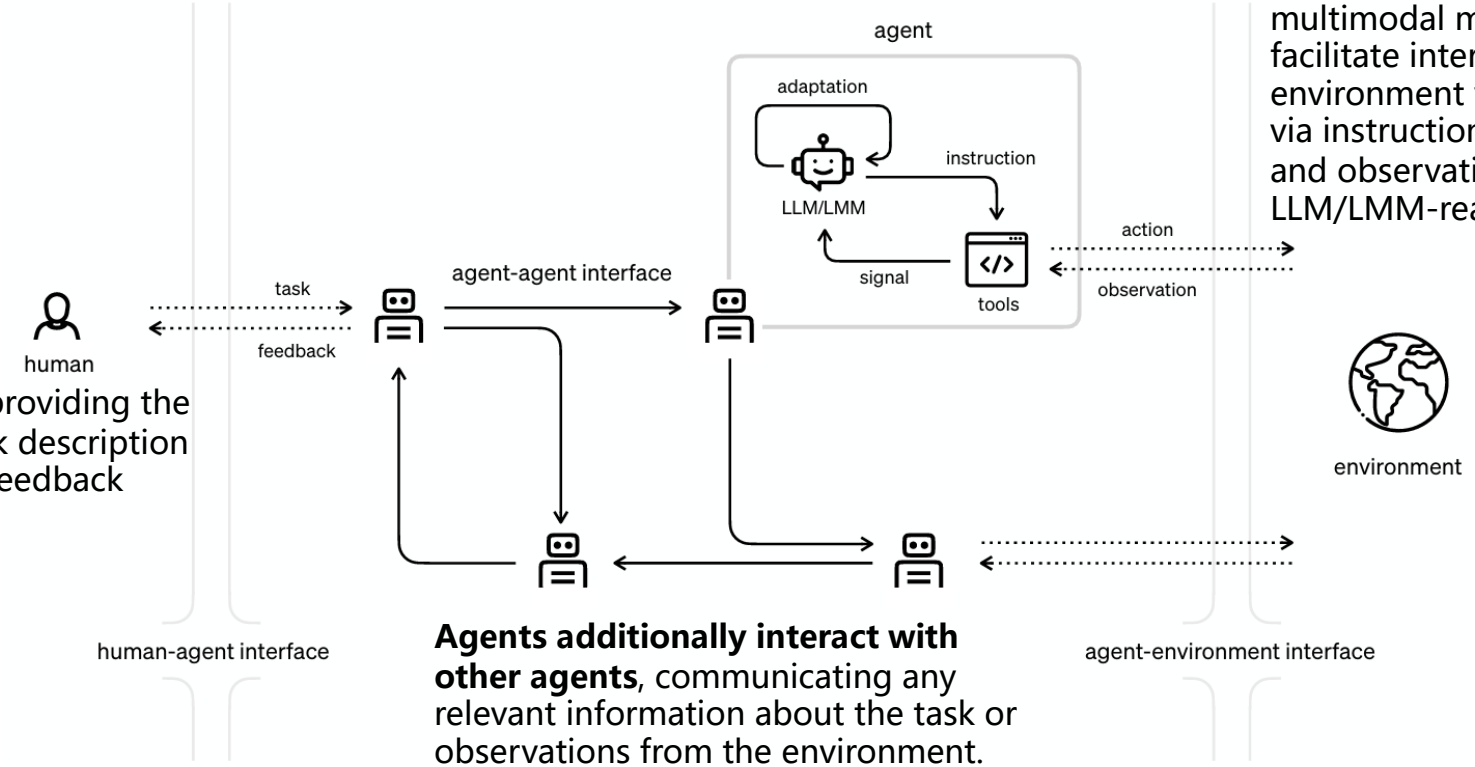
An Agentic AI System **coordinating multiple agents** to manage household automation tasks including temperature regulation, weather forecasting, energy pricing, appliance scheduling, and security monitoring.

Sapkota, Ranjan, Konstantinos I. Roumeliotis, and Manoj Karkee. "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." *arXiv:2505.10468*.

Definition of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

The Overall Pipeline of An Agentic System



agent: a LLM or LMM (large multimodal model), with tools that facilitate interaction with the external environment via actions (generated via instructions from the LLM/LMM) and observations (generating LLM/LMM-readable signals).

human: providing the initial task description and any feedback

Environment: defines reward and transition to next state.

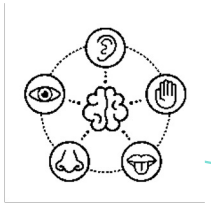
Agents additionally interact with other agents, communicating any relevant information about the task or observations from the environment.

Miehling, Erik, et al. "Agentic ai needs a systems theory." arXiv preprint arXiv:2503.00237 (2025).

Definition of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Key Concepts of Agentic AI System



Perception: Gathering information from surroundings and different sources.

Slow thinking

Reasoning: Powered by LLMs, agentic AI system analyzes the gathered data to understand the context, identify relevant information, and formulate potential solutions.

Planning: Using gathered information to develop a plan, which involves setting goals, splitting into smaller steps, and figuring out the best way to achieve them.

Fast thinking: perception to action directly.

Action: Performing tasks, making decisions, or interacting with other systems.

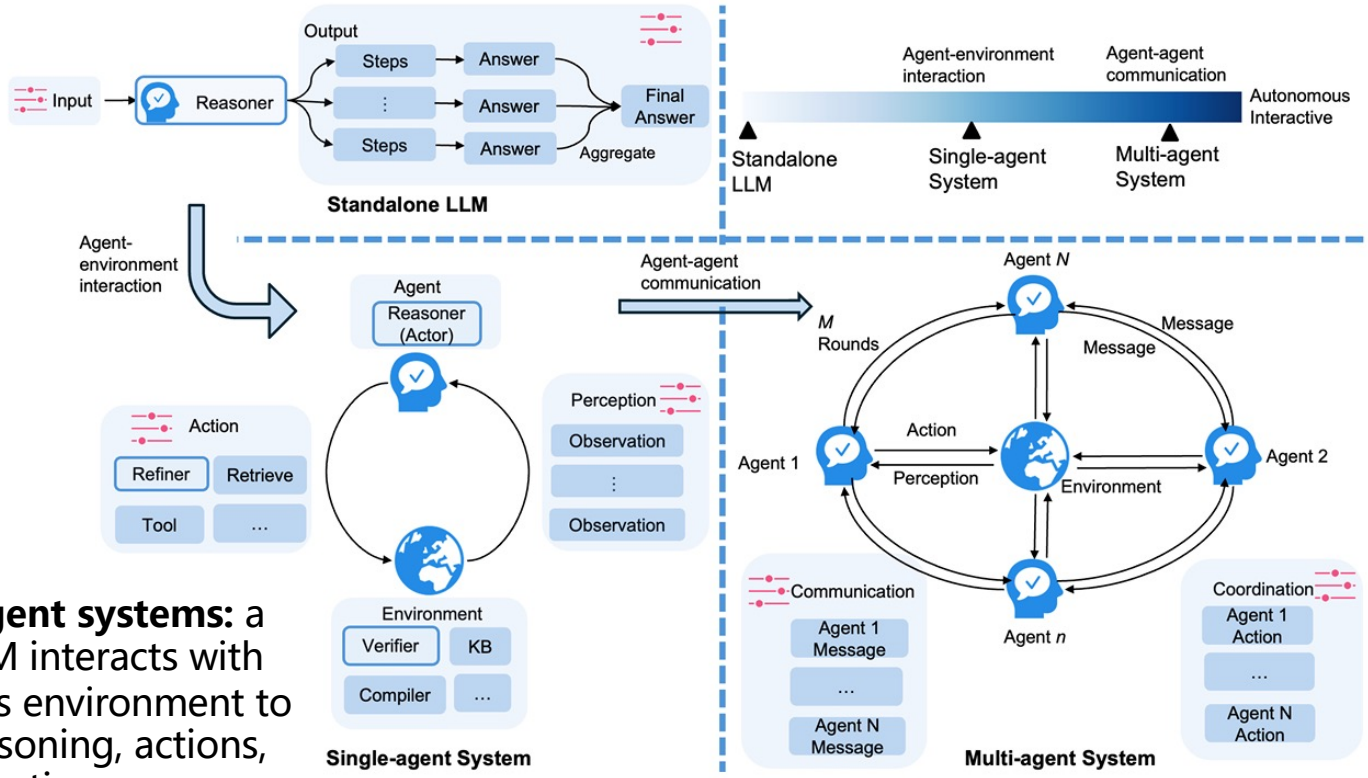
Reflection: After taking actions, the system learns from the results. It evaluates whether its actions were successful and uses this feedback to adjust its plans and actions in the future.

Google Cloud. "What is agentic AI?". https://cloud.google.com/discover/what-is-agentic-ai

Architecture of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Taxonomy of Agentic Systems



Single-agent systems: a single LLM interacts with tools in its environment to refine reasoning, actions, and perceptions.

Multi-agent systems: goes beyond agent-environment interactions by enabling agent-agent communication. Each agent takes on a distinct role and exchanges messages with others. Coordinations are need for collaborations.

Ke, Zixuan, et al. "A survey of frontiers in llm reasoning: Inference scaling, learning to reason, and agentic systems." arXiv preprint arXiv:2504.09037 (2025).

Architecture of Agentic System

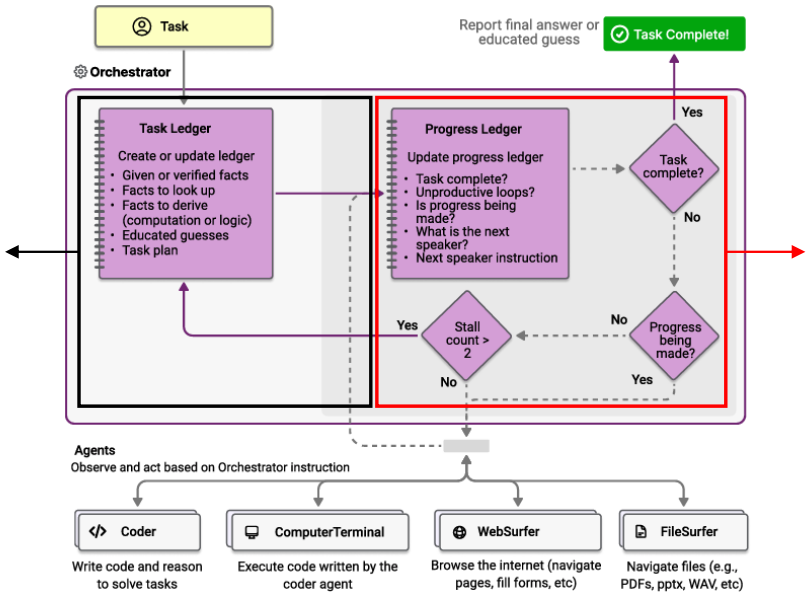
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

The Architectural Design of Multi-Agent System (MAS)

- **Centralized Structure.** This architecture follows **a manager-follower** paradigm where a central agent or higher-level coordinator handles planning, task decomposition, and delegation, while subordinate agents execute assigned subtasks.

Manager:

The outer loop manages the task ledger, containing facts, guesses, and plan.



Workers:

The inner loop manages the progress ledger, containing current progress, task assignment to agents.

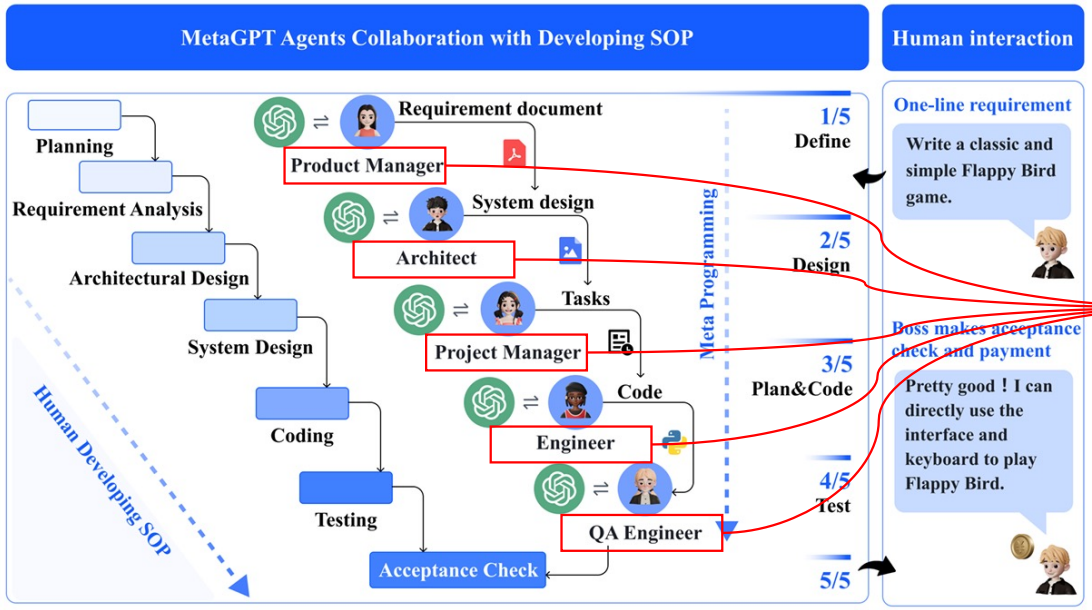
Fourney, Adam, et al. "Magentic-one: A generalist multi-agent system for solving complex tasks." *arXiv preprint arXiv:2411.04468* (2024).

Architecture of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

The Architectural Design of Multi-Agent System (MAS)

- **Hierarchical Structure.** These systems employ static hierarchical organizations, typically linear or tree-based, where tasks are explicitly decomposed and sequentially assigned to specific agents.



MetaGPT showcases its ability to decompose complex tasks into specific actionable procedures assigned to various roles.

Fang, Jinyuan, et al. "A comprehensive survey of self-evolving ai agents: A new paradigm bridging foundation models and lifelong agentic systems." arXiv. 2025.
Hong, Sirui, et al. "MetaGPT: Meta programming for a multi-agent collaborative framework." ICLR. 2023.

Architecture of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

The Architectural Design of Multi-Agent System (MAS)

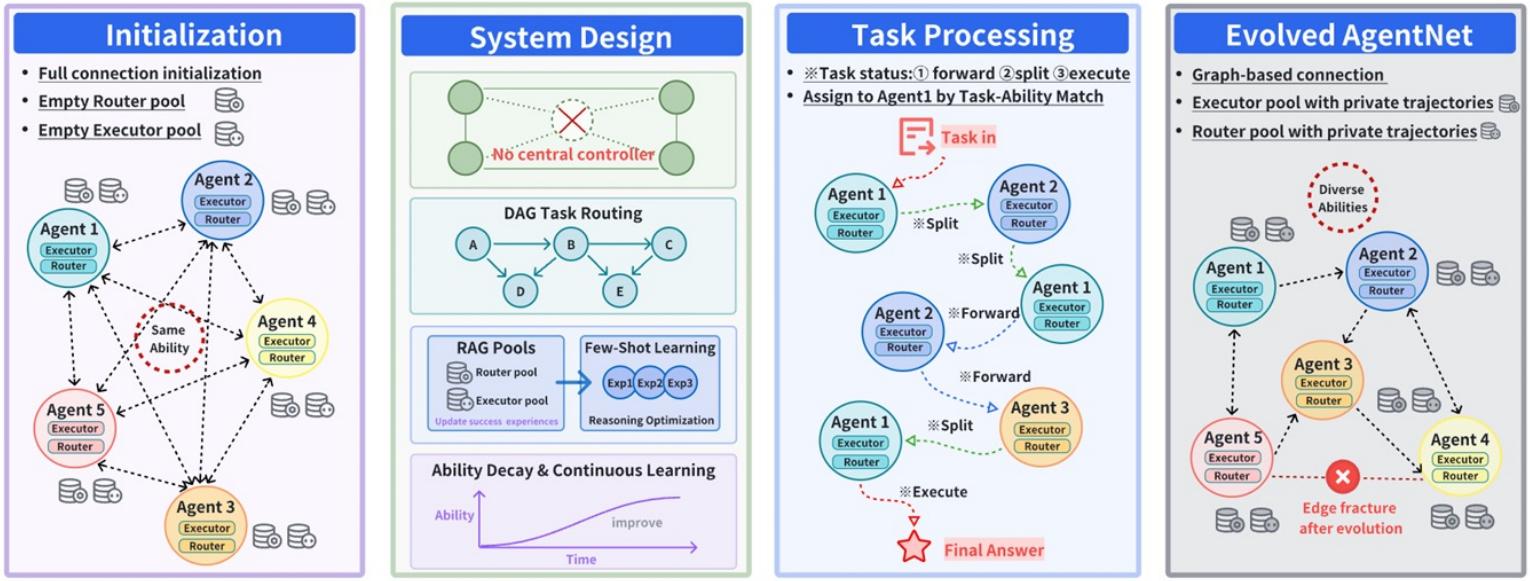
- **Decentralized Structure.** Agents collaborate as peers in a distributed network, widely adopted in world simulation applications, and any single-point failure does not paralyze the entire system.

Initially, agents are fully connected.

A DAG for dynamic task routing.

Executing tasks by following task routings.

Dynamnic evolving architecture

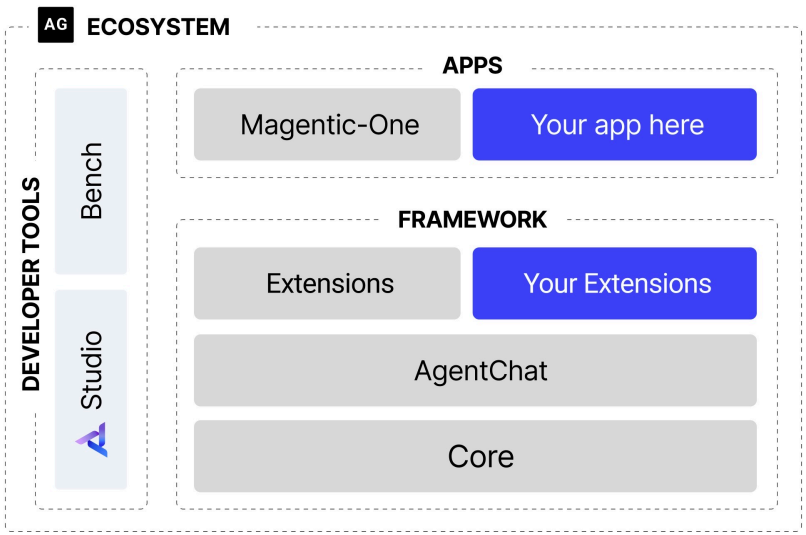


Yang, Yingxuan, et al. "Agentnet: Decentralized evolutionary coordination for llm-based multi-agent systems." *arXiv preprint arXiv:2504.00587* (2025).

Industry Efforts in Agentic System

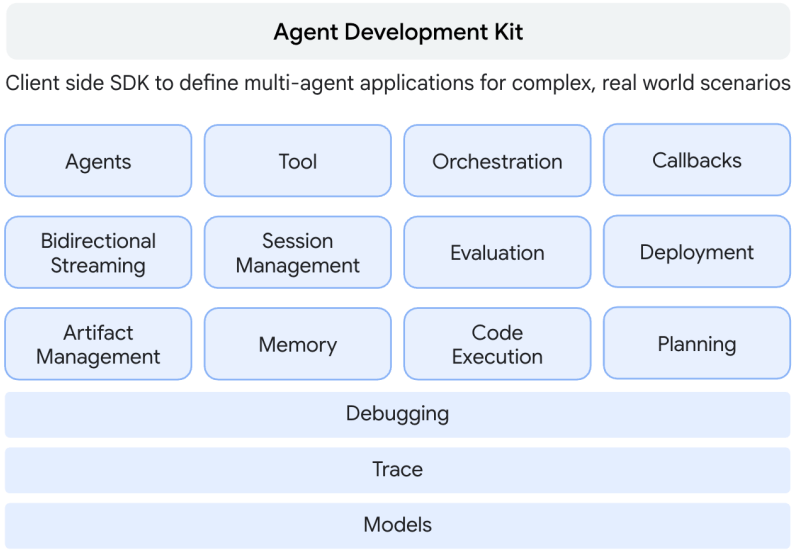
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Microsoft



AutoGen is a framework for creating multi-agent AI applications that can act autonomously or work alongside humans.

Google



Agent Development Kit (ADK) is a flexible and modular framework for developing and deploying AI agents.

Applications of Agentic System

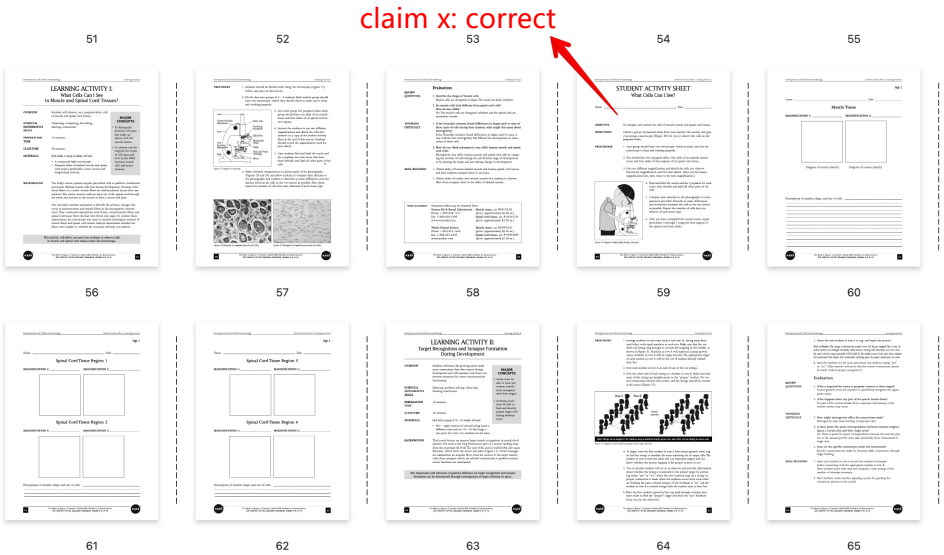
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Magentic-One: A Generalist Multi-Agent System for Solving Complex Tasks

Definition: A task is defined as a complex task if it requires, or significantly benefits from, a process involving planning, acting, observing, and reflecting, potentially multiple times.

For example, the input task is "**fact-check each claim in the file as correct or incorrect**" with a PDF file as an attachment. The desired output consists either of a textual answer (possibly representing a structured object), or a specific state of the environment to reach, e.g.,

```
{
   claim 1: correct,
   claim 2: incorrect,
   ...
}
```



[Microsoft Research AI Frontiers] Fourney, Adam, et al. "Magentic-one: A generalist multi-agent system for solving complex tasks." arXiv:2411.04468 (2024).

Applications of Agentic System

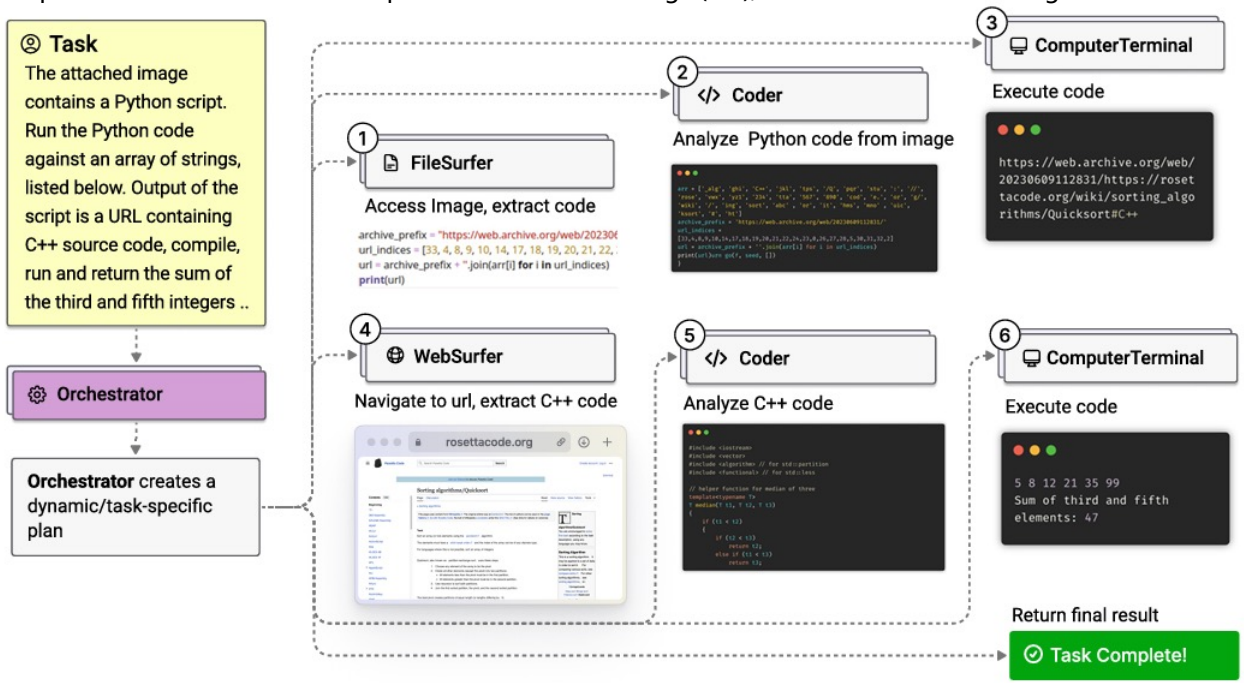
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Magentic-One Architecture Overview

- The **Orchestrator** agent creates a plan, delegates tasks to other agents, and tracks progress towards the goal, dynamically revising the plan as needed.

- The FileSurfer agent to read and handle files.
- The WebSurfer agent to operate a web browser.
- The Coder or Computer Terminal agent to write or execute code.

Sample task from GAIA benchmark: parse codes from the image (1-3), and execute the codes to get the final answer (4-6)



[Microsoft Research AI Frontiers] Fourney, Adam, et al. "Magentic-one: A generalist multi-agent system for solving complex tasks." arXiv:2411.04468 (2024).

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

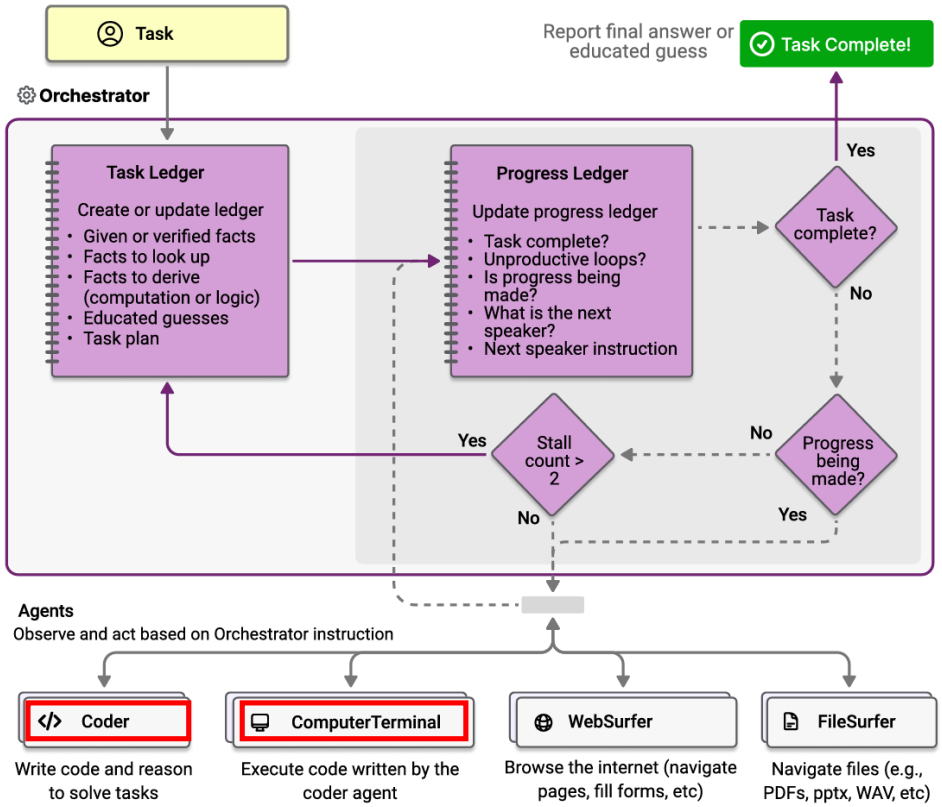
Magentic-One Agents

Coder

An LLM-based agent specialized through its system prompt for writing code, analyzing information collected from the other agents, or creating new artifacts.

ComputerTerminal

It provides the team with access to a console shell where the Coder's programs can be executed.



[Microsoft Research AI Frontiers] Fourney, Adam, et al. "Magentic-one: A generalist multi-agent system for solving complex tasks." arXiv:2411.04468 (2024).

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

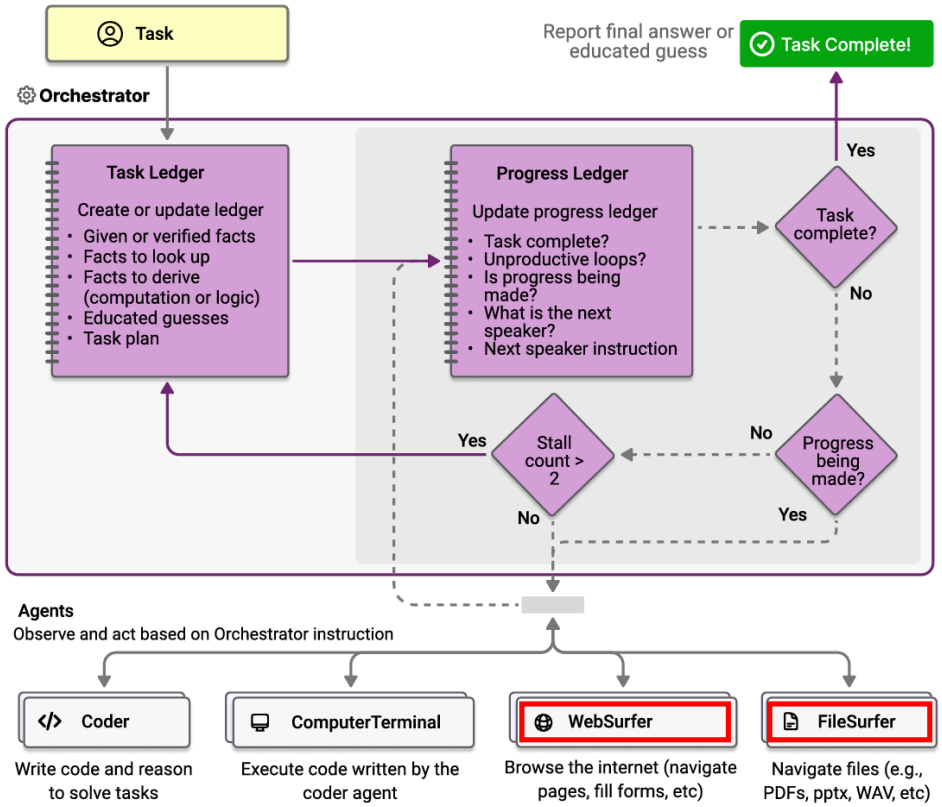
Magentic-One Workflow

WebSurfer

A highly specialized LLM-based agent that is proficient in commanding and managing the state of a Chromium-based web browser.

FileSurfer

It's similar to the WebSurfer, except that it commands a custom markdown-based file preview application rather than a web browser.



[Microsoft Research AI Frontiers] Fourney, Adam, et al. "Magentic-one: A generalist multi-agent system for solving complex tasks." arXiv:2411.04468 (2024).

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Magentic-One Evaluation

Benchmarks:

- **GAIA:** a benchmark for general AI assistants with 465 multi-modal question–answer pairs, which are real-world and challenging, requiring multiple steps and multiple tools to solve.

Level 1

Question: What was the actual enrollment count of the clinical trial on H. pylori in acne vulgaris patients from Jan-May 2018 as listed on the NIH website?
Ground truth: 90

Level 2

Question: If this whole pint is made up of ice cream, how many percent above or below the US federal standards for butterfat content is it when using the standards as reported by Wikipedia in 2020? Answer as + or - a number rounded to one decimal place.
Ground truth: +4.6

Level 3

Question: In NASA's Astronomy Picture of the Day on 2006 January 21, two astronauts are visible, with one appearing much smaller than the other. As of August 2023, out of the astronauts in the NASA Astronaut Group that the smaller astronaut was a member of, which one spent the least time in space, and how many minutes did he spend in space, rounded to the nearest minute? Exclude any astronauts who did not spend any time in space. Give the last name of the astronaut, separated from the number of minutes by a semicolon. Use commas as thousands separators in the number of minutes.
Ground truth: White; 5876

GAIA examples

- Level 1: questions generally **require no tools**, or at most one tool but no more than 5 steps.
- Level 2: questions generally **involve more steps**, roughly between 5 and 10 and combining different tools is needed.
- Level 3: questions are for a near-perfect general assistant, requiring to take **arbitrarily long sequences of actions**, use any number of tools, and access to the world in general.

Mialon, Grégoire, et al. "Gaia: a benchmark for general ai assistants." ICLR. 2023.

Magentic-One Evaluation

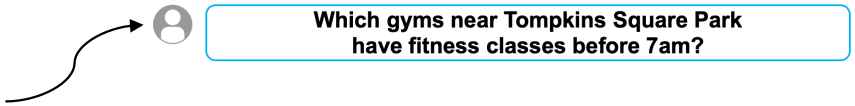
Benchmarks:

- **AssistantBench:** a set of 214 question–answer pairs that are realistic, time-consuming (requiring a human several minutes to perform), and require navigating real-world websites and multi-step reasoning.

Task: Which gyms near Tompkins Square Park (<200m) have fitness classes before 7am?

Answer: {CrossFit East River, Avea Pilates}

AssistantBench example: require navigating the web to find relevant information



Execution example: A gold trajectory for a task. A web agent can solve the task by first interacting with a map tool (e.g., Google Maps) to find nearby gyms and then browsing each website to find the relevant schedule.

Yoran, Ori, et al. "Assistantbench: Can web agents solve realistic and time-consuming tasks?." arXiv preprint arXiv:2407.15711 (2024).

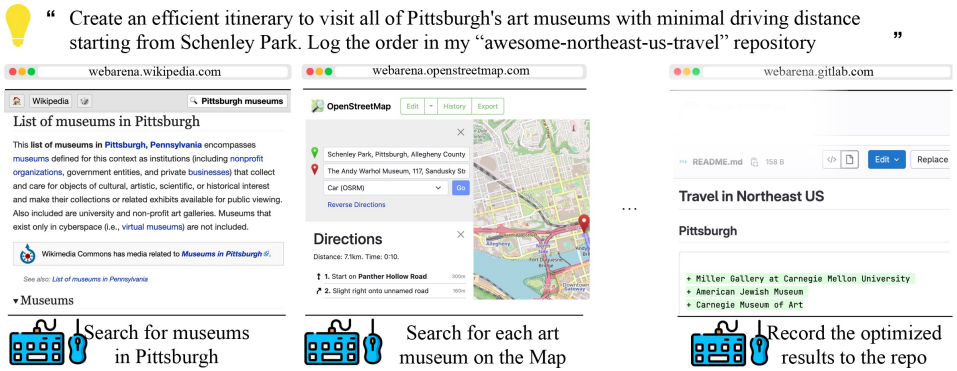
Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Magentic-One Evaluation

Benchmarks:

- **WebArena:** 812 tasks across five major website categories (e.g., shopping, forums, maps, etc.), and a sixth category that requires multi-step planning and interacting with multiple websites.



WebArena example: agents should finish each step to accomplish the task

Action Type	Description
`noop`	Do nothing
`click(elem)`	Click at an element
`hover(elem)`	Hover on an element
`type(elem, text)`	Type to an element
`press(key_comb)`	Press a key comb
`scroll(dir)`	Scroll up and down
`tab_focus(index)`	focus on i-th tab
`new_tab`	Open a new tab
`tab_close`	Close current tab
`go_back`	Visit the last URL
`go_forward`	Undo `go_back`
`goto(URL)`	Go to URL

Figure 4: Action Space of `WebArena`

Category	Example
Information Seeking	When was the last time I bought shampoo
	Compare walking and driving time from AMC Waterfront to Randyland
Site Navigation	Checkout merge requests assigned to me
	Show me the ergonomic chair with the best rating
Content & Config	Post to ask “whether I need a car in NYC”
	Delete the reviews from the scammer Yoke

Figure 5: Example intents from three categories.

WebArena action types and task categories

Zhou, Shuyan, et al. "Webarena: A realistic web environment for building autonomous agents." arXiv preprint arXiv:2307.13854 (2023).

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Magentic-One Evaluation

Performance:

Dataset	Category	Magentic-One (GPT-4o)	Magentic-One (GPT-4o, o1)	Best Baseline [75, 71]
GAIA [29]	Level 1	46.24	**54.84**	53.76 [75]
	Level 2	28.3	32.7	**37.11**
	Level 3	18.75	22.92	**26.53**
AssistantBench [71]	Easy	69.9	73.4	**81** [71]
	Medium	35.6	**47.1**	44.6
	Hard	**16.9**	14.8	13.3
WebArena [79]	Reddit	53.77	–	**65.1** [75]
	Shopping	33.16	–	**36.9**
	CMS	**29.1**	–	24.7
	Gitlab	27.78	–	**39.4**
	Maps	**34.86**	–	33.9
	Cross Site	14.6	–	–

Magentic-One complete better on hard tasks on AssistantBench.

[71] SPA(See-Plan-Act)-CB: compared to Magentic-One, it lacks Orchestrator and dynamic task management, without the ability to replan or handle errors dynamically

[75] WebPilot: focuses on global and local optimization for web tasks and lacks the same level of task orchestration and agent specialization across multiple domains

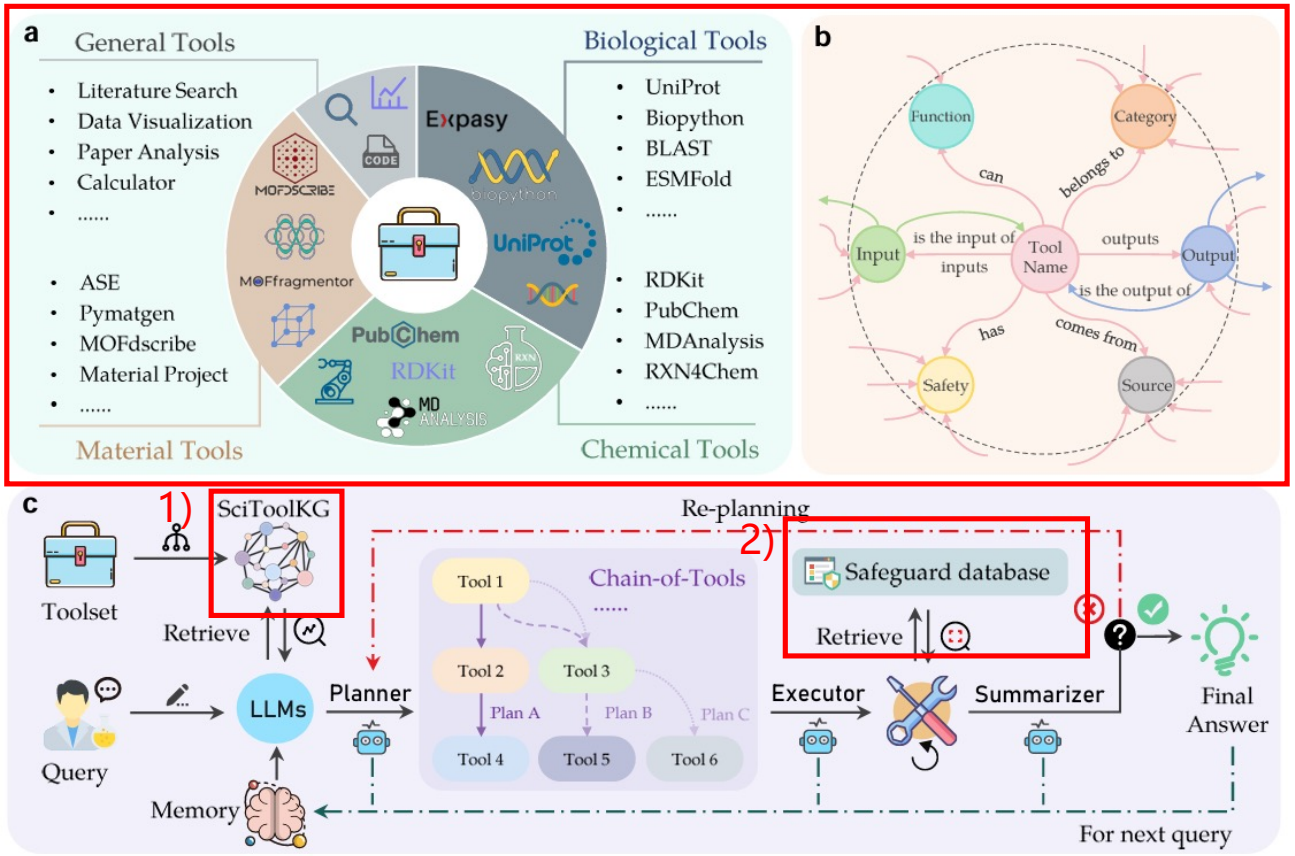
[Microsoft Research AI Frontiers] Fourney, Adam, et al. "Magentic-one: A generalist multi-agent system for solving complex tasks." arXiv:2411.04468 (2024).

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

SciToolAgent: an LLM-based agent in scientific research

1)



Why SciToolAgent?

Issues of existing methods integrating LLMs with domain-specific scientific tools:

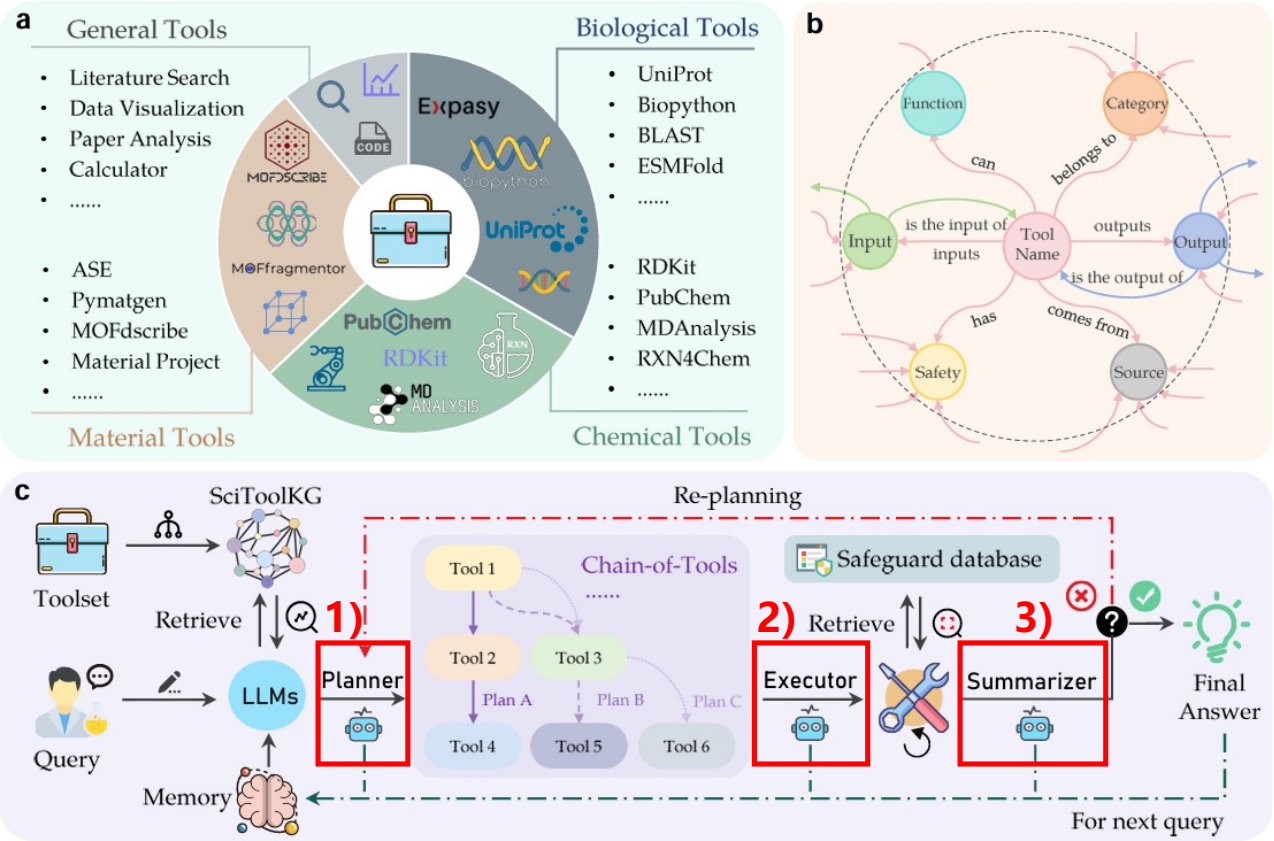
1) A restricted set of tools without semantic tool connections **(SciToolKG constructed)**.
2) Overlook crucial safety and ethical considerations in scientific research **(the security check in the Executor)**.

Ding, Keyan, et al. "SciToolAgent: a knowledge-graph-driven scientific agent for multitool integration." Nature Computational Science (2025): 1-11.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

SciToolAgent Overview



SciToolAgent Implementation

1) LLM-based Planner

- Retrieve relevant k tools
- Explore neighbor tools
- Generate Chain-of-Tools

2) LLM-based Executor

- Prepare necessary data
- Invoke tools with prepared inputs
- Error handling and retries
- Safety check

3) LLM-based Summarizer

- Collect and verify the outputs
- Refine the interation

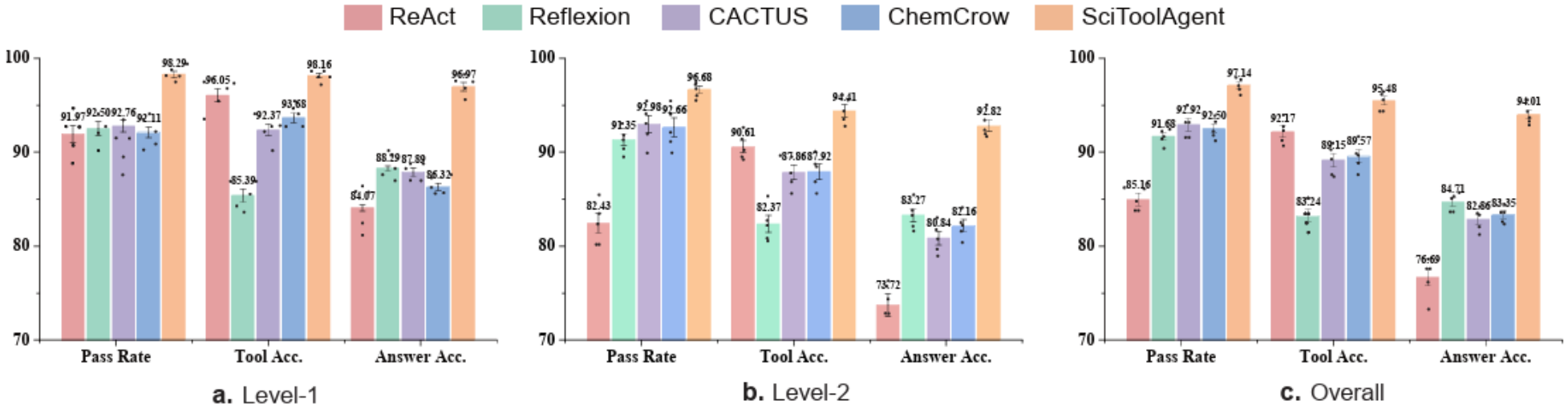
Ding, Keyan, et al. "SciToolAgent: a knowledge-graph-driven scientific agent for multitool integration." Nature Computational Science (2025): 1-11.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

SciToolAgent Performance

Level-1: 152 simple questions
Level-2: 379 complex questions



Evaluation Metrics:

1) **Pass Rate:** the proportion of successfully completed queries
2) **Tool Planning Accuracy:** the alignment between the tool selection and sequencing generated by the agent and a reference plan verified by human experts
3) **Final Answer Accuracy:** the final solution generated by agents compared with the standard answer

Ding, Keyan, et al. "SciToolAgent: a knowledge-graph-driven scientific agent for multitool integration." Nature Computational Science (2025): 1-11.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

SciToolAgent: raw data samples for classification

	B_Fingerprint
2	0000000000000000000000000100010000000000001100000000000100000011000000000000000000000000010000000000000000000000000000010000000000000000
3	0000000000000000000100000000000001001000000000000001000000000000000000000000010000000000100101000000000000000010110000010000000000000000000000
4	0000000000000000000000010000010100000100000000000000100000000001100010000100010000000000000000000000000100000000000000000000010000000010000000
5	0000000000000001000000000000000000100100000010000000000000000000010000000000100001000000001010010000000000000000000000000000000000000010010000

molecule Morgan fingerprints

	B_Fingerprint
2	1111011000110100000101110011110001001111100110001100011010110010111111010111010011011111011111111001011110101101010011011100001110111010001101110111001000001101100
3	0100100100101000101100100011011100000000000011100011100101000011111110011001001011111100111111110001001010010010000100001111100000010110010011100111101100101001
4	1111111101100111101111001110011011110101111010111101111011101010111110110010111101111111101010100101111111111100101111101110110111111110011111111011111111111111101111101010111101
5	1111111100111010111010110101010110001011000110110110110111011111110110101101001011101111100010111111111011101111111110101110110101100011110101111111110010011111101000110101

molecule RDK fingerprints

	Reactant_B_MaxEStateIndex	Reactant_B_MinAbsEStateIndex	Reactant_B_MinEStateIndex	Reactant_B_qed	Reactant_B_NumValenceElectrons	Reactant_B_Num
2	8.88616654069035	0.38319727891156496	-1.465312736205594	0.6272909554190677	70	0
3	12.353055555555557	0.26587962962962997	-4.474398148148147	0.6347040256381961	86	0
4	9.01600297992118	0.4396862139917699	-1.455849605274209	0.6585766158605998	88	0
5	11.203027210884354	0.1460997732426299	-0.5010185185185188	0.7089373711883955	86	0

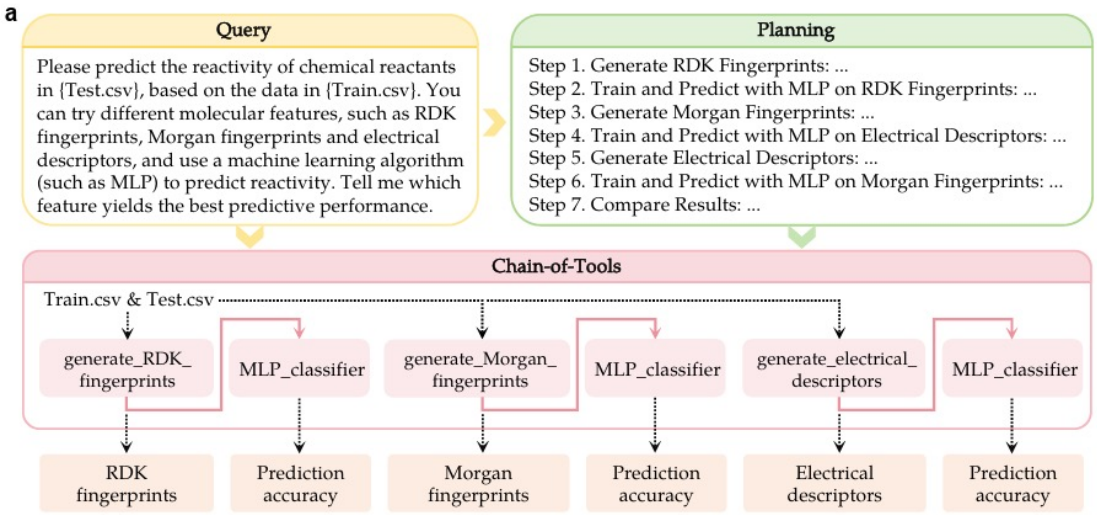
molecule electrical descriptions

Ding, Keyan, et al. "SciToolAgent: a knowledge-graph-driven scientific agent for multitool integration." Nature Computational Science (2025): 1-11.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

SciToolAgent: case study

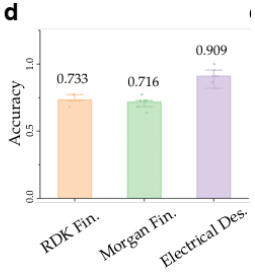


(a) Overview of the input query, tool planning, chain-of-tools, and corresponding outcomes.

c

Reactant_A	Reactant_B	Product	Reactivity
OB(c1c[n](c2ccccc 2)nc1)O	Cc1c(Br)[n]2 c(ncc2)cc1	Cc1ccc2nccn2c1- c1cnn(c2ccccc2)c1	59.2% (mid.)
C=C(C(F)(F)F)B1O C(C)(C)CC(C)O1	Cc1c(Br)[n]2 c(ncc2)cc1	C=C(c1c(C)ccc2n ccn12)C(F)(F)F	25.7% (low)
OCc1c(B(O)O)cc (F)cc1	Brc1c[n]2c(c nc2)cc1	OCc1ccc(F)cc1- c1ccc2cncn2c1	95.3% (high)
...	...	...	...

d



(c) The dataset used for training and testing.
(d) The prediction accuracy with different molecular features.

Open Question: How to evaluate the planning quality?

Ding, Keyan, et al. "SciToolAgent: a knowledge-graph-driven scientific agent for multitool integration." Nature Computational Science (2025): 1-11.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Preacher: Paper-to-Video Agentic System

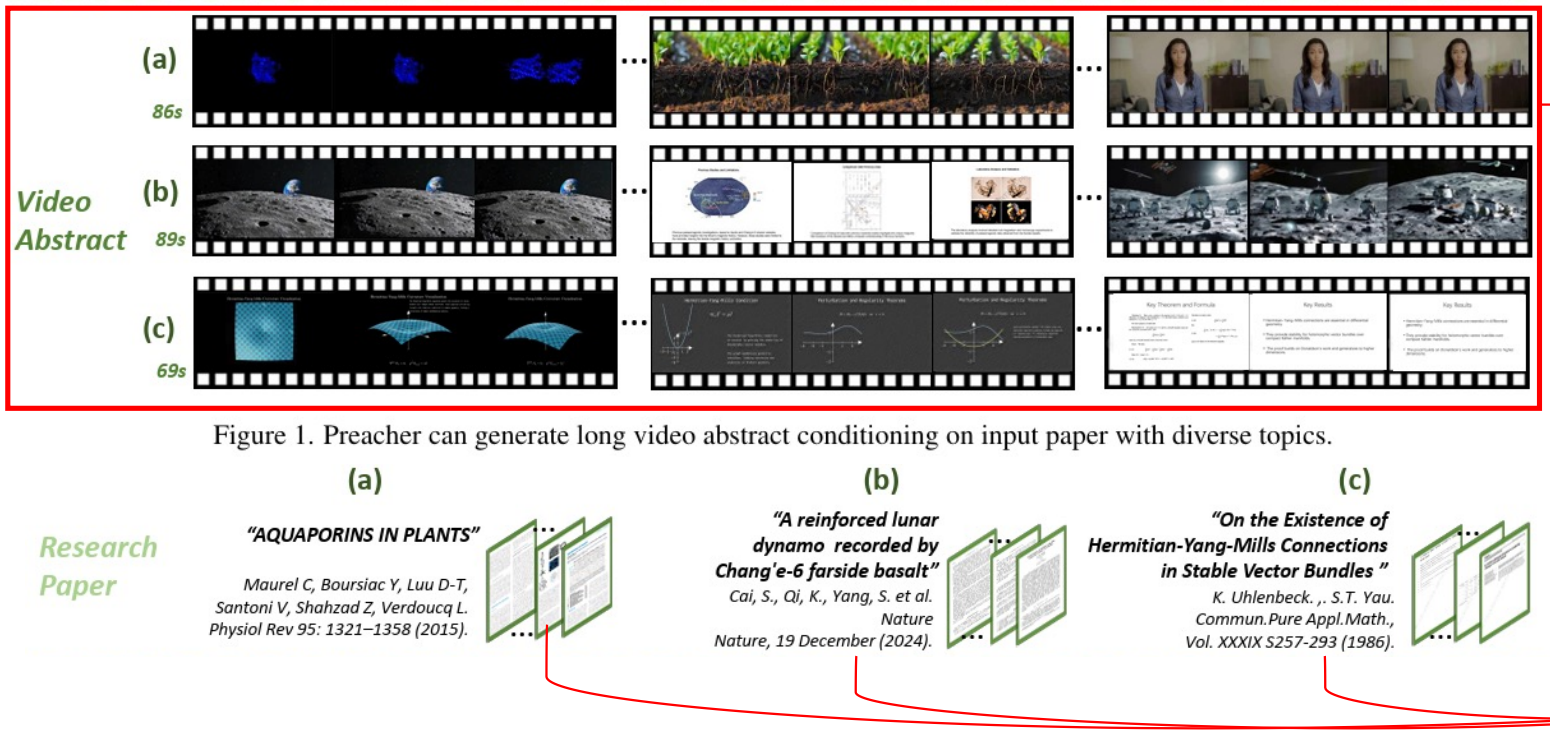


Figure 1. Preacher can generate long video abstract conditioning on input paper with diverse topics.

Motivations:

- Studies have shown that papers accompanied by video abstracts receive 15% more citations.
- Producing video abstracts remains resource-intensive, requiring both domain-specific expertise and professional video production skills, making it a costly process.
- Existing methods are hard to transfer research papers containing embedded multimodal elements and long contexts to videos

Liu, Jingwei, et al. "Preacher: Paper-to-video agentic system." ICCV. 2025.

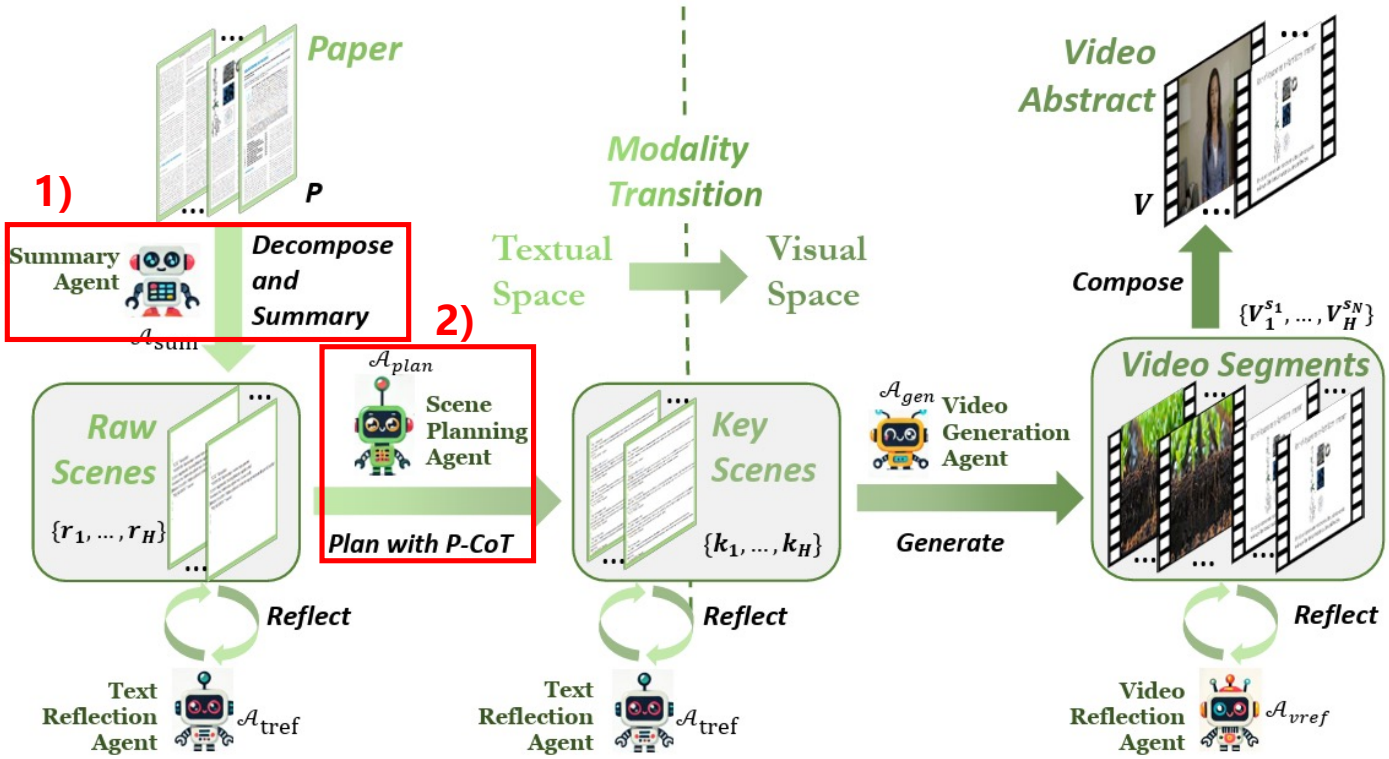
Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Preacher Overview

Top-Down Structure

- Paper undergoes 1) decomposition and summarization into multiple raw scenes $\{r_1, \ldots, r_H\}$, which are chunks of the paper, serving as anchors for content segmentation.
- 2) Planning process continuously references the original paper, ensuring precise semantic alignment. Progressive Chain of Thought (P-CoT) incorporates reflection mechanisms to enhance planning coherence. Planning results termed key scenes $\{k_1, \ldots, k_H\}$.

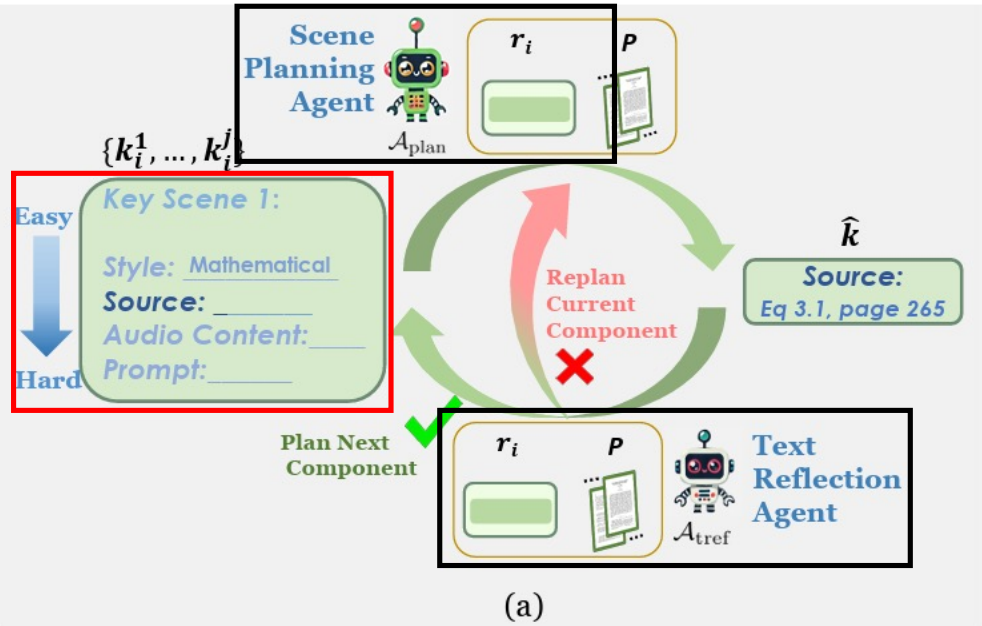


Liu, Jingwei, et al. "Preacher: Paper-to-video agentic system." ICCV. 2025.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Progressive Chain-of-Thought Planning



- The **approved components $\{k_i^1, \ldots, k_i^j\}$ in i-th key scene** requiring systematic planning.
- **Scene Planning Agent** devises a structured plan $\hat{k}$ for i-th key scene based on raw chunk r_i in paper P
- **Text Reflection Agent** evaluates the planned scenes based on content r_i in paper P.

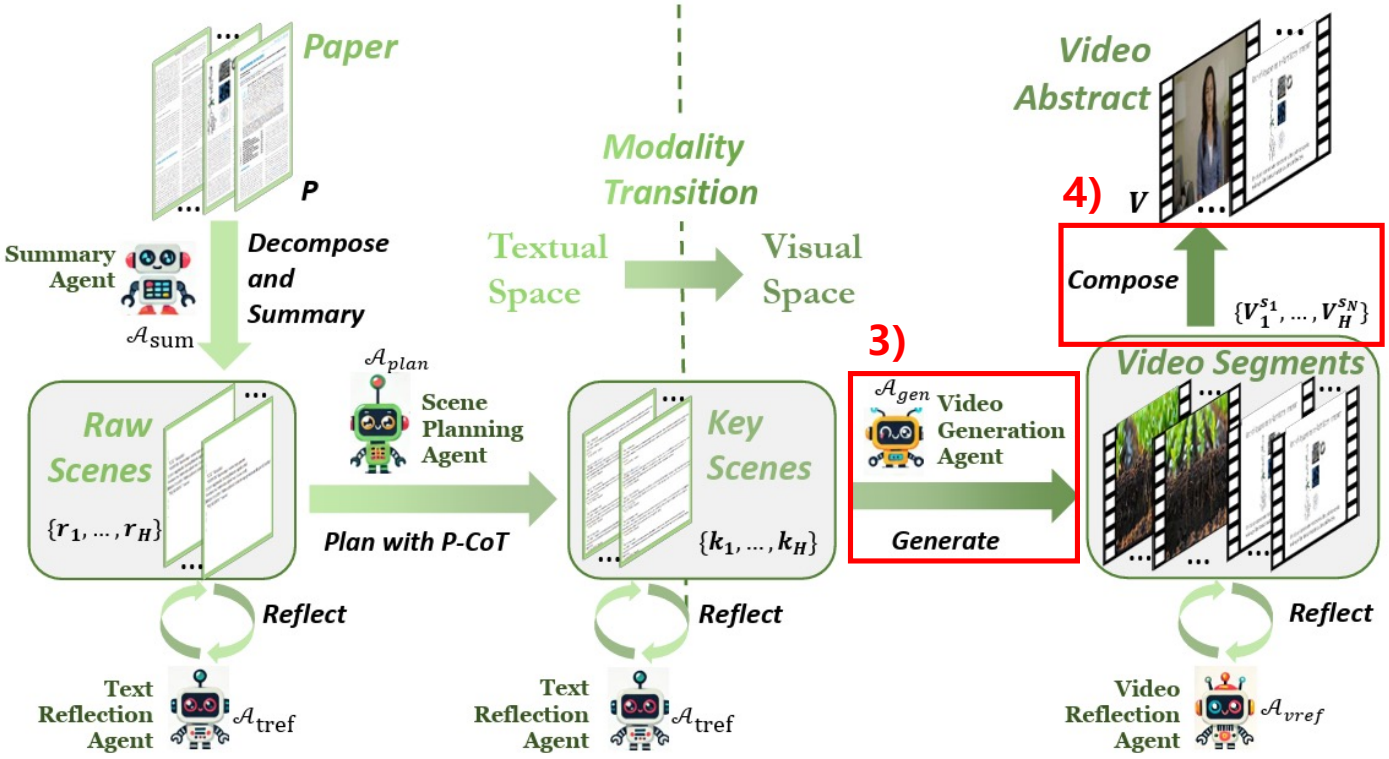
- With the reflection outcome, the scene planning agent either advances to the next component using the existing plan, or revises the current component, iterating this process until all components are effectively planned.

Liu, Jingwei, et al. "Preacher: Paper-to-video agentic system." ICCV. 2025.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Preacher Overview



Bottom-up Structure

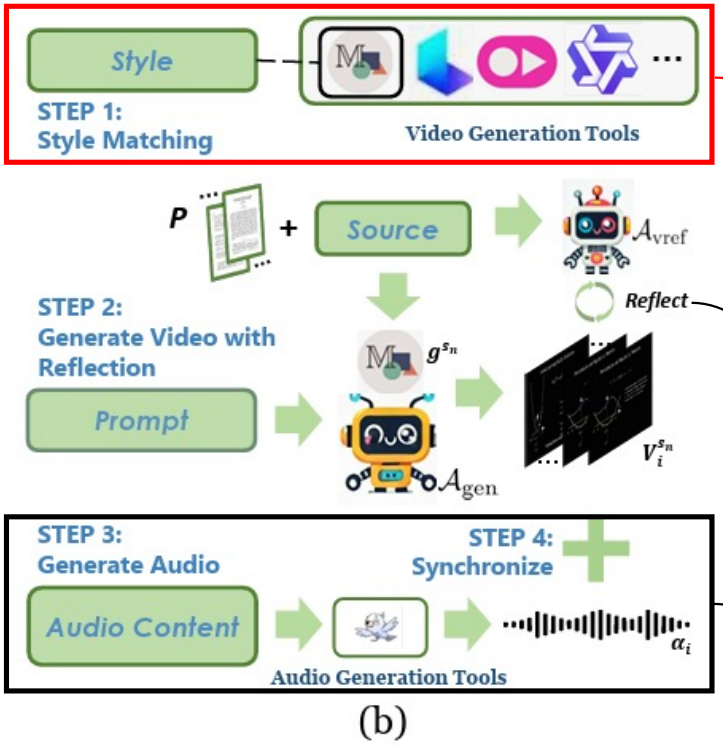
- The video generation agent 3) reconstruct the key scenes, generating both video and corresponding audio.
- Each video segment is synthesized from these elements, and all segments $\{V_1^{S_1}, \ldots, V_H^{S_N}\}$ are 4) integrated into the final video abstract.

Liu, Jingwei, et al. "Preacher: Paper-to-video agentic system." ICCV. 2025.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Generating Professional Video Abstracts



(b)

- Multiple video generation tools are integrated in **Video Generation Agent**.
- Six **video styles** are supported in Preacher: “talking heads,” “general,” “static concept,” “molecular visualization,” “slides,” and “mathematics”.
- **Evaluation criteria of reflect** for generated video segments are Accuracy, Professionalism, and Alignment between the video content and the paper.
- The corresponding **audio** will be also generated by the agent and finish the synchronization.

Liu, Jingwei, et al. "Preacher: Paper-to-video agentic system." ICCV. 2025.

Applications of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Preacher Evaluations

- **Dataset:** 40 research papers spanning five distinct fields: Mathematics, Molecular Biology, Geology, Machine Learning, and Climate Science.
- **Evaluation Metrics:**
  1) Accuracy: Correctness of the video content, free from errors.
  2) Professionalism: Use of domain-specific knowledge and expertise.
  3) Aesthetic Quality: Visual appeal, design, and overall presentation.
  4) Alignment with the Paper: Semantic Alignment with the paper.
  5) CLIP: text-image similarity score.
  6) Aesthetic Score (AE): evaluate the consistency with the prompt and aesthetic quality.

Method	GPT Evaluation				Human Evaluation				CLIP ↑	AE ↑
	Accuracy ↑	Professionalism ↑	Aesthetic ↑	Alignment ↑	Accuracy ↑	Professionalism ↑	Aesthetic ↑	Alignment ↑		
OpenAI-o3-mini [36] + StreamingT2V [19]	3.35(0.98)	4.03(0.87)	4.00(0.77)	3.60(0.91)	3.13(0.93)	3.83(0.96)	3.10(1.21)	3.87(0.86)	0.23(0.04)	4.99(0.67)
OpenAI-o3-mini + Wan 2.1-14B [60]	3.75(0.43)	4.53(0.48)	4.15(0.41)	4.33(0.69)	3.63(0.91)	4.45(0.49)	**4.33**(0.51)	4.23(0.69)	0.29(0.07)	5.29(0.47)
OpenAI-o3-mini + Kling 1.6 [27]	3.70(0.61)	4.18(0.79)	3.98(0.73)	4.05(0.83)	3.40(1.06)	4.23(0.87)	4.13(0.69)	3.78(0.89)	0.26(0.07)	5.18 (0.63)
OpenAI-o3-mini + Sora[35]	4.33(0.94)	4.45 (0.49)	**4.18**(0.67)	4.30(0.73)	3.88(0.86)	4.50(0.67)	4.30(0.49)	4.38(0.59)	**0.31**(0.06)	**5.31**(0.53)
Preacher (**Ours**)	**4.50**(0.55)	**4.63**(0.44)	4.17(0.69)	**4.35**(0.98)	**4.80**(0.46)	**4.78**(0.46)	4.25(0.58)	**4.75**(0.43)	0.26(0.09)	5.20(0.83)

Liu, Jingwei, et al. "Preacher: Paper-to-video agentic system." ICCV. 2025.

Summary of Agentic System

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Takeaways

- Most agentic systems are based on LLMs / LMMs (large multimodal models) with general and domain-specific tools.
- Agentic systems have been applied in general complex task planning, scientific research, et al.
- The workflow of agentic systems mainly contain planning, executing, reflecting and summarizing steps.
- The benchmarks for evaluating agentic systems are still limited.

Next section: How to gaurantee the responsibility of agentic systems?