



2025 AI+ Development  
Digital Summit

# AI+ 研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站





# 2025 AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

08/08-09 | 北京站

11/14-15 | 深圳站



查看会议详情

## 北京站论坛设置

大模型和 AI 应用评测

智能存储与检索技术

下一代知识工程

AI+ 金融业务创新

智能需求工程

智能体与研发效率工具

AI 产品运营与出海策略

大模型安全与对齐

大模型应用开发框架与实践

智能体经济 (Agentic Economy)

智能测试工具的开发与应用

具身智能与机器人

代码生成及其改进

AI+ 新能源汽车

AI 前沿技术探索与实践

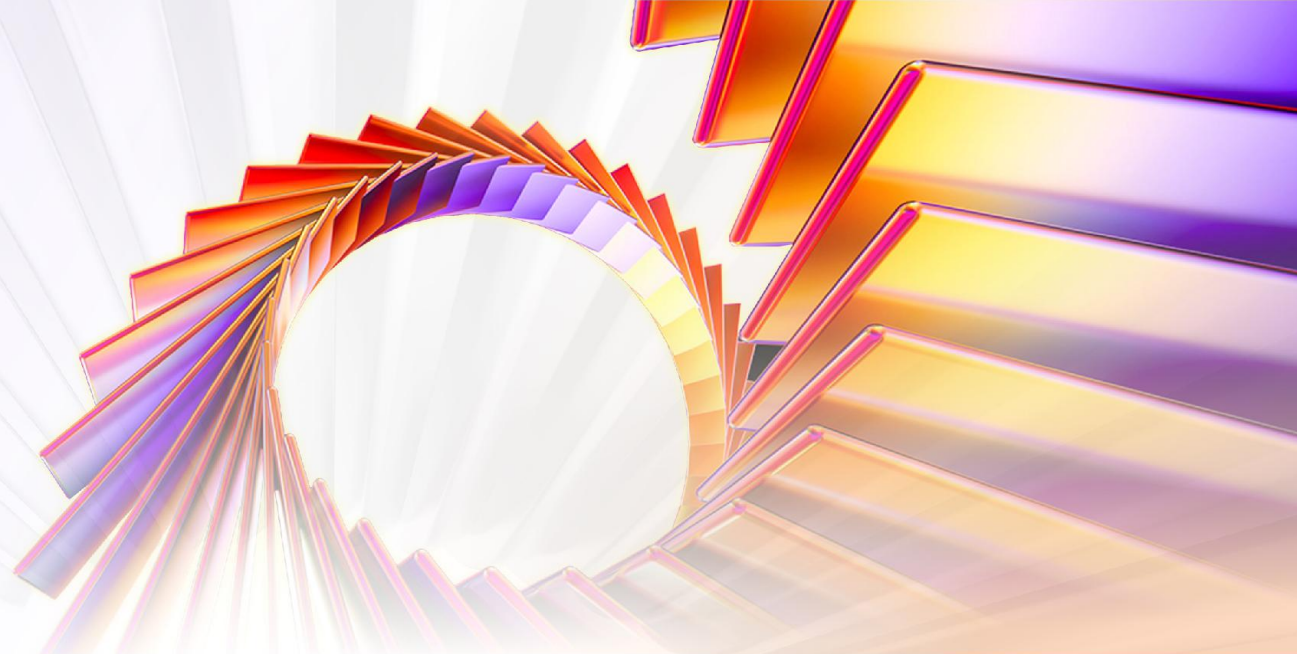


| 05/23-24 | 上海站

**2025** AI+ Development  
Digital Summit

# AI+研发数字峰会

拥抱AI 重塑研发



# 多模态、多端手机智能体 Mobile-Agent

徐海洋 | 阿里巴巴通义实验室



## 徐海洋

阿里巴巴通义实验室 高级算法专家

---

阿里通义实验室高级算法专家，负责通义多模态大模型mPLUG、Mobile-Agent系列工作，包括基础多模态模型mPLUG/mPLUG-2，多模态对话大模型mPLUG-Owl/Owl2，多模态文档大模型mPLUG-DocOwl，多模态智能体Mobile-Agent、PC-Agent等，其中 mPLUG 工作在 VQA 榜单首超人类的成绩，Mobile-Agent工作CCL2024 Best Demo，获得多个多模态榜单第一和Best Paper。在国际顶级期刊和会议ICML/NeurIPS/ICLR/CVPR/ICCV/ACL/EMNLP等发表论文50多篇，并担任多个顶级和会议AC/PC/Reviewer。主导参与开源项目mPLUG，Mobile-Agent，AliceMind，DELTA。

# 目录

## CONTENTS

- I. 大模型智能体背景
- II. 多模态手机智能体Mobile-Agent
- III. 多模态PC智能体PC-Agent
- IV. Mobile-Agent开源应用

# PART 01

## 大模型智能体背景



“如果一篇论文提出了某种不同的训练方法，我们内部的Slack上会嗤之以鼻，认为都是我们玩剩下的。但是当新的AI Agents论文出来的时候，我们会认真兴奋的讨论”

– Andrej Karpathy



“AI Agent不仅会改变每个人与计算机交互方式。它还将颠覆软件行业，带来自我们从键入命令到点击图标以来最大的计算变革” – 比尔盖茨

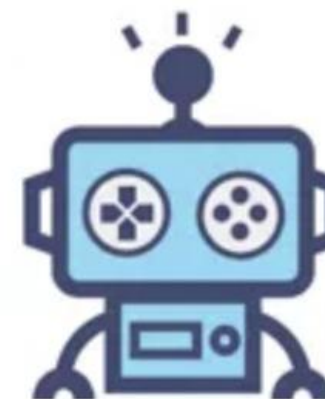




OpenAI Five



DeepMind AlphaStar



LLM Agent with ChatGPT

## 传统基于RL的智能体的局限性

数据采样专有环境和低效

面向特定任务

稀疏奖励和长时段问题

## 大模型智能体的优势

丰富的世界知识

推理/规划能力

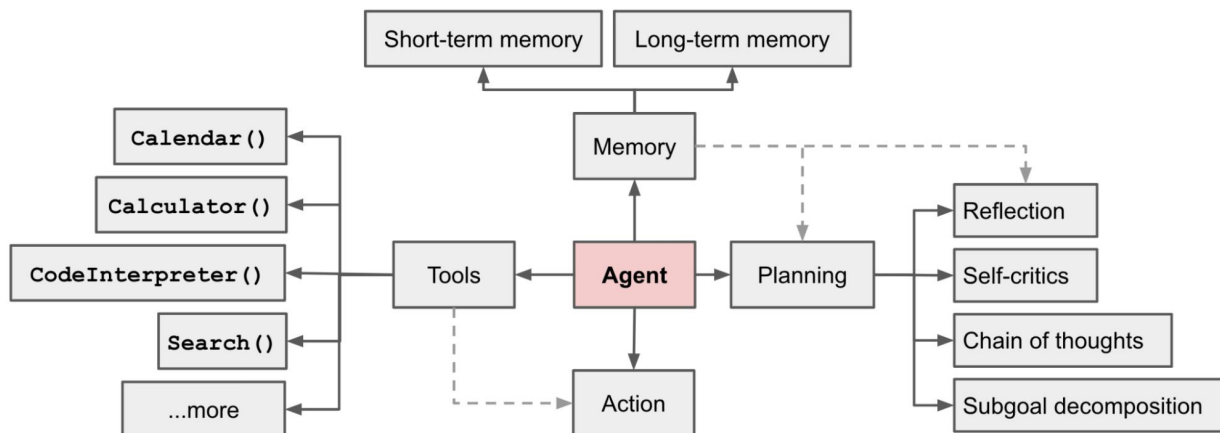
工具使用（检索、code等）

In-context Learning

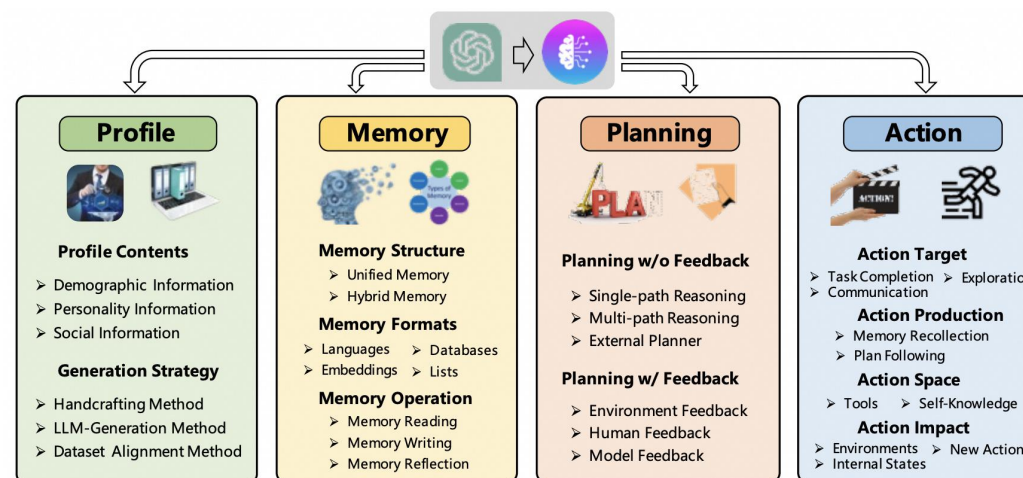


在人工智能领域，AI智能体指可以观察周遭 **环境** 并作出 **行动** 以达致 **目标** 的 **自主** 实体

## Agent System Overview from Lilian Weng's blog

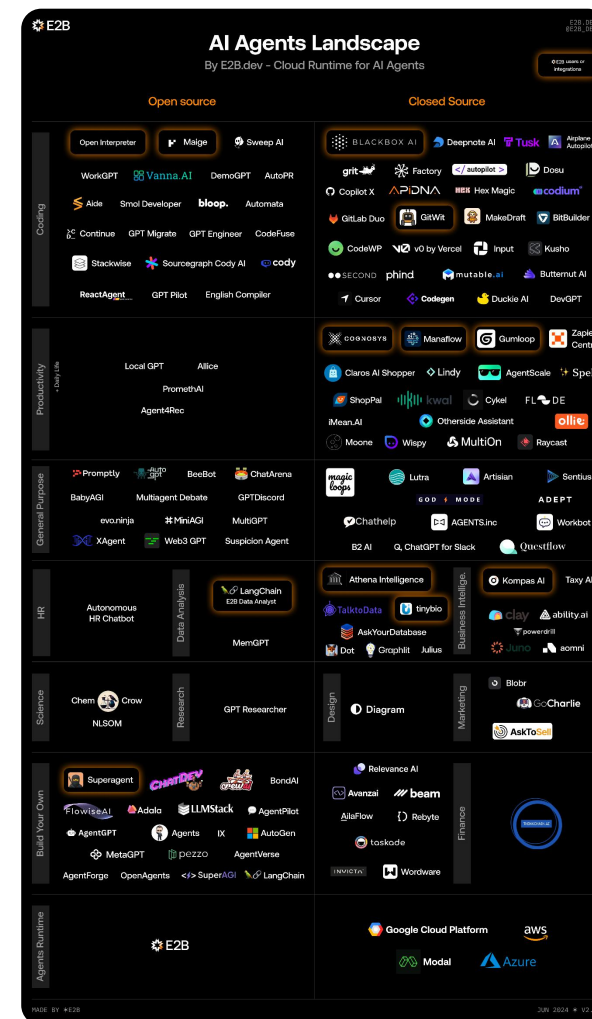
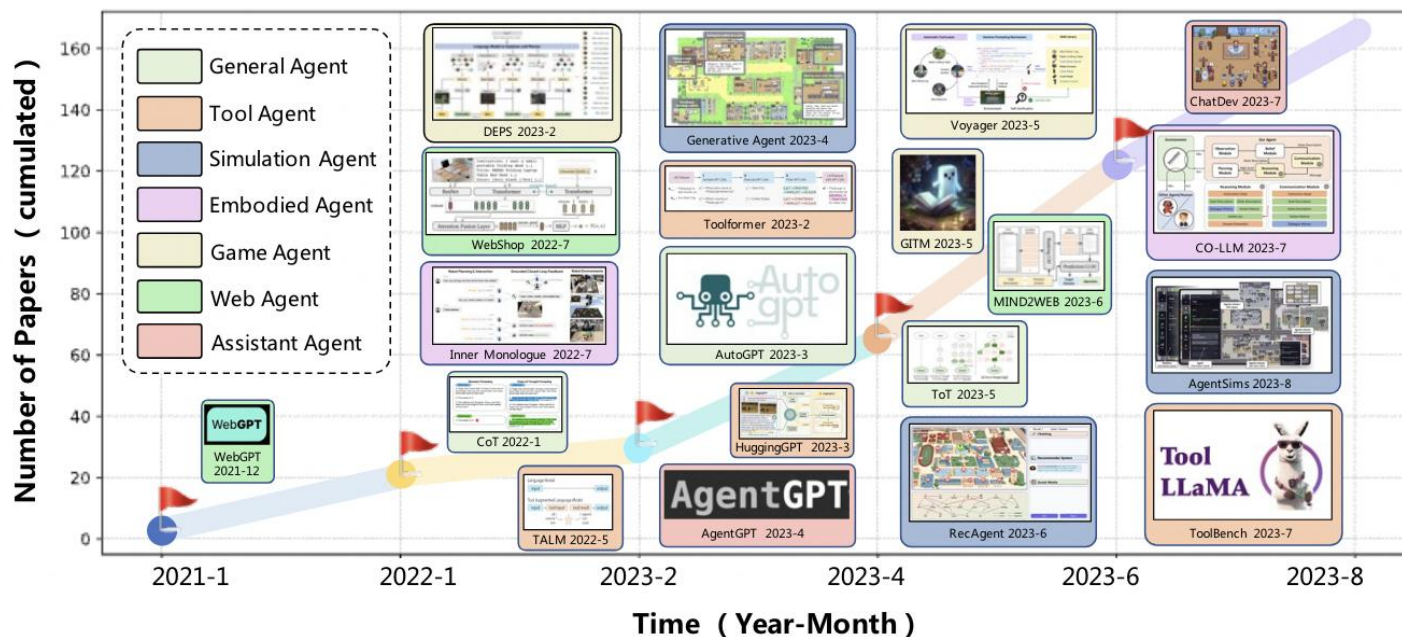


## Wang et al. A Survey on Large Language Model based Autonomous Agents

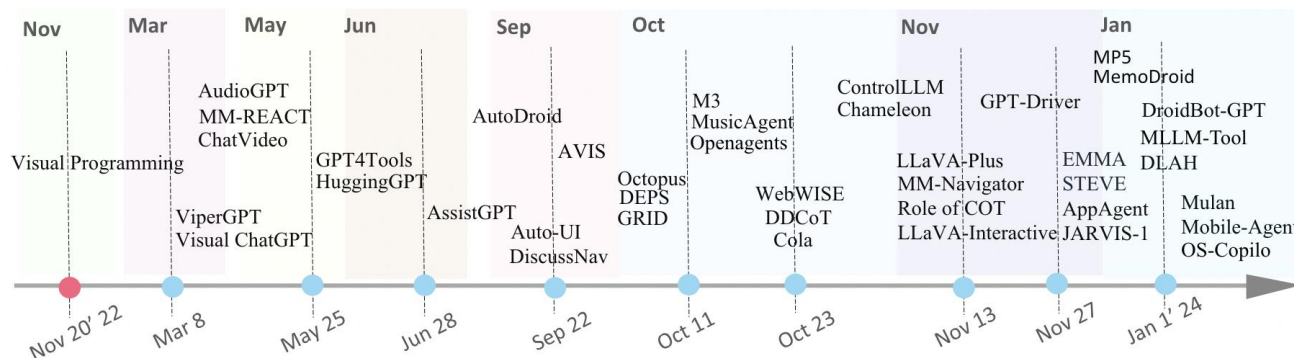
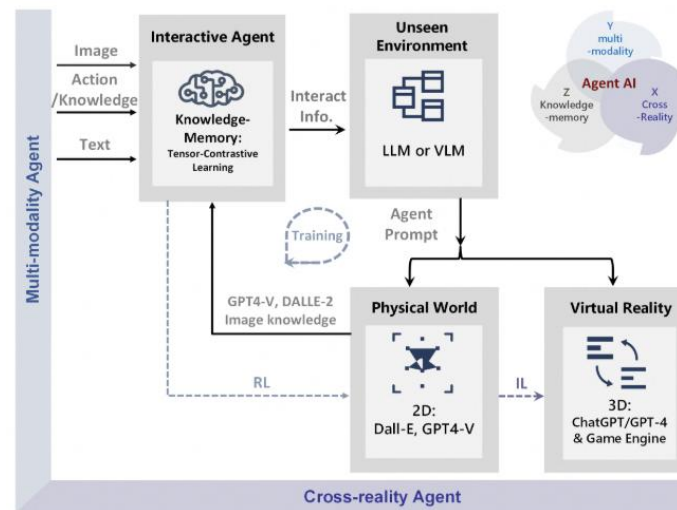
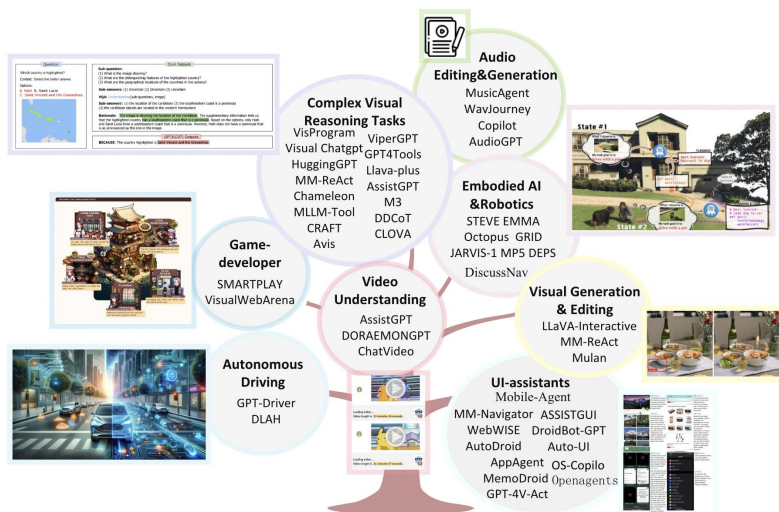


# 大模型智能体发展迅速

大模型广泛使用后，各类大模型智能体模型、框架、应用呈现井喷趋势



现实世界是需要 **多模态环境交互** 的，多模态智能体可能衍生出更多Super、Fancy应用



现实世界是需要 **多模态环境交互** 的，多模态智能体可能衍生出更多Super、Fancy应用



### Benchmark Comparison

How WritingBench compares to existing writing evaluation benchmarks

BENCHMARK	QUERIES	PRIMARY DOMAINS	SECONDARY DOMAINS	STYLE REQ.	FORMAT REQ.	LENGTH REQ.	AVG. INPUT TOKENS
EQ-Bench	241	1	—	✗	✗	✗	130
LongBench-Write	100	7	—	✗	✗	✓	87
HelixBench	647	5	38	✗	✗	✓	1,270
WritingBench	1,239	6	100	✓	✓	✓	1,546

**Comprehensive Coverage**  
WritingBench covers 6 primary domains and 100 secondary subdomains, ensuring a thorough evaluation of writing capabilities across diverse contexts.

**Multi-Dimensional Requirements**  
Unique among benchmarks, WritingBench evaluates style, format, and length requirements, assessing the complexity of real-world writing tasks.

**Extended Input Context**  
With an average input of 1,546 tokens and maximum of 9,361 tokens, WritingBench tests models' ability to handle complex, lengthy contexts.

### Introduction

Recent advancements in large language models (LLMs) have significantly enhanced text generation capabilities, yet evaluating their performance in generative writing remains a challenge. Existing benchmarks primarily focus on generic text generation or are limited in writing tasks, failing to capture the diverse requirements of high-quality written content across various domains.

WritingBench addresses this gap by providing a comprehensive evaluation framework designed specifically for assessing writing capabilities across multiple dimensions, including style, format, and length requirements.

#### Key Contributions

- A comprehensive benchmark with 1,239 queries across 6 primary domains and 100 subdomains
- A query-dependent evaluation framework with instance-specific criteria generation
- A fine-tuned critic model for criteria-aware scoring with 83% human alignment
- Demonstration of framework effectiveness through data curation and model enhancement.

#### Example Query from WritingBench

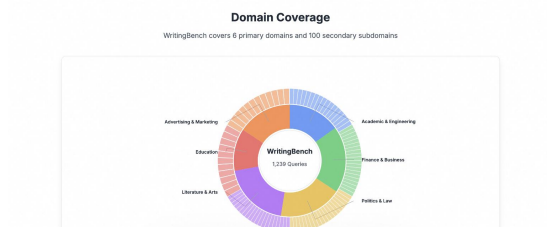
I need to write a research proposal focused on quantum computing applications in financial risk modeling. The proposal should follow the IEEE conference template and be approximately 2,500 words. Use a formal academic tone with precise technical language.

Include the following sections:  
- Abstract (300 words)  
- Introduction with problem statement  
- Literature Review  
- Methodology  
- Expected Results  
- Timeline  
- Budget Justification  
- References (using IEEE citation style)

Please reference these recent papers on quantum algorithms for financial modeling:  
[1] Chen et al., "Quantum Risk Modeling: A New Paradigm in Finance," *IEEE Transactions on Financial Engineering and Applied Computing*, 2023.  
[2] Wang and Liu, "Hybrid Quantum-Classical Algorithms for Portfolio Optimization," *Quantum Information Science*, 2024.

Content specificity: Normal requirement Length constraint: 2500 words  
Style requirement: IEEE conference template  
Domain relevance: Finance

Each WritingBench query is designed with multiple requirements and evaluation dimensions, simulating real-world writing scenarios.



Find and book me the highest rated one-day tour of Rome on Tripadvisor.

I'll search for the highest-rated tour of historic Rome on TripAdvisor. Once I find a suitable option, I'll provide you with the details. Let's begin.

Worked for 2 minutes ^

Navigating to TripAdvisor website  
Selecting "Things to Do" category  
Searching for historic Rome tours  
Closing pop-up, continuing tour search  
Searching for Historic Rome tours  
Exploring all historic Rome tour options  
Closing Colosseum tab, resuming tour search  
Closing tour pop-up, tab afterward  
Exploring options for top-rated tours  
Sorting results by tour ratings  
Exploring filters for top-rated tours  
Scrolling for sorting options, finding tours

## Claude 3.7 sonnet (computer use)

参照人类思考系统的快速反应与慢反思结合的工作模式，将LLM快速响应和思维链深度思考

[https://mp.weixin.qq.com/s/JVxbkR587\\_Mxpu02k29Kw](https://mp.weixin.qq.com/s/JVxbkR587_Mxpu02k29Kw)

## Operator

基于Computer-Using Agent 模型，结合GPT-4o的视觉理解能力和强化学习习得的推理能力，自动执行鼠标和键盘的组合操作，无需API，具备推理思维链和自动纠错能力

## 搜索索引能力 + 生成式 AI 技术



### GPT-4o

Help me find iOS and android adoption rates, % who want to learn another language, and change in mobile penetration, over the past 10 years, for top 10 developed and top 10 developing countries by GDP. Lay this info out in a table and separate stats into columns, and include recommendations on markets to target for a new iOS translation app from ChatGPT, focusing on markets ChatGPT is currently active in.

Compiling comprehensive data on iOS and Android adoption rates, the percentage of populations interested in learning another language, and changes in mobile penetration over the past decade for the top 10 developed and developing countries by GDP is challenging due to data limitations. However, based on

### Deep research

Help me find iOS and android adoption rates, % who want to learn another language, and change in mobile penetration, over the past 10 years, for top 10 developed and top 10 developing countries by GDP. Lay this info out in a table and separate stats into columns, and include recommendations on markets to target for a new iOS translation app from ChatGPT, focusing on markets ChatGPT is currently active in.

Top 10 Developed Economies (by GDP) – Smartphone OS Share, Language-Learning Interest, and Mobile Penetration Growth

Country	iOS	Android	Interest in New	Mobile
---------	-----	---------	-----------------	--------

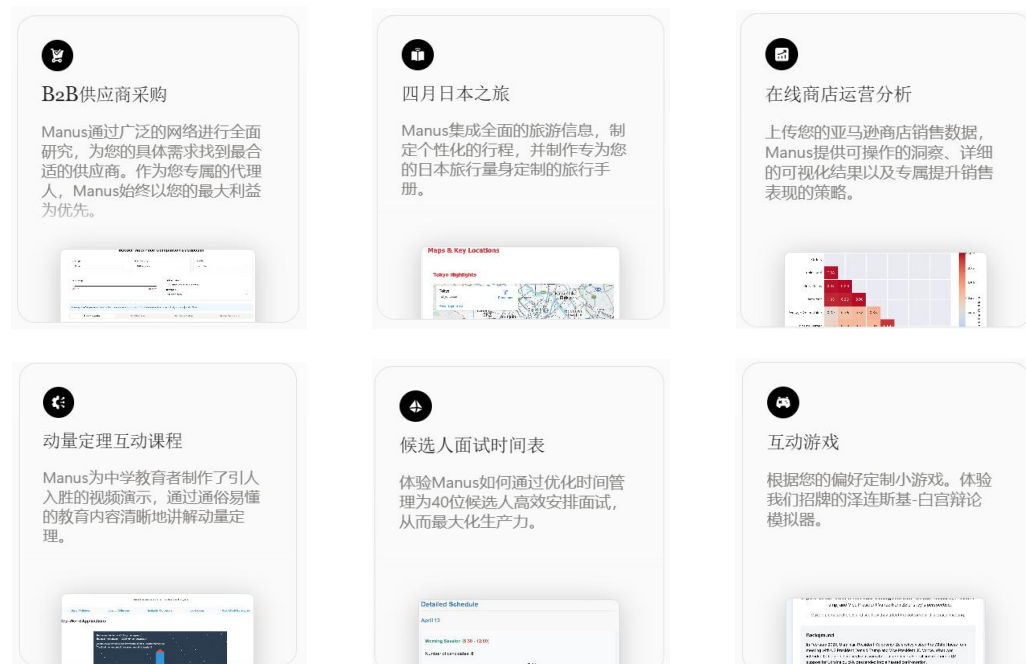
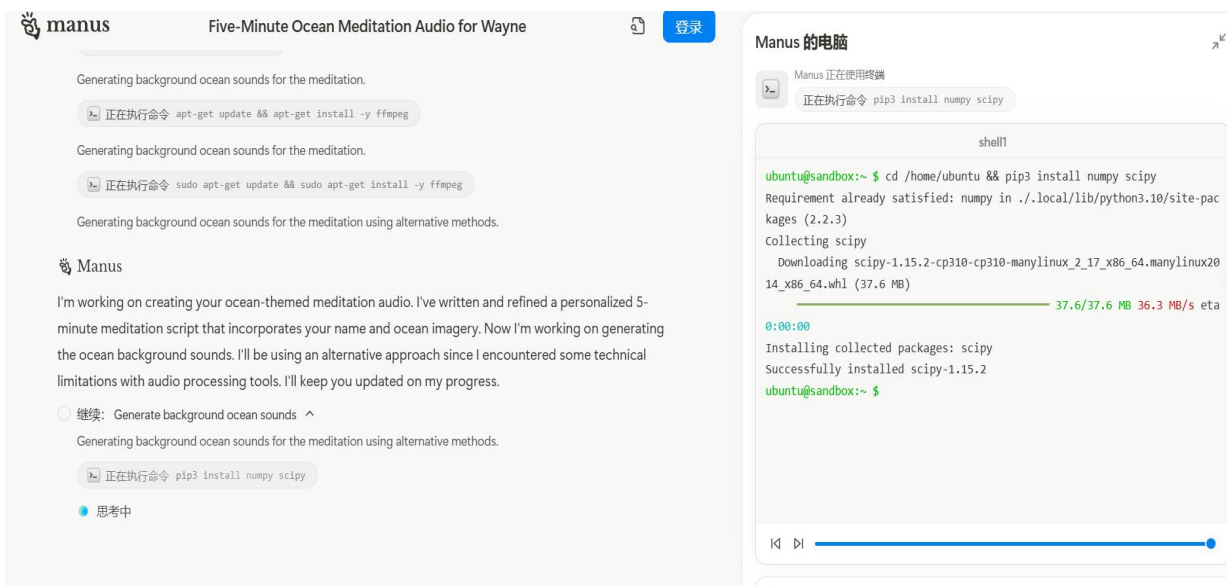
## DeepResearch

整合Google 的搜索索引能力和 OpenAI 的生成式 AI 技术，实现从简单查询到复杂推理的跨越，优化多模态搜索（如文本、图像、视频的综合检索）和实时知识更新能力，持续进行多次搜索-分析-优化的循环，完成专业级报告生成



从基于检索提供信息，到Agent执行任务的本质进阶

(1) 规划-执行Tool-反思; (2) 操作上网; (3) 快操作 + 慢思考



## Manus/Open Manus

Manus强调“需求→规划→执行→交付”全流程自动化，无需用户持续指导便可能直接生成可交付成果，动态调整执行路径，在解决现实世界问题方面表现卓越

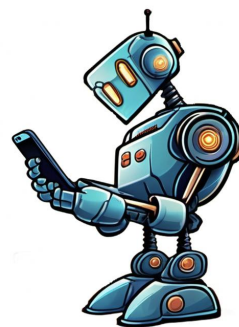


## PART 02

# 多模态手机智能体Mobile-Agent

## 一句指令实现自动操作手机

1. 纯视觉方案，不依赖系统数据
2. 可以多个应用之间操作
3. 感知、规划、反思三者结合
4. 无需训练、即插即用



# Mobile-Agent

Mobile-Agent: The Powerful Mobile Device Operation Assistant Family



<https://github.com/X-PLUG/MobileAgent>



CCL2024 唯一 Best Demo

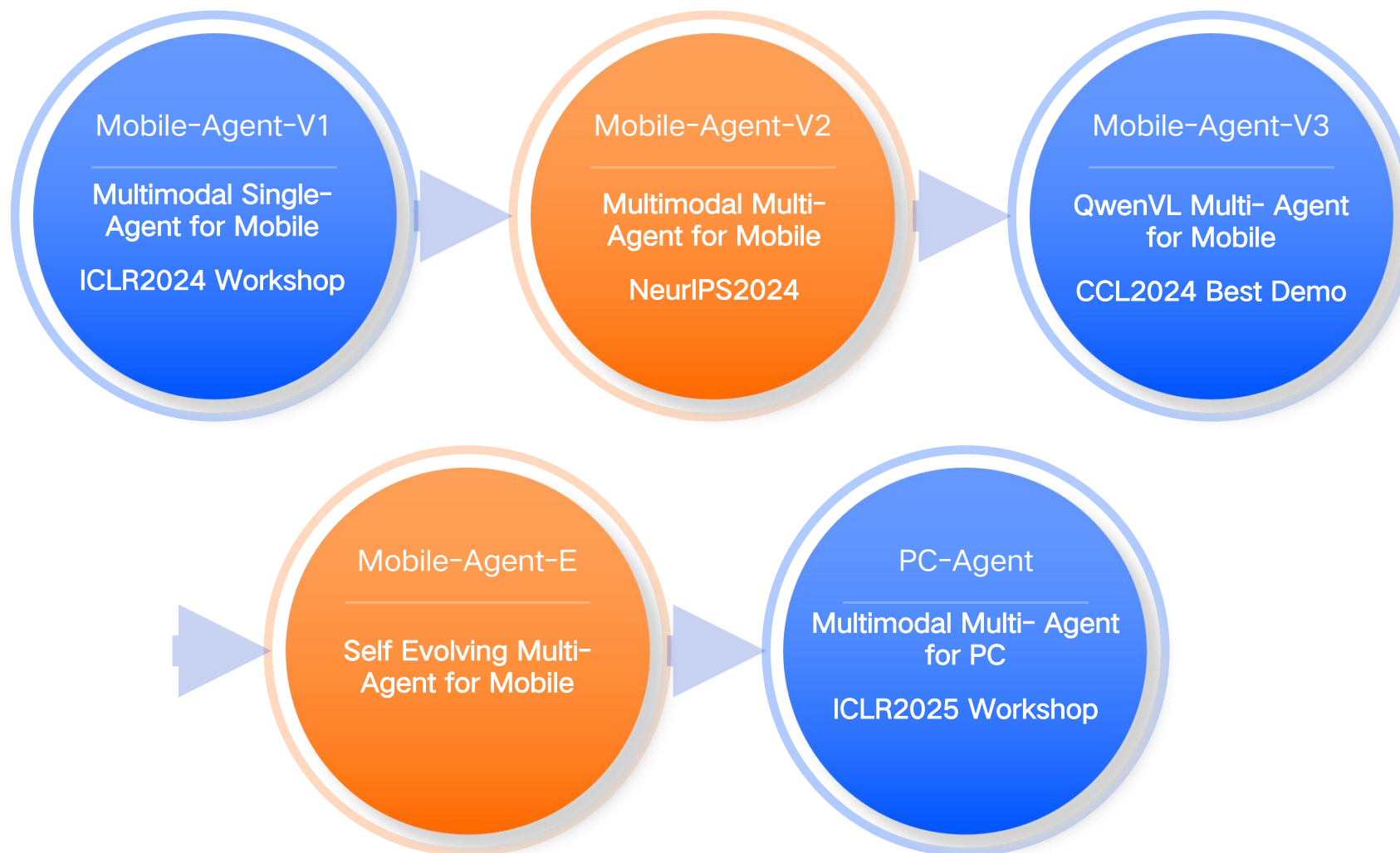




# 多模态智能体 Mobile-Agent

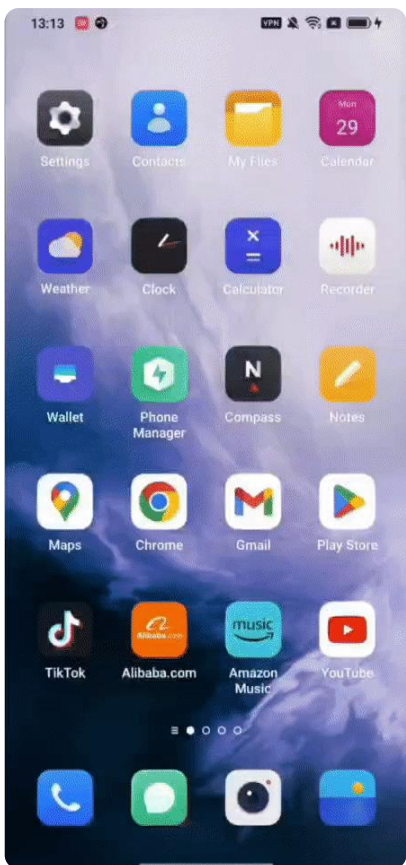


# 多模态智能体Mobile-Agent

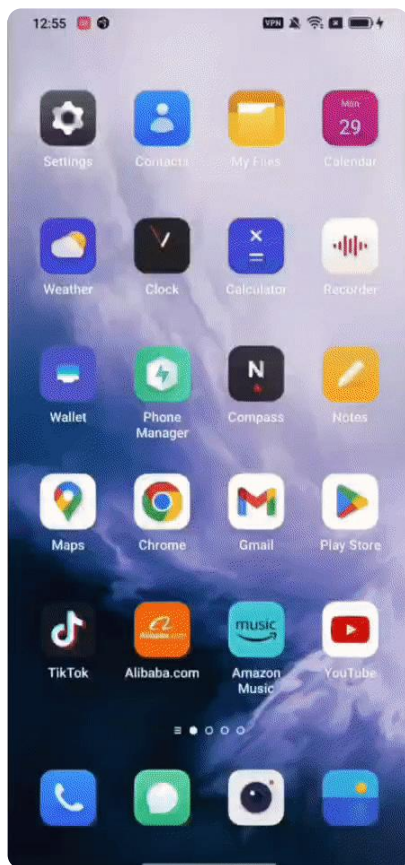


# 多模态手机智能体 Mobile-Agent-V1

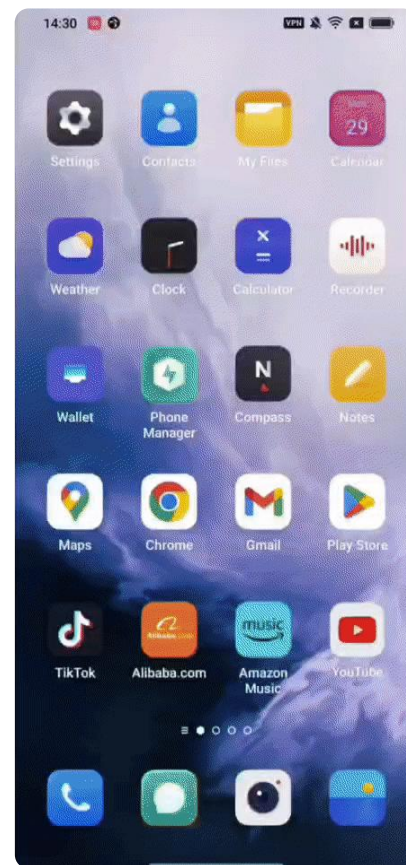
## 分析天气



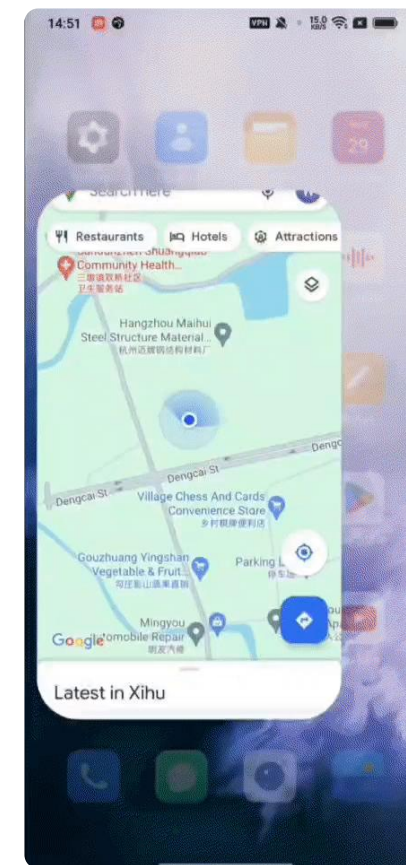
## 搜索视频并评论



## 刷短视频并点赞



## 导航



# 多模态手机智能体Mobile-Agent-V1

Instruction: [Rules and operations of the game].  
Help me play this game.

click text (5)



click text (出牌)



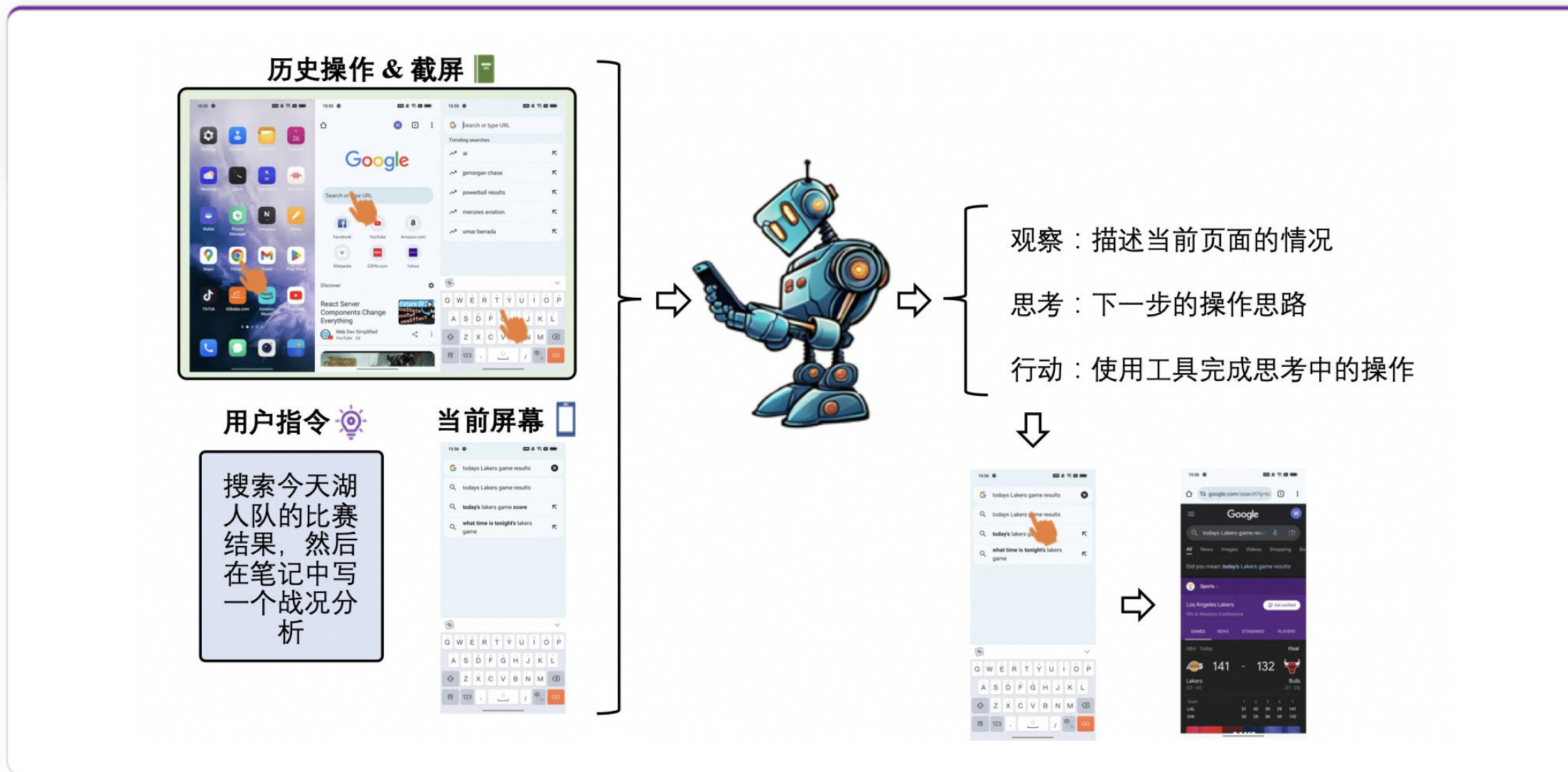
click text (7)



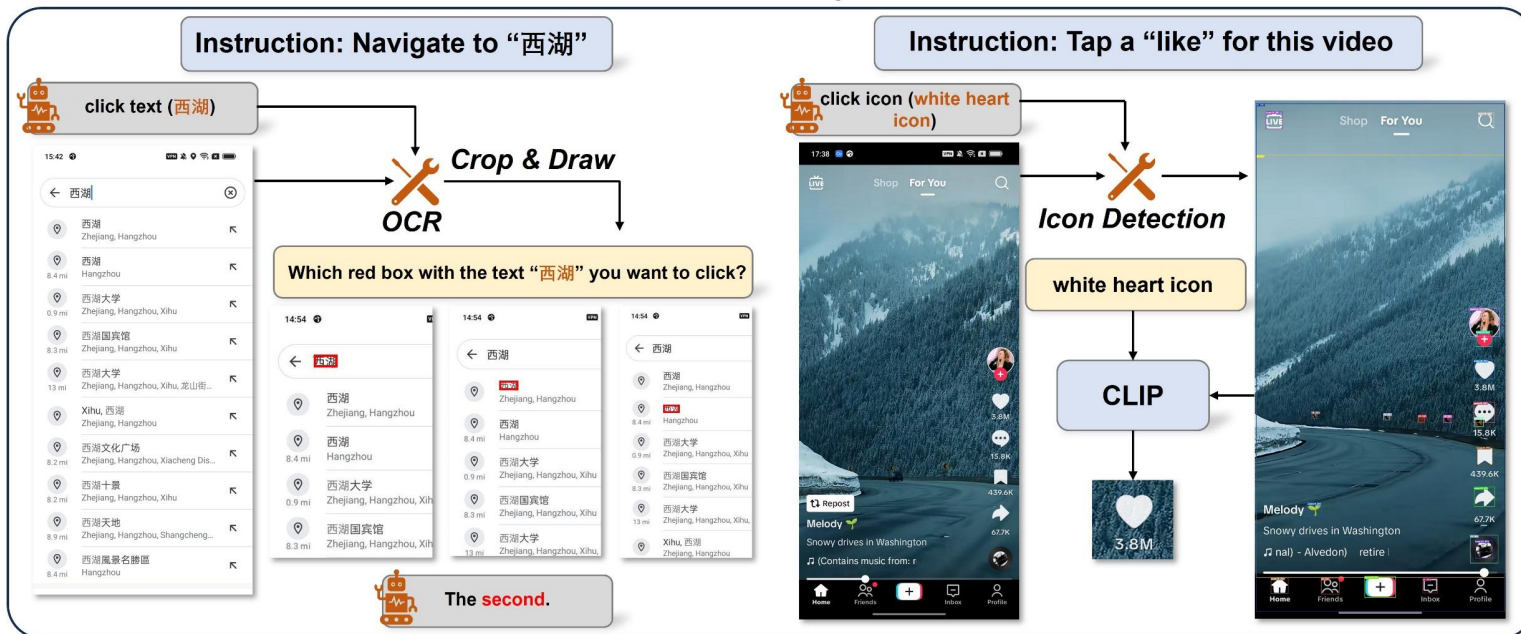
click text (出牌)



# 多模态手机智能体 Mobile-Agent-V1



## Visual Perception



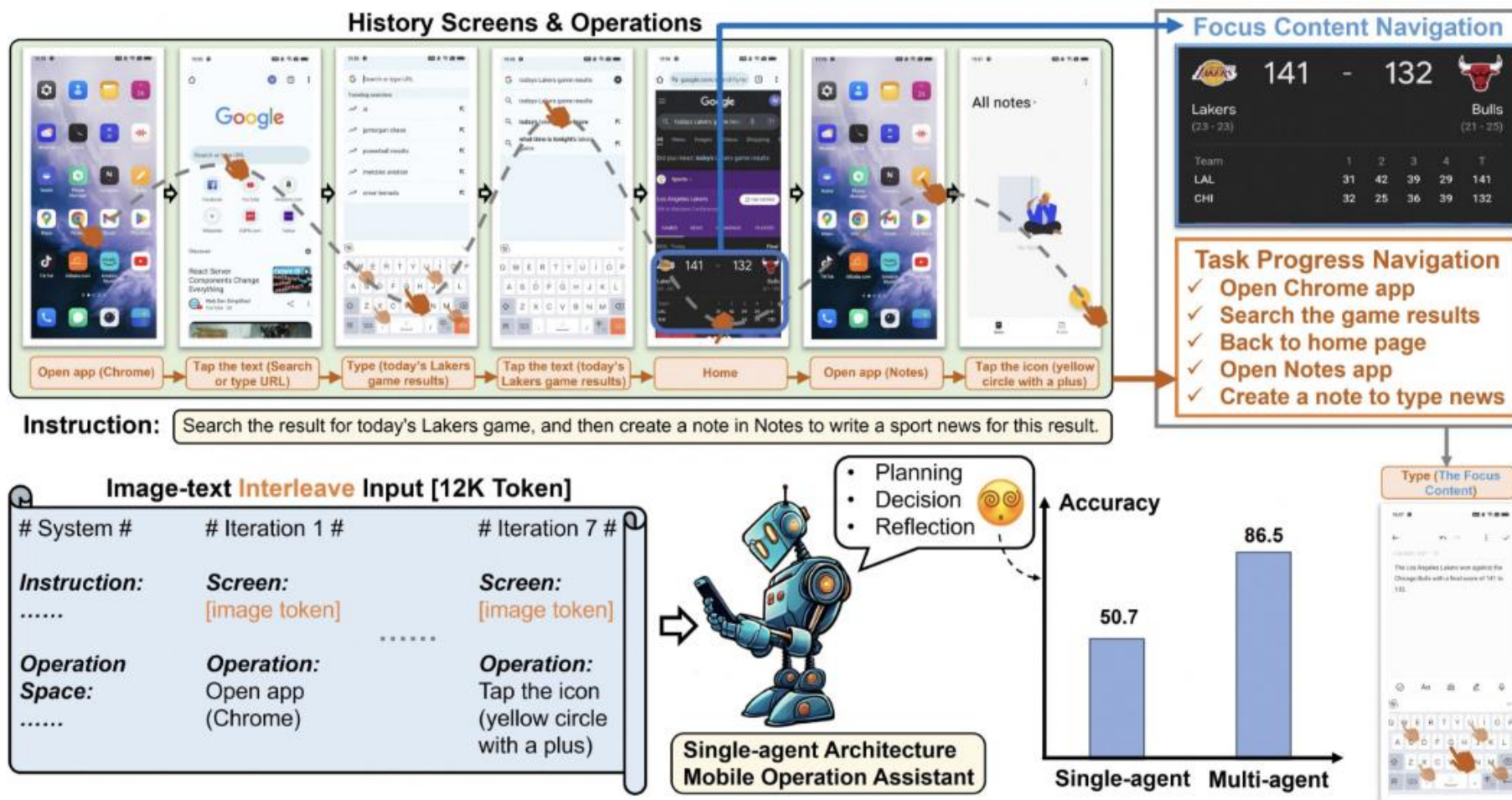
## 行为空间

1. 点击文本
2. 点击图标
3. 打字
4. 上划 & 下划
5. 返回上一页面
6. 返回桌面
7. 结束

## 大模型缺乏输出精确坐标的grounding能力

- 屏幕文本定位：使用OCR工具检测识别文本框
- 图标定位：使用图标分割检测工具检测所有图标和位置

冗长并且图文交错格式的操作历史，会大大增加智能体追踪任务进度的难度



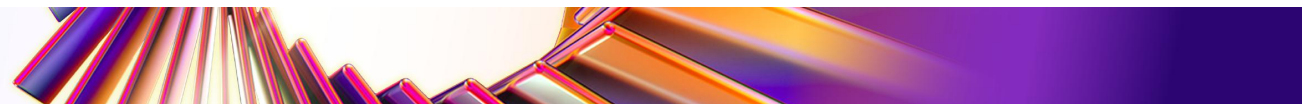
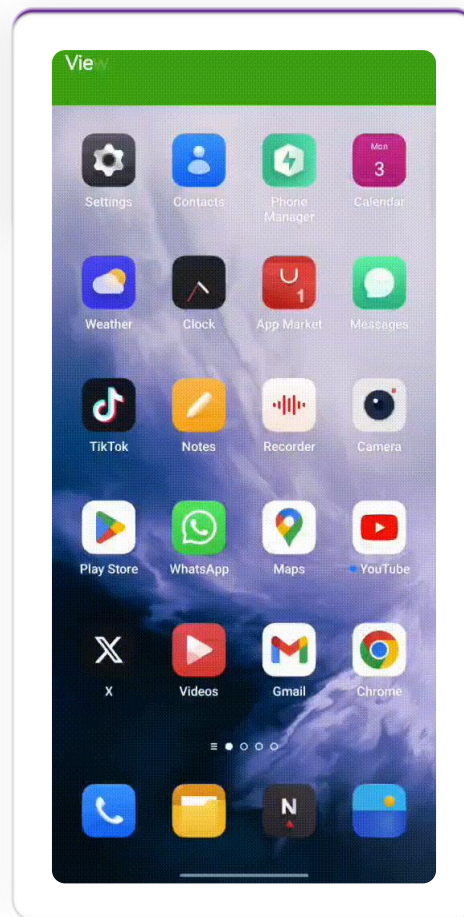
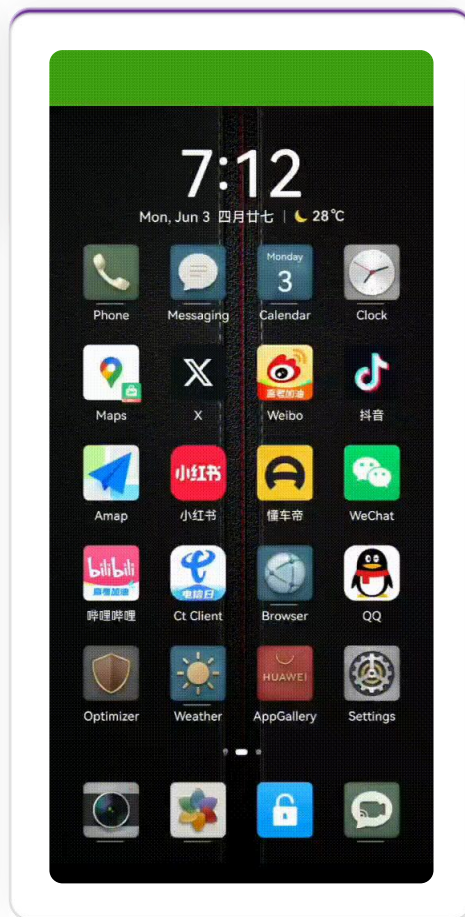
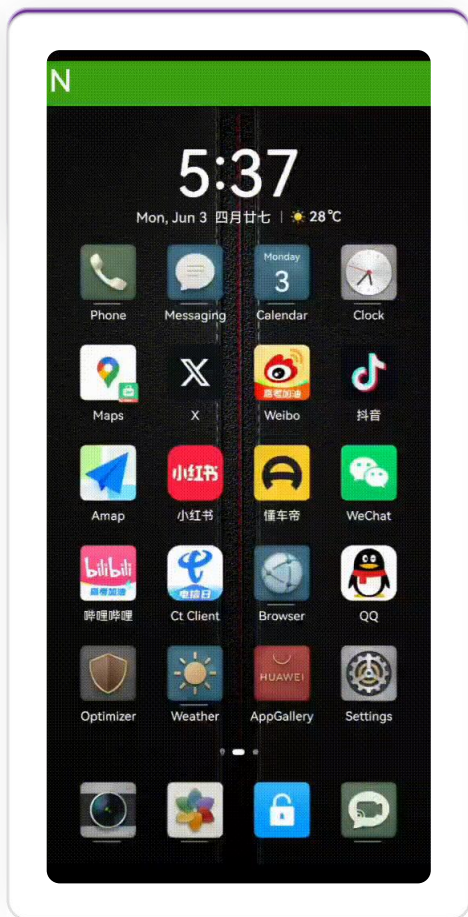
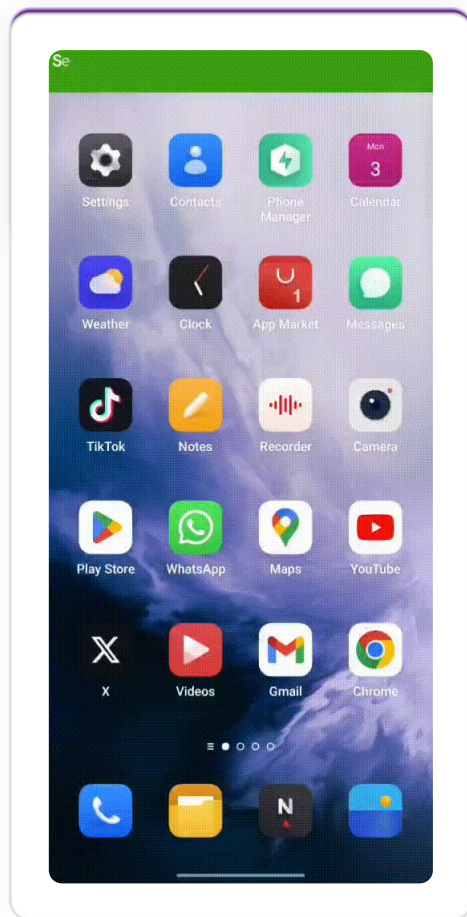
# 多模态手机智能体 Mobile-Agent-V2



- 首次在手机操作任务上采用多智能体架构，并延续了一代的纯视觉方案
- 多智能体各司其职，实现了更有效的任务进度追踪、任务相关信息记忆和操作反思
- 更强大的复杂指令拆解能力、跨应用操作能力和多语言场景操作能力

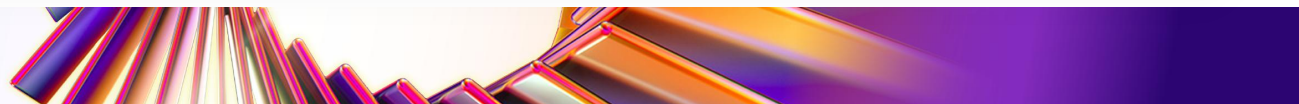
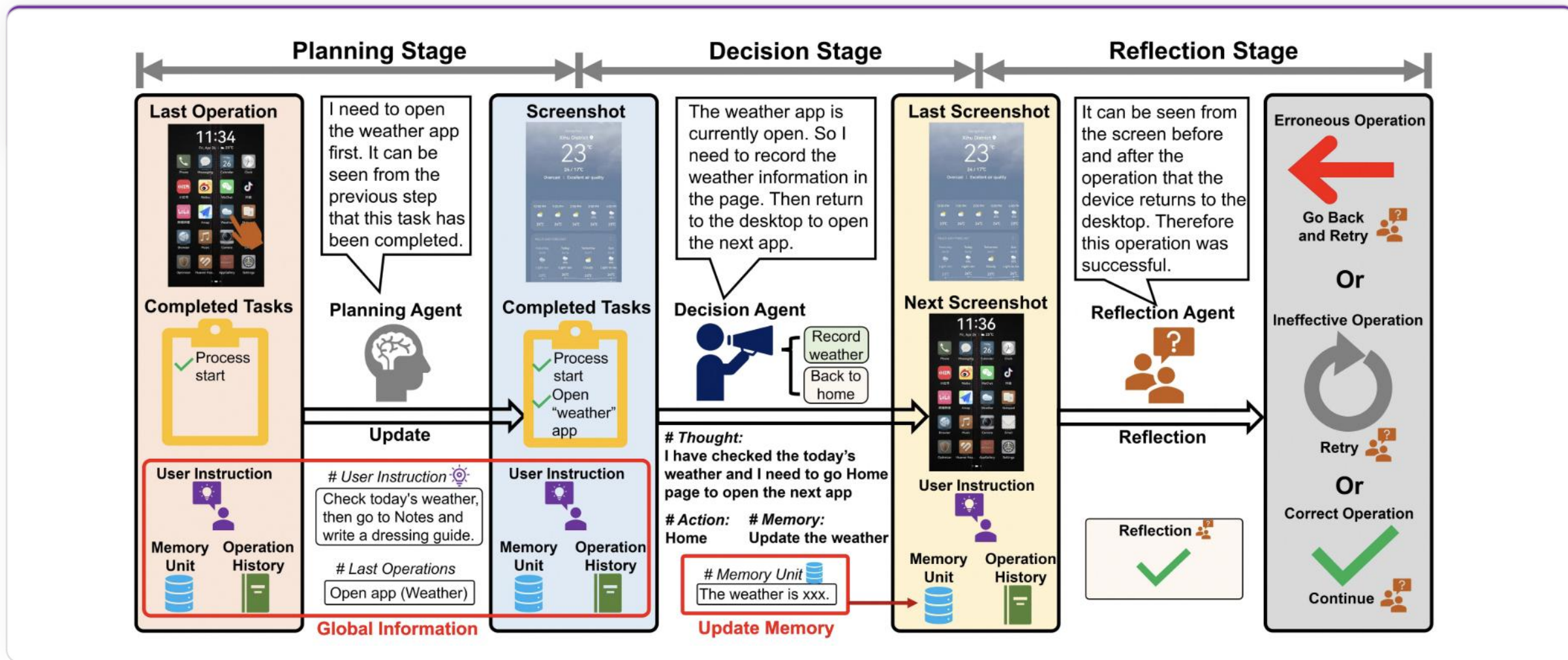


# 多模态手机智能体 Mobile-Agent-V2



# 多模态手机智能体Mobile-Agent-V2

采用多智能体框架，包括Planning Agent、Decision Agent、Reflection Agent



**动态评测：** 5个系统内置应用和5个第三方应用，每个APP和多个APP各2条基础指令和2条进阶指令

Method	Basic Instruction				Advanced Instruction			
	SR	CR	DA	RA	SR	CR	DA	RA
	<i>System app</i>							
Mobile-Agent	5/10	41.2	37.6	-	3/10	37.3	32.9	-
Mobile-Agent-v2	9/10	86.8	82.5	93.3	6/10	82.7	78.2	84.4
Mobile-Agent-v2 + Know.	10/10	97.5	98.2	98.9	8/10	88.9	87.2	91.4
	<i>External app</i>							
Mobile-Agent	2/10	38.3	35.4	-	1/10	29.2	27.0	-
Mobile-Agent-v2	8/10	97.9	94.0	92.5	5/10	77.9	74.1	78.8
Mobile-Agent-v2 + Know.	10/10	99.1	95.6	97.3	8/10	87.8	83.0	85.9
	<i>Multi-app</i>							
Mobile-Agent	1/2	52.8	50.0	-	0/2	33.3	31.4	-
Mobile-Agent-v2	2/2	100	92.9	91.6	2/2	100	93.8	92.9
Mobile-Agent-v2 + Know.	-	-	-	-	-	-	-	-

Table 1: Dynamic evaluation results on non-English scenario, where the *Know.* represents manually injected operation knowledge.

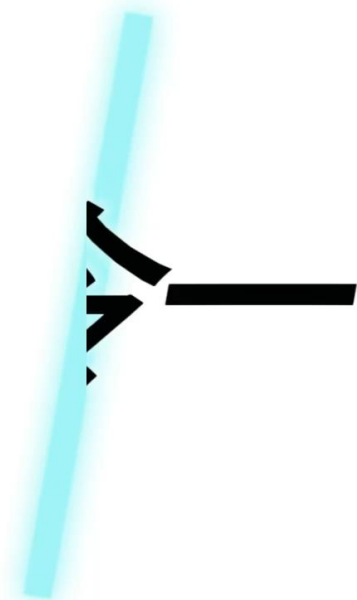
Method	Basic Instruction				Advanced Instruction			
	SR	CR	DA	RA	SR	CR	DA	RA
	<i>System app</i>							
Mobile-Agent	9/10	92.5	89.7	-	4/10	62.0	71.3	-
Mobile-Agent-v2	9/10	95.0	92.9	96.5	6/10	76.0	77.6	88.4
Mobile-Agent-v2 + Know.	10/10	100	96.2	98.7	8/10	85.3	87.9	92.0
	<i>External app</i>							
Mobile-Agent	7/10	79.7	72.0	-	3/10	45.3	38.7	-
Mobile-Agent-v2	9/10	97.1	93.8	96.2	7/10	89.7	91.0	93.4
Mobile-Agent-v2 + Know.	10/10	100	98.2	97.4	9/10	97.1	94.2	98.5
	<i>Multi-app</i>							
Mobile-Agent	2/2	100	91.2	-	1/2	86.7	92.9	-
Mobile-Agent-v2	2/2	100	97.4	100	1/2	93.3	93.3	80.0
Mobile-Agent-v2 + Know.	-	-	-	-	2/2	100	100	100

Table 2: Dynamic evaluation results on English scenario, where the *Know.* represents manually injected operation knowledge.

**Metrics.** We design the following four metrics for dynamic evaluation:

- **Success Rate (SR):** When all the requirements of a user instruction are fulfilled, the agent is considered to have successfully executed this instruction. The success rate refers to the proportion of user instructions that are successfully executed.
- **Completion Rate (CR):** Although some challenging instructions may not be successfully executed, the correct operations performed by the agent are still noteworthy. The completion rate refers to the proportion of correct steps out of the ground truth operations.
- **Decision Accuracy (DA):** This metric reflects the accuracy of the decision by the decision agent. It is the proportion of correct decisions out of all decisions.
- **Reflection Accuracy (RA):** This metric reflects the accuracy of reflection by the reflection agent. It is the proportion of correct reflections out of all reflections.





- **纯视觉理解方案**，模拟人对屏幕的感知，可操作**跨多个APP**的复杂任务
- 多智能体协作机制，通过规划、决策、反思，实现**自主推理**，达到泛化能力和容错机制
- **强安全干预机制**，避免高危操作
- **端+云协同部署**，端侧模型处理本地图片，云上模型思考推理，避免隐私安全问题



## 邀请函

APSARA 云栖大会 | 通义

### AI时代智能终端 分论坛

阿里云通义实验室

2024年9月20日 云栖小镇国际会展中心D馆 D3-5

随着模型技术的飞速发展，智能终端设备作为模型和用户的最直接入口载体，拥有巨大的想象空间。过去的一年，从新形态硬件（AI Pin、Rabbit）、智能玩具、智能穿戴、智能手机、智能电视到智能汽车，这些设备通过集成先进的大模型算法，不仅提升了用户体验，也为企业带来了新的增长点和产品竞争优势。本次分论坛将聚焦于通义大模型在终端领域的技术探索和最新实践，以及这些技术带来的行业创新发展潮流。

## 论坛议程

- 09:30-09:45 **大模型时代的智能终端和芯片发展**  
杨斌  
阿里云通义大模型终端发展总经理
- 09:45-10:00 **签约&发布仪式**  
战略合作签约仪式  
Mobile Agent产品发布
- 10:00-10:20 **大语言模型在NVIDIA Drive 平台的应用**  
卓睿  
NVIDIA软件高级总监
- 10:20-10:40 **边缘为主的混合计算成为生成式AI应用的新趋势**  
陆忠立 博士  
联发科技计算与人工智能技术事业群副总经理
- 10:40-11:00 **智舱AI，人车关系新范式**  
蔡明  
斑马智行首席产品官

# 自主进化手机智能体 Mobile-Agent-E

**Mobile-Agent-E**

$A_M$   
 $\uparrow$   
 $A_O - A_R - A_N$

Hierarchical Agents

Self-Evolution

Shortcuts

**Name:** Tap\_Type\_and\_Enter  
**Arguments:** ["x", "y", "text"]  
**Description:** Tap an input box at position (x, y), Type the "text", and then perform the Enter operation. Very useful for searching ...  
**Precondition:** There is a text input box on the screen with no previously entered content.

Tips

1. By default, no apps are opened in the background ...  
 14. If a tap action does not work as expected, consider tapping alternative areas of the screen ...

Long-term Memory

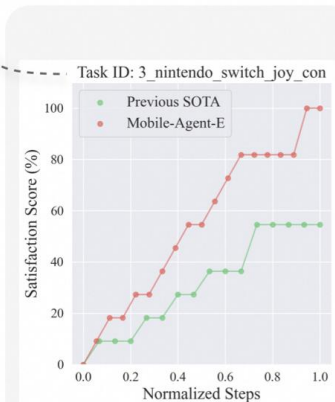
**Task Query:** I want to buy a brand-new Nintendo Switch Joy-Con. Any color is fine. Please compare the prices on Amazon, Walmart, and Best Buy. Find the cheapest option and stop at the screen where I can add it to the cart.

Low-Level Actions: Tap, Swipe, Home, Tap\_Type\_and\_Enter

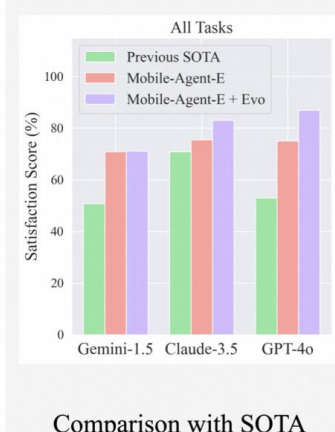
High-level Plans:

1. Open Amazon and search for Nintendo Switch
2. Note the price and proceed to the next App
3. Open Walmart ... 4. Note the Price ...
5. Open Best Buy ... 6. Note the Price ...
7. Open the app with the cheapest price ...

Mobile-Agent-E on a long-horizon, reasoning-intensive, multi-app task



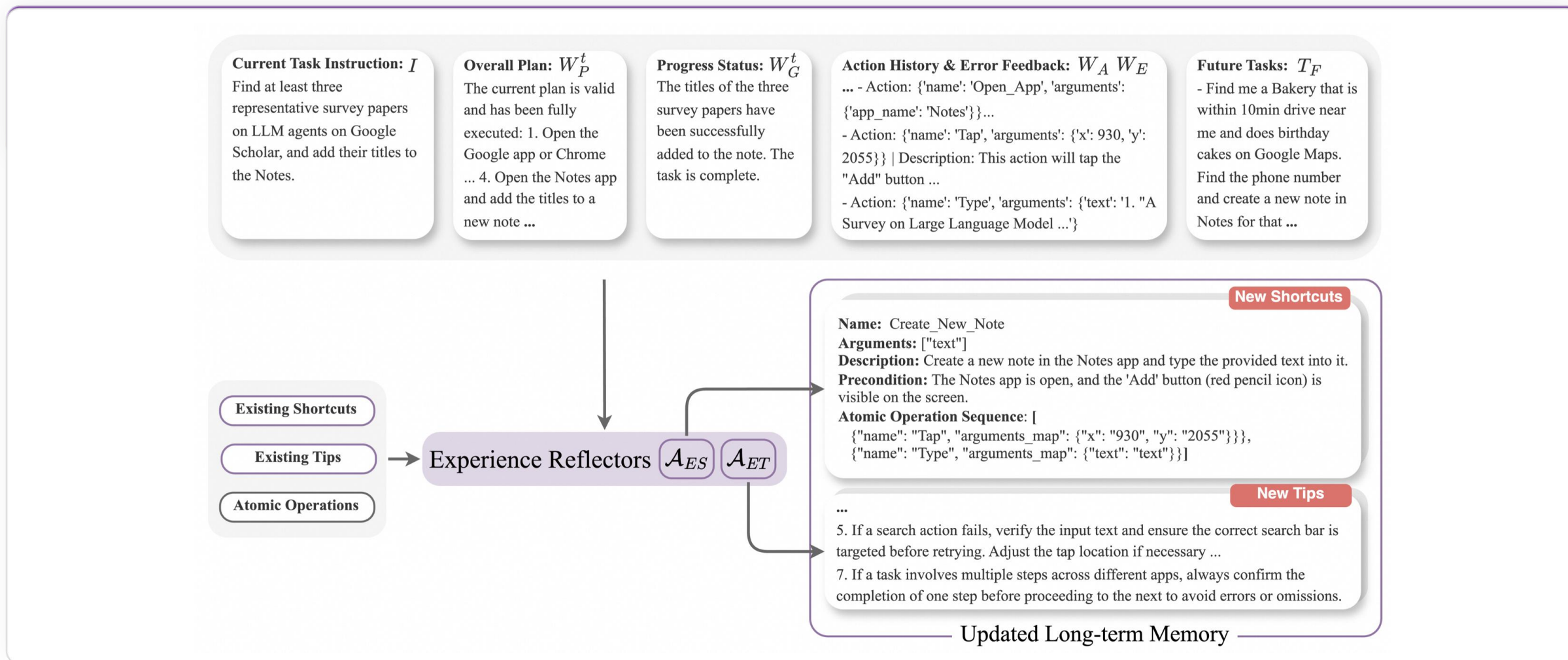
**复杂指令:**  
 执行复杂推理、多步规划以及跨App操作



**自我进化:**  
 反思过往的任务记录，从经验中学习，自动生成Tips和Shortcut

多平台比价任务示例

两个经验 Experience Reflectors 会根据当前任务的操作记录和错误日志等信息，对 Tips 和 Shortcuts 进行可能的优化和更新



## PART 03

# 多模态PC智能体PC-Agent



# PC Agent

**Institute for Intelligent Computing of Alibaba Group**

<https://arxiv.org/abs/2502.14282>

[Github: https://github.com/X-PLUG/MobileAgent/tree/main/PC-Agent](https://github.com/X-PLUG/MobileAgent/tree/main/PC-Agent)



# 多模态PC智能体PC-Agent

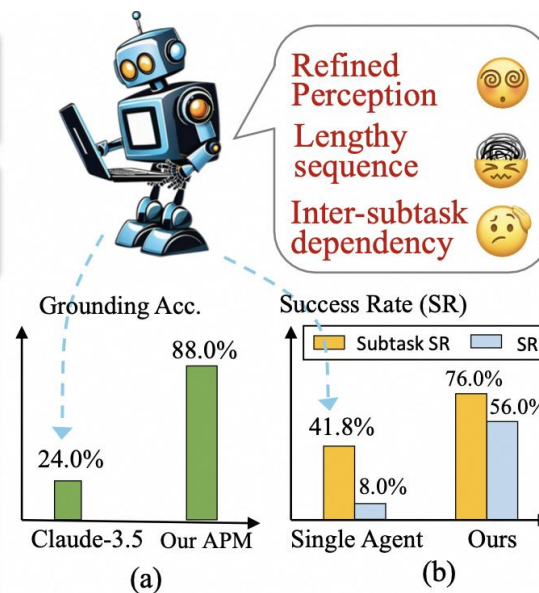
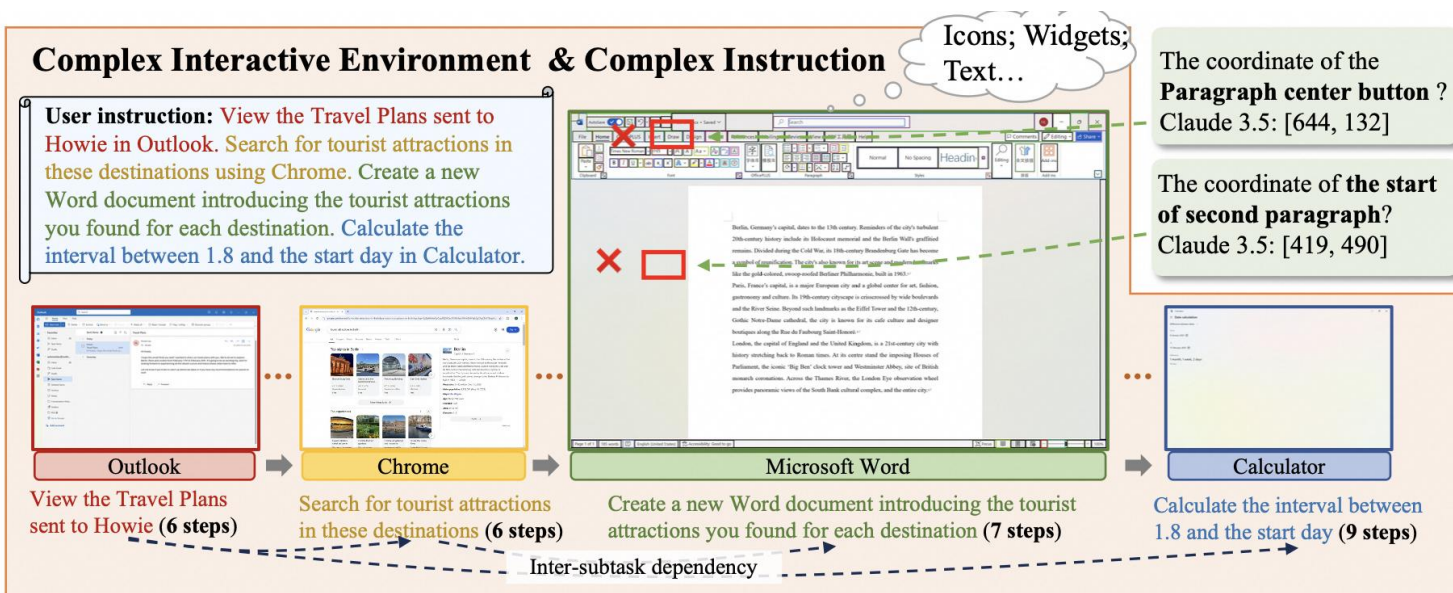
与Mobile场景相比，PC场景有两个难点：

## 1. 更复杂的交互环境

更密集多样的可交互元素，以及不同布局的文本

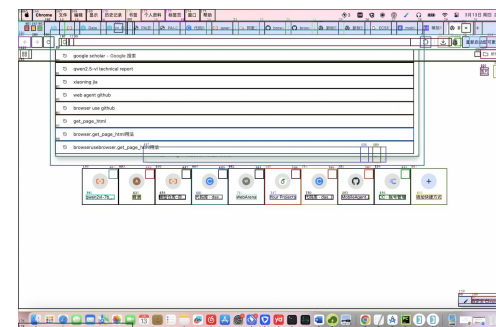
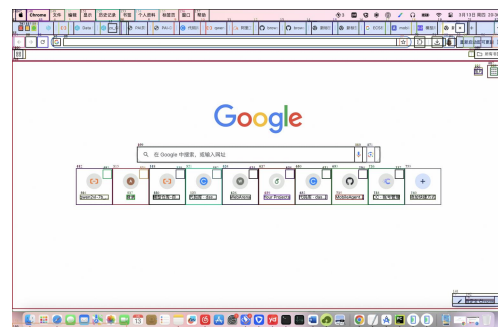
## 2. 更复杂的操作序列

PC常用于生产力场景，涉及更多复杂的App内及跨App workflow，因此包含更复杂的任务序列



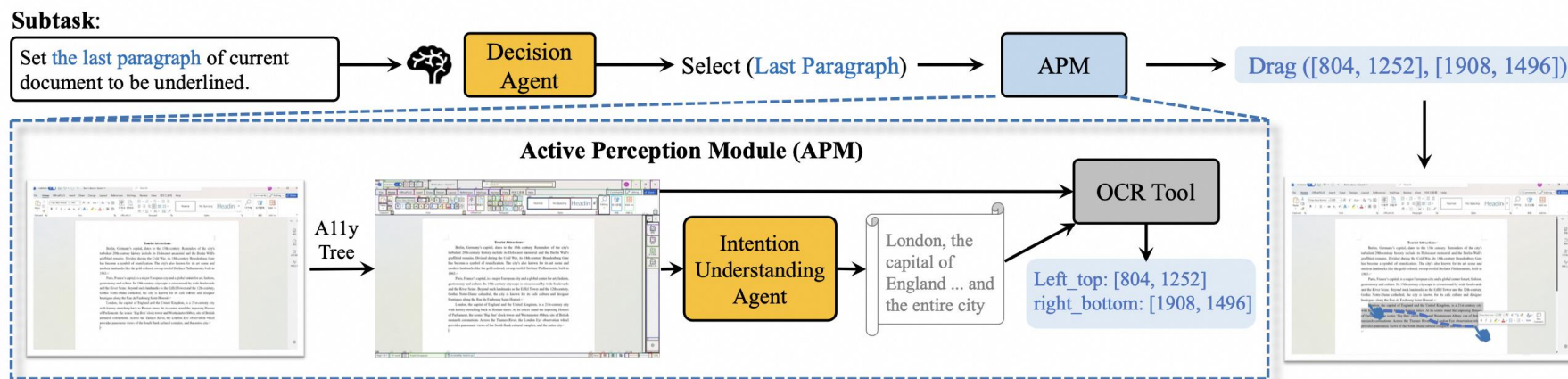
## 更复杂的交互环境

1. 更密集多样的可交互元素  
(Accessibility Tree)



```
{
  "text": "mark number: 85 icon: 关闭",
  "coordinates": [
    616,
    116
  ]
},
{
  "text": "mark number: 86 icon: 新标签页",
  "coordinates": [
    688,
    117
  ]
},
{
  "text": "mark number: 87 icon: 搜索标签页"
  "coordinates": [
    2984,
    117
  ]
}
```

2. 主动感知模块 (APM) : 引入select动作, 来选中/操作不同布局的文本



## 更复杂的操作序列

Manager Agent将复杂指令的执行分解为3个层次：**指令-子任务-动作**

指令层级

Manager Agent

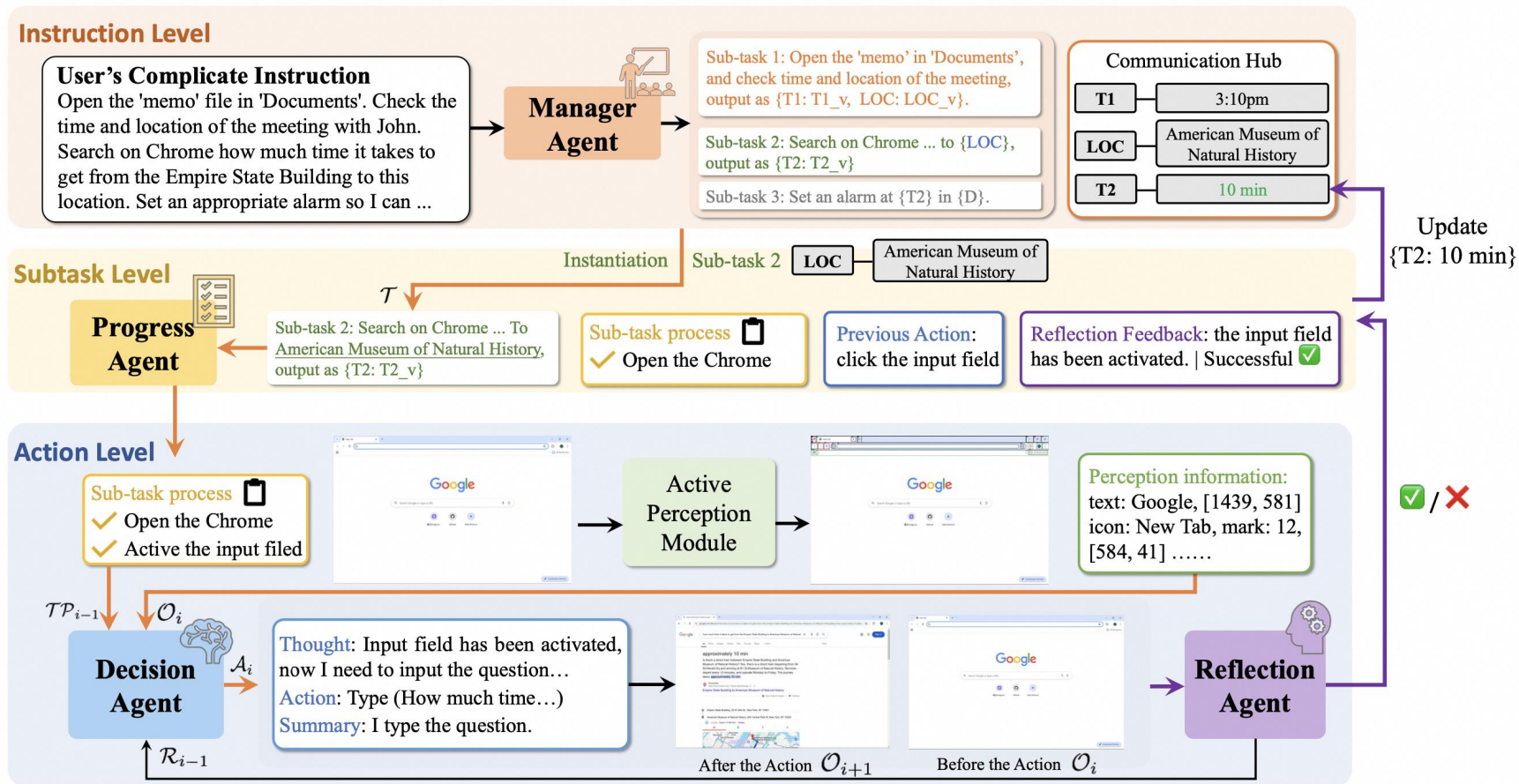
子任务层级

Progress Agent

动作层级

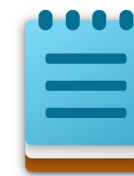
Decision Agent 决策

Reflection Agent 反思



## PC-Eval复杂指令测评集

8 个常用PC应用的 25 条复杂用户指令。每条指令由若干具有依赖关系的子任务构成，强调精细化操作及长程决策，并与现实场景 workflow 相对应。



Applications	Instruction	Steps
File Explorer Notepad, Clock Calculator	In the Notepad app, open the 'travel_plan' file in 'Documents', and check the time and location of the travel plans. Add the travel destination to the World Clock list on the Clock app. Calculate the interval between February 18 and the start time of the travel on the Calculator.	20
Chrome Excel	Search on Chrome for the total population of China, the United States, and India in 2024 respectively. Create a new spreadsheet in Excel, write the three countries' names in column A in descending order of population, and the corresponding populations in column B.	23
File Explorer Word	Open the 'test_doc1' file in 'Documents' in File Explorer, set the title to be bold, and set the line spacing of the first two paragraphs to 1.5x in Word.	8



## PC-Eval复杂指令测评集

Model	Type	Subtask SR (%) ↑	Success Rate (%) ↑
Gemini-2.0	Single-Agent	35.4%	0.0%
Claude-3.5		15.2%	0.0%
Qwen2.5-VL		46.8%	12.0%
GPT-4o		41.8%	8.0%
UFO (Zhang et al., 2024)	Multi-Agent	43.0%	12.0%
Agent-S (Agashe et al., 2024)		55.7%	24.0%
<b>PC-Agent (Ours)</b>		<b>76.0%</b>	<b>56.0%</b>

Ablation study			Subtask Success Rate	Success Rate
APM	Manager Agent	Reflection Agent		
	✓	✓	58.2%	20.0%
✓		✓	50.6%	12.0%
✓	✓		48.1%	12.0%
✓	✓	✓	<b>76.0%</b>	<b>56.0%</b>



## PART 04

# Mobile-Agent开源应用

# Mobile-Agent PC-Agent开源应用

The screenshot shows the GitHub repository page for Mobile-Agent. At the top, there is a table of files: LICENSE (last year), README.md (3 days ago), README\_ja.md (3 days ago), README\_zh.md (3 days ago), and index.html (2 months ago). Below this is the README section, which features a blue robot illustration and the text "Mobile-Agent: The Powerful Mobile Device Operation Assistant Family". It also includes a badge for "#5 Repository Of The Day" and a list of recent releases on Arxiv. On the right side, there are sections for Releases (no releases published), Packages (no packages published), Contributors (11), and Deployments (41).

<https://github.com/X-PLUG/MobileAgent>



The screenshot shows the file browser view of the Mobile-Agent repository, specifically the PC-Agent directory. The file list includes: .., PCAgent (Update crop.py), PCAgent\_v1 (update), .DS\_Store (v2), PC-Eval.json (Create PC-Eval.json), README.md (Update README.md), README\_v1.md (c1), config.json (v2), pywin.py (v2), and requirements.txt (Update requirements.txt). The left sidebar shows the repository structure with folders for Mobile-Agent-E, Mobile-Agent-v2, Mobile-Agent-v3, Mobile-Agent, and PC-Agent.

<https://github.com/X-PLUG/MobileAgent/tree/main/PC-Agent>







## 开始

! 目前仅安卓和鸿蒙系统（版本号 <= 4）支持工具调试。其他系统如iOS暂时不支持使用Mobile-Agent。

## 安装依赖

```
pip install -r requirements.txt
```



## 准备通过ADB连接你的移动设备

1. 下载 [Android Debug Bridge](#) (ADB)。
2. 在你的移动设备上开启“USB调试”或“ADB调试”，它通常需要打开开发者选项并在其中开启。
3. 通过数据线连接移动设备和电脑，在手机的连接选项中选择“传输文件”。
4. 用下面的命令来测试你的连接是否成功: `/path/to/adb devices`。如果输出的结果显示你的设备列表不为空，则说明连接成功。
5. 如果你用的是MacOS或者Linux，请先为 ADB 开启权限: `sudo chmod +x /path/to/adb`。
6. `/path/to/adb` 在Windows电脑上将是 `xx/xx/adb.exe` 的文件格式，而在MacOS或者Linux则是 `xx/xx/adb` 的文件格式。

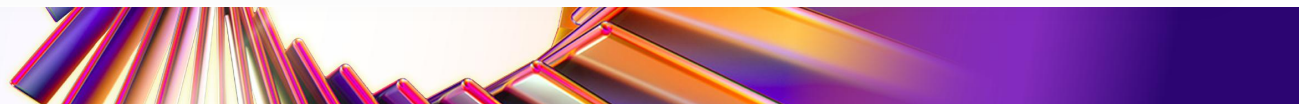
## 在你的移动设备上安装 ADB 键盘

1. 下载 ADB 键盘的 [apk](#) 安装包。
2. 在设备上点击该 apk 来安装。
3. 在系统设置中将默认输入法切换为“ADB Keyboard”。

## 选择适合的运行方式

1. 在 `run.py` 的22行起编辑你的设置，并且输入你的 ADB 路径，指令，GPT-4 API URL 和 Token。
2. 选择适合你的设备的图标描述模型的调用方法：
  - 如果您的设备配备了高性能GPU，我们建议使用“local”方法。它是指在本地设备中部署图标描述模型。如果您的设备足够强大，则通常具有更好的效率。
  - 如果您的设备不足以运行7B 大小的 LLM，请选择“api”方法。我们使用并行调用来确保效率。
3. 选择图标描述模型：
  - 如果选择“local”方法，则需要在“qwen-vl-chat”和“qwen-vl-chat-int4”之间进行选择，其中“qwen-vl-chat”需要更多的GPU内存，但提供了更好的性能与“qwen-vl-chat-int4”相比。同时，“qwen\_api”可以是空置的。
  - 如果您选择“api”方法，则需要在“qwen-vl-plus”和“qwen-vl-max”之间进行选择，其中“qwen-vl-max”需要更多的费用，但与“qwen-vl-plus”相比提供了更好的性能。此外，您还需要申请 [Qwen-VL 的 API-KEY](#)，并将其输入到“qwen\_api”。
4. 您可以在“add\_info”中添加操作知识（例如，完成您需要的指令所需的特定步骤），以帮助更准确地运行移动设备。
5. 如果您想进一步提高移动设备的效率，则可以将“reflection\_switch”和“memory\_switch”设置为“False”。
  - “reflection\_switch”用于确定是否在此过程中添加“反思智能体”。这可能会导致操作陷入死周期。但是您可以将操作知识添加到“add\_info”中以避免它。
  - “memory\_switch”用于决定是否将“内存单元”添加到该过程中。如果你的指令中不需要在后续操作中使用之前屏幕中的信息，则可以将其关闭。

[https://github.com/X-PLUG/MobileAgent/blob/main/Mobile-Agent-v2/README\\_zh.md](https://github.com/X-PLUG/MobileAgent/blob/main/Mobile-Agent-v2/README_zh.md)





# Mobile-Agent开源应用

**Mobile-Agent-v2: Mobile Device Operation Assistant with Effective Navigation via Multi-Agent Collaboration**

Github Code | Arxiv 2406.01014 | Stars 2.3k

If you like our project, please give us a star on Github for latest update.

**Terms of use**

1. Input your instruction in "Instruction", for example "Turn on the dark mode".
2. You can input helpful operation knowledge in "Knowledge".
3. Click "Submit" to get the operation. You need to operate your mobile device according to the operation and then upload the screenshot after your operation.
4. The 5 cases in "Examples" are a complete flow. Click and submit from top to bottom to experience.
5. Due to limited resources, each operation may take a long time, please be patient and wait.

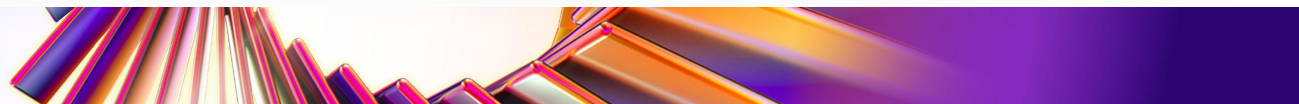
**使用说明**

1. 在"Instruction"中输入你的指令，例如“打开深色模式”。
2. 你可以在"Knowledge"中输入帮助性的操作知识。
3. 点击"Submit"来获得操作。你需要根据输出来操作手机，并且上传操作后的截图。
4. "Example"中的5个例子是一个任务。从上到下点击它们并且点击"Submit"来体验。
5. 由于资源有限，每次操作的时间会比较长，请耐心等待。

Screenshot	Instruction
	Turn on the dark mode
	Turn on the dark mode
	Turn on the dark mode
	Turn on the dark mode
	Turn on the dark mode

QR Code:

<https://modelscope.cn/studios/wangjunyang/Mobile-Agent-v2>



## Introduction

- PC-Agent is a multi-agent collaboration system, which can achieve automated control of productivity scenarios (e.g. Chrome, Word, and WeChat) based on user instructions.
- Active perception module designed for dense and diverse interactive elements are better adapted to the PC platform.
- The hierarchical multi-agent cooperative structure improves the success rate of more complex task sequences.

## Getting Started

### Installation

Now Windows is supported.

```
conda create --name pcagent python=3.10
source activate pcagent

# For Windows
pip install -r requirements.txt

git clone https://github.com/Topdu/OpenOCR.git
pip install openocr-python
```

### Configuration

Edit config.json to add your API keys and customize settings:

```
# API configuration
{
  "vl_model_name": "GPT-4o",
  "llm_model_name": "GPT-4o",
  "token": "sk-...", # Replace with your actual API key
  "url": "https://api.openai.com/v1"
}
```

### Test on your computer

1. Run the `run.py` with your instruction and your GPT-4o api token. For example,

```
python run.py --instruction="Create a new doc on Word, write a brief introduction of Alibaba, and save the doc"
```

2. Optionally, you can add specific operational knowledge via the `--add_info` option to help PC-Agent operate more accurately.
3. To further improve the operation efficiency of PC-Agent, you can set `--disable_reflection` to skip the reflection process. Note that this may reduce the success rate of the operation.
4. If the task is not very complex, you can set `--simple 1` to skip the task decomposition.

<https://github.com/X-PLUG/MobileAgent/blob/main/PC-Agent/README.md>



## 未来可能方向

### 技术角度：

- 更通用泛化的执行方式，API与操作相结合
- Agent RL训练提高决策能力
- 多模态内容的理解，image video audio
- 云端异步处理
- 手机和电脑端的联动（信息差）

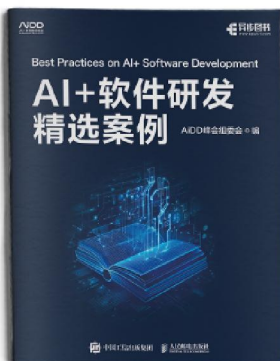
### 应用角度：

- 更好的人机交互过程
- 用户偏好记忆、个人隐私保护

# Agent !



## 参与调研您将优先获得



AiDD定制版  
《AI+软件研发精选案例》



专属学习顾问  
1对1需求对接

# AiDD会后小调研

AiDD峰会致力于协助企业利用AI技术深化计算机对现实世界的理解，推动研发进入智能化和数字化的新时代。作为峰会的重要共建者，您的真知灼见对我们至关重要。衷心感谢您的参与支持！

# 2025 AI+研发数字峰会

## 拥抱 AI 重塑研发



扫码参与调研

# 科技生态圈峰会 + 深度研习

—1000+ 技术团队的选择



**K+峰会** **敦煌站**  
**K+ 思考周®研习社**  
时间: 2025.08.29-30

**K+峰会** **上海站**  
**K+ 金融专场**  
时间: 2025.09.26-27

**K+峰会** **香港站**  
**K+ 思考周®研习社**  
时间: 2025.11.17-18



K+峰会详情



**AIDD峰会** **上海站**  
**AI+研发数字峰会**  
时间: 2025.05.23-24

**AIDD峰会** **北京站**  
**AI+研发数字峰会**  
时间: 2025.08.08-09

**AIDD峰会** **深圳站**  
**AI+研发数字峰会**  
时间: 2025.11.14-15



AIDD峰会详情



2025 AI+研发数字峰会  
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

