

目录

2025 年 12 月 19 日	2
解决登录问题	3
读读文档	3
2025 年 12 月 18 日	3
2025 年 12 月 16 日	5
目标：录屏	5
问题：花屏	5
确认是偶然现象还是必然现象	5
结论：4 次出现 1 次。	5
重新录屏	5
问题：无有效操作便终止	5
试试设置成中文输入法	5
结论：无效	5
结论：ui-tars-desktop-ubuntu 不稳定	5
目标：看看 AtoGLM-GUI 能不能跑	5
结论：不能。仅限安卓	5
目标：测试 ui-tars-desktop-browser-use-ubuntu	5
结论：未执行有效操作却错误判定成功完成任务	5
目标：录制 ui-tars-desktop-windows	5
问题：win + ctrl + r 只能录制单个窗口，而我想录制整个屏幕。	5
尝试：使用 PPT 录屏	5
结论：成功	5
结论：ui-tars-desktop-windows 不稳定	5
2025 年 12 月 14 日	5
我想跑通 CUA	5
我想跑通 open-interpreter	8
我想试试 Agent S3 能不能跑通	8
实验 ui-tars-desktop (local computer)	13
回顾所有能用的 agent	13
GUI Agent 部署	13
规划器与执行器的解耦	13
解耦的必要性：运行环境的不同	13
最佳部署方式	13
cua	14
项目介绍	14
执行器的使用	14
控制端	14
受控端	14
规划器的使用	14
我想试试 ui-tars-desktop 能否在 ubuntu 上运行。	15
2025 年 12 月 13 日	18
我想弄懂 GUI Agent 的规划器和执行器的分离	18
我想弄懂 cua	18
安装 CUA	18
我想测试速度与精度，通过 OSWorld	18

我想看看 ui-tars-desktop 能不能跑	19
我想试试 agent-tars-cli 能不能跑通	21
我想试试 UFO 能不能跑通? 还试吗?	22
我想试试 Agent-S 能不能跑通	23
2025 年 12 月 12 日	23
今日目标	23
我想读 fara 论文, 并做一个简介的图	23
我想通过 typst 文档了解作图	23
我想绘制 UFO 图	24
我想整理 Surfer 2 PPT	24
我想研究 GLM 的 GUI Agent 能力	24
我想研究机器人进展	24
我想找点事做度过时间	24
我想横向对比不同框架的可用性	24
2025 年 12 月 11 日	25
今日目标	25
整理 ReAct 到 PPT 中	25
整理 UFO 到 PPT 中	25
深入理解 open-interpreter	25
深入理解 OSWorld	25
Typst 如何处理数学公式中下标有多字母的情况?	25
OS-Copilot 论文阅读	25
The Dawn of GUI Agent: A Preliminary Case Study with Claude 3.5 Computer Use 论文阅读	25
我想整理 Cradle PPT	25
我想记录明天要做的事情	25
2025 年 12 月 10 日	26
typst 的 BUG	26
问题描述	26
原因	26
解决方案: 使用代码块	26

2025 年 12 月 19 日

已经在内网跑通结果, 需要录屏。麒麟系统没有自带录屏软件。经查询资料, 得知 ffmpeg 能够录屏:

```
ffmpeg -f x11grab -video_size 1920x1080 -i :0.0 output.mp4
```

现在需要确认麒麟系统能不能用 ffmpeg。能。

成功向领导汇报。

我意识到一个重大改进空间: 从行业调研角度了解这一问题。

我需要一个术语定义: 什么行业?

ui 自动化?

我需要定义评价指标

rpa? (有应用专有 api 等。 web 自动化只是其中一部分。)

gui agent 是 rpa 的升级版?

- chat gpt 的升级版是?

苹果手机是功能机的升级版么?

- 功能机 + 相机 + mp3

RPA (Robotic Process Automation) = 浏览器自动化 (基于规则解析 DOM 的) + 应用程序自动化 (基于应用内部 API 的)

- 技术成熟度: 高。商业化化成熟。

BUA (Browser Use Agent) = 浏览器自动化 (基于大模型理解 DOM 的)

- 技术成熟度: 中。商业化不成熟。

(CUA) Computer Use Agent = 浏览器自动化 (基于视觉的) + 应用程序自动化 (基于视觉的)

- 技术成熟度: 低。

连我对这些概念都很难弄懂, 如何给别人讲清楚?

“流程自动化” Process Automation

解决登录问题

登录本来就不该是自动化的事情。

把代码下载到本地, 问 qwen 怎么使用已经打开的浏览器窗口, 或者使用已经登录的。这是一个很典型的问题。

自动打开浏览器, 手动登录, 登录后自动执行命令。

做了 4 次, 都没有自动启动浏览器。

读读文档

<https://browser-use.com/> Persistent_Profiles Login to any website. Stay authenticated.

完成: 通过指定浏览器路径实现记住登录信息的功能。

2025 年 12 月 18 日

在 GA 网上, 调用千帆模型, 报 400 错。我在互联网上调用千帆看看能否成功。如果互联网上也报相同错误, 说明需要改代码从而支持千帆模型。

千帆 api key: bce-v3/ALTAK-

WEtfPuAdoIz3APYflYveE/5f92d54547ab2be4834c6f6bc9671e4e1f48f4de (<https://console.bce.baidu.com/qianfan/ais/console/onlineTest/LLM>)

base_url = 'https://qianfan.baidubce.com/v2'

model= "ernie-5.0-thinking-preview"

api 输入错误: WARNING [Agent] ⚠️ LLM error (ModelError: Error code: 401 - { 'error' : { 'code' : 'invalid_iam_token' , 'message' : 'invalid_iam_token' , 'type' : 'invalid_request_error' } , 'id' : 'as-ssusauigqj' }) but no fallback_llm configured

地址输入错误报错： base_url = ‘https://qianfan.baidubce.com/v2/chat/completions’
WARNING [Agent]  Result failed 1/4 times: 404 page not found 这说明代码中的地址和 curl 中的地址不能是同一个（内网上再次验证了这一点）

试一下 curl 能否跑通

```
curl -location -request POST 'https://qianfan.baidubce.com/v2/chat/completions'  
-header 'Content-Type: application/json'  
-header 'Authorization: Bearer bce-v3/ALTAK-  
WEtfPuAdoIz3APYflYveE/5f92d54547ab2be4834c6f6bc9671e4e1f48f4de'  
-data-raw '{ "model" : "ernie-5.0-thinking-preview" , "messages" :[{ "role" :  
"user" , "content" : "hi" }], "fps" :2, "web_search" :{ "enable" :true}}' 成功。  
  
{"id":"as-  
yqqxsmi5np","object":"chat.completion","created":1766059499,"model":"ernie-5.0-  
thinking-preview","choices": [{"index":0,"message":  
{"role":"assistant","content":"Hello! How can I help you  
today?", "reasoning_content": "The user just said \"hi.\" This is a very simple  
greeting. My goal is to respond in a way that is friendly, welcoming, and encourages  
further interaction. I should acknowledge their greeting and then prompt them to tell  
me what they need.\n\nHere's my thought process for crafting the response:\n1.  
**Acknowledge the greeting**: A simple \"Hello!\" or \"Hi there!\" is perfect. It's  
direct and friendly.\n2. **Offer assistance**: The core of my purpose is to help. I  
need to ask how I can do that. Phrases like \"How can I help you today?\" or \"What  
can I do for you?\" are standard and effective.\n3. **Combine them**: Putting it  
together creates a complete, polite, and helpful opening. \"Hello! How can I help you  
today?\" is a classic for a reason. It's clear, friendly, and gets straight to the  
point.\n\nI'll go with that. It's a safe and effective way to start the  
conversation."}, "finish_reason": "stop", "flag": 0}], "usage":  
{"prompt_tokens": 1, "completion_tokens": 227, "total_tokens": 228, "completion_tokens_details":  
{"reasoning_tokens": 218}}}
```

使用正确的 api 之后，就能成功访问了。这说明不用改代码专门适配千帆大模型。

2025 年 12 月 16 日

目标：录屏

问题：花屏

确认是偶然现象还是必然现象

结论：4 次出现 1 次。

重新录屏

问题：无有效操作便终止

试试设置成中文输入法

结论：无效

结论：ui-tars-desktop-ubuntu 不稳定

目标：看看 AtoGLM-GUI 能不能跑

结论：不能。仅限安卓

目标：测试 ui-tars-desktop-browser-use-ubuntu

结论：未执行有效操作却错误判定成功完成任务

目标：录制 ui-tars-desktop-windows

问题：win + ctrl + r 只能录制单个窗口，而我想录制整个屏幕。

尝试：使用 PPT 录屏

结论：成功

结论：ui-tars-desktop-windows 不稳定

2025 年 12 月 14 日

我想跑通 CUA

关键是能调用什么模型。

openrouter - 一下就用了 5 毛钱。最低冲 40

试试能不能自定义

发生错误: litellm.BadRequestError: LLM Provider NOT provided. Pass in the LLM provider you are trying to call. You passed model=doubao-1-5-thinking-vision-pro-250428
Pass model as E.g. For 'Huggingface' inference endpoints pass in
'completion(model='huggingface/starcoder',...)` Learn more: <https://docs.litellm.ai/docs/providers>

充钱吧.充了 41.

🚀实验详情:

- 平台: win11

- Agent: CUA
- 大模型: glm-4.5v
- 模型平台: openrouter
- 任务: “在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条”
- 结果: 成功
- 用时: 150s

已连接到本地电脑，Agent 开始思考...

Thinking: 我需要完成的任务是：在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条。

首先分析当前屏幕情况：

- 右侧是一个Google Chrome浏览器窗口，显示的是Google主页
- 我需要在Google搜索框中输入“百度百科 武林外传”来找到相关页面

我的行动计划是：

1. 点击Google搜索框
2. 在搜索框中输入“百度百科 武林外传”
3. 按回车键或点击搜索按钮进行搜索
4. 从搜索结果中找到并点击百度百科链接
5. 进入百度百科后搜索《武林外传》词条

让我开始执行第一步：点击Google搜索框。

AI: 我需要完成在右侧浏览器中进入百度百科并搜索《武林外传》的任务。目前我看到右侧浏览器显示的是Google首页，首先需要点击搜索框。

Thinking: 我需要在右侧浏览器中进入百度百科并搜索电视剧《武林外传》词条。根据当前截图，我可以看到：

1. 右侧是一个Chrome浏览器窗口，显示的是Google首页
2. 搜索框已经处于激活状态（有下拉建议列表显示）
3. 我需要在搜索框中输入“百度百科 武林外传”

我的下一步操作应该是在搜索框中输入相关内容。由于搜索框已经激活，我可以直接开始输入文本。

我应该输入的内容是“百度百科 武林外传”，这样可以直接搜索到百度百科上的《武林外传》词条。

AI: 现在我需要在已经激活的搜索框中输入“百度百科 武林外传”来搜索相关信息。搜索框已经处于活跃状态，可以直接进行输入。

Thinking: 我正在查看当前截图，可以看到右侧浏览器窗口中有一个Google搜索页面。我已经在搜索框中输入了“百度百科 武林外传”，并且现在显示了搜索建议和结果。

根据任务要求，我需要进入百度百科并搜索《武林外传》词条。现在我需要执行这个搜索，然后从搜索结果中选择合适的链接以访问百度百科上的《武林外传》条目。

在当前的搜索结果中，我可以看到几个与《武林外传》相关的选项：

1. “武林外传（2006年尚敬执导的古装武... - baike.baidu.com/item/...”
2. “武林外传（2006年尚敬... - baike.baidu.com/lemma/api/entry?w...”
3. “武林外传（2006年尚敬执导... - baike.baidu.com/item/武林外传?...”

这些看起来都是指向百度百科上《武林外传》条目的链接。我应该点击第一个链接，因为它似乎是最完整的条目链接。

因此，我的下一步操作应该是按下回车键或点击其中一个搜索结果来执行搜索并导航到相关页面。

AI: 我已在搜索框中输入了“百度百科 武林外传”，现在需要执行搜索。我可以看到搜索下拉菜单中显示了几个与《武林外传》相关的百度百科链接。为了完成任务，我应该按下回车键或点击搜索按钮来执行搜索，然后选择最相关的百度百科链接。

Thinking: 让我分析当前的情况：

1. 当前屏幕显示：右侧浏览器已经打开了Google搜索页面，并且在搜索框中已经输入了“百度百科 武林外传”。但是看起来搜索还没有被执行，因为页面仍然显示的是Google首页而不是搜索结果页面。

2. 根据历史操作：

- 第一步：点击了Google搜索框
- 第二步：在搜索框中输入了“百度百科 武林外传”
- 第三步：按下了回车键（从截图看，可能没有成功执行搜索）

3. 从当前截图来看，搜索框中确实有文字“百度百科 武林外传”，但页面还是Google的主页，这说明搜索还没有被执行。

4. 下一步应该是执行搜索。我可以有两种方式：

- 按下回车键
- 点击搜索按钮

由于之前尝试按回车键可能没有成功，或者需要再次确认，我应该再次按回车键来执行搜索。

不过，从截图中看，搜索框中的内容是可见的，而且URL栏显示“google.com/search?q=百度百科+武林外传&oq=百度百科+武林外传&g...”，这表明搜索请求已经被发送，但页面可能还在加载中，或者需要等待搜索结果出现。

考虑到这一点，我认为最好的行动是等待页面加载完成，或者如果页面没有变化，再次按回车键确保搜索被执行。

AI：我注意到搜索框中已经输入了“百度百科 武林外传”，但从当前页面来看，搜索似乎还没有被执行，因为页面仍显示为Google首页而非搜索结果页面。虽然URL栏显示有搜索参数，但页面内容还未更新。我需要执行搜索操作，最直接的方式是按回车键来触发搜索。

Thinking：让我分析当前的屏幕截图和任务情况：

1. 当前屏幕显示的是Google搜索结果页面，搜索关键词是“百度百科 武林外传”。这表明之前的搜索操作已经成功执行。

2. 在搜索结果中，我看到了多个与“武林外传”相关的百度百科条目：

- 第一个结果是“武林外传_百度百科”
- 第二个也是“武林外传_百度百科”
- 第三个同样是“武林外传_百度百科”
- 还有“天真派武林外传”等其他相关条目

3. 我的任务是“在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条”。

根据当前的情况，我已经完成了搜索步骤，现在需要点击第一个最相关的百度百科链接来进入《武林外传》的词条页面。

从搜索结果来看，第一个链接应该是我们要找的电视剧《武林外传》的百科词条。我应该点击这个链接来完成任务。

AI：我现在看到Google搜索结果页面显示了多个关于《武林外传》的百度百科链接。根据任务要求，我需要点击第一个最相关的百度百科链接来进入《武林外传》的词条页面。

Thinking：让我分析当前的屏幕截图和任务进展：

1. 当前状态：我在右侧浏览器中看到了Google搜索结果页面，显示了多个关于《武林外传》的百度百科链接。

2. 任务目标：进入百度百科并搜索电视剧《武林外传》词条。

3. 历史操作：

- 已经在Google搜索框中输入了“百度百科 武林外传”
- 已经执行了搜索，现在看到了搜索结果
- 上一步操作是点击了一个链接（left_click(start_box='[1458,458]')）

4. 当前屏幕分析：

- 我看到搜索结果中有多个百度百科链接，其中第一个链接是“武林外传_百度百科”，描述为“《武林外传》是由北京联盟影业投资有限公司出品，宁财神编剧，尚敬执导，闫妮、姚晨、倪虹洁、沙溢、喻恩泰、姜超、王莎莎等人主演的古装武侠喜剧。”

- 这个链接看起来是最相关的，应该是我们要找的电视剧《武林外传》的百度百科词条。

5. 下一步操作：

- 我需要点击这个最相关的百度百科链接来进入《武林外传》的词条页面。

从历史操作看，上一步已经点击了一个链接，但现在仍然显示搜索结果页面，可能是因为点击后页面还在加载，或者点击的位置不准确。我需要重新点击正确的链接。

根据当前屏幕，我应该点击第一个“武林外传_百度百科”链接，即位于页面上方的那个链接。

AI：我现在看到Google搜索结果页面显示了多个关于《武林外传》的百度百科链接。根据任务要求，我需要点击最相关的第一个百度百科链接来进入《武林外传》的词条页面。从搜索结果来看，第一个链接就是我们要找的电视剧《武林外传》的百科词条。

Thinking：让我分析一下当前的屏幕截图：

我看到右侧浏览器已经成功打开了百度百科的《武林外传》词条页面。这可以从以下几个方面确认：

1. 浏览器标签页显示“武林外传（2006年尚敬执导的...）”，表明我们已经进入了相关页面
2. URL地址栏显示“baike.baidu.com/item/武林外传/5298”，这是百度百科《武林外传》词条的标准URL格式
3. 页面主体内容显示了《武林外传》的相关信息，包括：
 - 标题“武林外传”
 - 副标题“2006年尚敬执导的古装武侠喜剧”
 - 剧情简介和相关信息
 - 演员信息等

这表明我们已经成功完成了任务：在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条。

AI：我可以看到右侧浏览器已经成功加载了百度百科《武林外传》的词条页面。从URL栏可以看到“baike.baidu.com/item/武林外传/5298”，页面内容也显示了《武林外传》的相关信息，包括它是2006年尚敬执导的古装武侠喜剧等内容。任务已经完成，我已成功在右侧浏览器中进入百度百科并搜索到了《武林外传》词条。
150.86535143852234

我想跑通 open-interpreter

```
conda create -n open_interpreter_py310 python=3.10
conda activate open_interpreter_py310
pip install open-interpreter
# export OPENROUTER_API_KEY=sk-or-v1-
e71e35a3fb685d3734744753ab98ae820242a5d047bb24443f37799d1fe107ce
$env:OPENROUTER_API_KEY = "sk-or-v1-
e71e35a3fb685d3734744753ab98ae820242a5d047bb24443f37799d1fe107ce"
interpreter --model openrouter/openai/gpt-4
```

非 os mode 没有视觉能力。

os mode 报错：ImportError: cannot import name ‘StrEnum’ from ‘enum’ 因为 python 需要 311

```
conda create -n open_interpreter_py311 python=3.11
conda activate open_interpreter_py311
pip install open-interpreter
pip install pyautogui
pip install uvicorn fastapi
$env:OPENROUTER_API_KEY = "sk-or-v1-
e71e35a3fb685d3734744753ab98ae820242a5d047bb24443f37799d1fe107ce"
interpreter --os --model openrouter/openai/gpt-4
```

不行。必须只能使用 Anthropic API。这是闭源的。没有意义。

我想试试 Agent S3 能不能跑通

```
conda create -n agent_s3_py311 python=3.11
conda activate agent_s3_py311
```

```
pip install gui-agents
$env:OPENROUTER_API_KEY = "sk-or-v1-
e71e35a3fb685d3734744753ab98ae820242a5d047bb24443f37799d1fe107ce"
$env:OPEN_ROUTER_ENDPOINT_URL="https://openrouter.ai/api/v1"
agent_s --provider open_router --model gpt-5-2025-08-07 --ground_provider open_router
--ground_url https://openrouter.ai/api/v1 --ground_model gpt-5-2025-08-07 --
grounding_width 1920 --grounding_height 1080
```

运行实录

```
agent_s --provider open_router --model gpt-5-2025-08-07 --ground_provider open_router
--ground_url https://openrouter.ai/api/v1 --ground_model gpt-5-2025-08-07 --
grounding_width 1920 --grounding_height 1080
[2025-12-14 13:31:31,511 INFO code_agent/102-MainProcess] CodeAgent initialized with
budget=20
Query: 在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条
```

🔄 Step 1/15: Getting next action from agent...

Response success!

[2025-12-14 13:32:07,088 INFO worker/361-MainProcess] PLAN:

(Previous action verification)

No previous action was performed, so there is nothing to verify.

(Screenshot Analysis)

The screen shows a Windows desktop with Google Chrome open on the right in a new tab page (Chinese UI). The omnibox/search bar is visible with Google logo, and several quick links are shown. The task requires using this already-open browser to go to Baidu Baike and search for the TV series “武林外传”.

(Next Action)

Focus the browser's address bar to type the Baidu Baike URL.

(Grounded Action)

```
python agent.hotkey([ 'ctrl' , 'l' ])
```

EXECUTING CODE: import pyautogui; pyautogui.hotkey('ctrl', 'l')

🔄 Step 2/15: Getting next action from agent...

Response success!

[2025-12-14 13:32:23,772 INFO worker/175-MainProcess] REFLECTION THOUGHTS: Case 2:

The trajectory is going according to plan.

You correctly identified the open browser and initiated focusing the address bar, which aligns with the task. Continue proceeding as planned.

[2025-12-14 13:32:23,772 INFO worker/176-MainProcess] REFLECTION: Case 2: The trajectory is going according to plan.

You correctly identified the open browser and initiated focusing the address bar, which aligns with the task. Continue proceeding as planned.

Response success!

Response success!

RAW GROUNDING MODEL RESPONSE: 920,96

[2025-12-14 13:33:21,728 INFO worker/361-MainProcess] PLAN:

(Previous action verification)

The previous action pressed Ctrl+L to focus the browser's address bar. The Chrome new tab page is visible, and focusing the address bar is the correct next step to navigate to Baidu Baike.

(Screenshot Analysis)

On the right side, Google Chrome is open to a new tab page with the search/omnibox ready. The left side shows a PowerShell window, which is irrelevant. We need to navigate within the browser to Baidu Baike and then search for the TV series entry.

(Next Action)

Type the Baidu Baike URL into the focused address bar and press Enter to navigate there.

(Grounded Action)

```
python agent.type( "The Chrome address bar at the top is focused and ready for input" , text= "baike.baidu.com" , enter=True)
```

Response success!

RAW GROUNDING MODEL RESPONSE: 930,86

```
EXECUTING CODE: import pyautogui;
try:
    import pyperclip
except ImportError:
    import subprocess
    subprocess.run('echo "osworld-public-evaluation" | sudo -S apt-get install -y
xclip xsel', shell=True, check=True)
    subprocess.check_call([subprocess.sys.executable, '-m', 'pip', 'install',
'pyperclip'])
    import pyperclip
```

```
pyautogui.click(1240, 127); pyautogui.write('baike.baidu.com');
pyautogui.press('enter');
```

↳ Step 3/15: Getting next action from agent...

Response success!

[2025-12-14 13:34:17,081 INFO worker/175-MainProcess] REFLECTION THOUGHTS: Case 2:
The trajectory is going according to plan. You correctly focused the Chrome address bar and initiated navigation to baike.baidu.com, which aligns with the task. Continue proceeding as planned.

[2025-12-14 13:34:17,081 INFO worker/176-MainProcess] REFLECTION: Case 2: The trajectory is going according to plan. You correctly focused the Chrome address bar and initiated navigation to baike.baidu.com, which aligns with the task. Continue proceeding as planned.

Response success!

[2025-12-14 13:35:08,773 INFO worker/361-MainProcess] PLAN:

(Previous action verification)

The previous step intended to navigate to baike.baidu.com via the address bar, but the screen still shows the Chrome new tab page, indicating the navigation likely did not occur. The address bar may not have been focused or the URL was not entered successfully.

(Screenshot Analysis)

Chrome remains on the Google new tab page with the Google logo and search/omnibox. No Baidu Baike page is visible yet. We need to ensure the address bar is focused to enter the Baidu Baike URL.

(Next Action)

Refocus the Chrome address bar to guarantee it's ready for typing the URL.

(Grounded Action)

```
python agent.hotkey([ 'ctrl' , 'l' ])
```

```
EXECUTING CODE: import pyautogui; pyautogui.hotkey('ctrl', 'l')
```

🔄 Step 4/15: Getting next action from agent...

Response success!

```
[2025-12-14 13:35:36,351 INFO worker/175-MainProcess] REFLECTION THOUGHTS: Case 2:  
The trajectory is going according to plan. You correctly refocused the Chrome address  
bar to prepare for navigation. Continue proceeding as planned.
```

```
[2025-12-14 13:35:36,352 INFO worker/176-MainProcess] REFLECTION: Case 2: The  
trajectory is going according to plan. You correctly refocused the Chrome address bar  
to prepare for navigation. Continue proceeding as planned.
```

Response success!

Response success!

```
RAW GROUNDING MODEL RESPONSE: 930,100
```

```
[2025-12-14 13:36:31,770 INFO worker/361-MainProcess] PLAN:
```

(Previous action verification)

The previous action used Ctrl+L to focus the Chrome address bar. This is appropriate
and prepares for entering a URL. No navigation has occurred yet.

(Screenshot Analysis)

Chrome is on the Google new tab page on the right. The address bar should be focused
after the last action. The task is to go to Baidu Baike and search for the TV series
“武林外传”.

(Next Action)

Type the Baidu Baike URL into the focused address bar and press Enter to navigate.

(Grounded Action)

```
python agent.type( "The Chrome address bar at the top, which is currently focused  
for input" , text= "https://baike.baidu.com" , enter=True)
```

Response success!

```
RAW GROUNDING MODEL RESPONSE: 920,76
```

```
EXECUTING CODE: import pyautogui;
```

```
try:
```

```
    import pyperclip
```

```
except ImportError:
```

```
    import subprocess
```

```
    subprocess.run('echo "osworld-public-evaluation" | sudo -S apt-get install -y  
xclip xsel', shell=True, check=True)
```

```
    subprocess.check_call([subprocess.sys.executable, '-m', 'pip', 'install',  
'pyperclip'])
```

```
    import pyperclip
```

```
pyautogui.click(1227, 113); pyautogui.write('https://baike.baidu.com');  
pyautogui.press('enter');
```

🔄 Step 5/15: Getting next action from agent...

Response success!

```
[2025-12-14 13:37:40,904 INFO worker/175-MainProcess] REFLECTION THOUGHTS: Case 1:  
The trajectory is not going according to plan.
```

You've repeatedly tried to focus the Chrome address bar and type the Baidu Baike URL, but the page remains on the Google new tab, indicating the input isn't reaching the address bar or navigation isn't being triggered. Continuing the same sequence is causing a loop with no progress. Modify the approach to ensure the address bar is actually focused and that navigation to Baidu Baike is confirmed before proceeding to the search.

[2025-12-14 13:37:40,905 INFO worker/176-MainProcess] REFLECTION: Case 1: The trajectory is not going according to plan.

You've repeatedly tried to focus the Chrome address bar and type the Baidu Baike URL, but the page remains on the Google new tab, indicating the input isn't reaching the address bar or navigation isn't being triggered. Continuing the same sequence is causing a loop with no progress. Modify the approach to ensure the address bar is actually focused and that navigation to Baidu Baike is confirmed before proceeding to the search.

Response success!

Response success!

RAW GROUNDING MODEL RESPONSE: 930,87

[2025-12-14 13:38:36,026 INFO worker/361-MainProcess] PLAN:

(Previous action verification)

The previous attempt intended to type the Baidu Baike URL, but the browser still shows the Google new tab page, indicating the address bar was not engaged or navigation did not occur.

(Screenshot Analysis)

Google Chrome is open on a new tab with the omnibox showing the placeholder “在 Google 中搜索，或输入网址”. We need to ensure the address bar gets focus before typing.

(Next Action)

Click directly inside the Chrome address/omnibox field to ensure it is focused.

(Grounded Action)

python agent.click(“The Chrome address bar at the top center that shows the placeholder text ‘在 Google 中搜索，或输入网址’” , 1, “left”)

Response success!

RAW GROUNDING MODEL RESPONSE: 935,292

EXECUTING CODE: import pyautogui; import pyautogui; pyautogui.click(1247, 433, clicks=1, button='left');

⌚ Step 6/15: Getting next action from agent...

Attempt 1 failed: Expecting value: line 173 column 1 (char 946)

Response success!

[2025-12-14 13:40:23,421 INFO worker/175-MainProcess] REFLECTION THOUGHTS: Case 2:

The trajectory is going according to plan. You corrected the earlier loop by explicitly focusing the Chrome omnibox. Continue proceeding as planned.

[2025-12-14 13:40:23,422 INFO worker/176-MainProcess] REFLECTION: Case 2: The trajectory is going according to plan. You corrected the earlier loop by explicitly focusing the Chrome omnibox. Continue proceeding as planned.

Response success!

♦ Agent-S Workflow Paused ♦

=====

Options:

- Press Ctrl+C again to quit

```
• Press Esc to resume workflow  
=====  
[PAUSED] Waiting for input...  
Unknown command: '  
    ' (ord: 12)  
[PAUSED] Waiting for input...
```

```
● Exiting Agent-S...  
(agent_s3_py311) PS C:\Users\admin>
```

运行了 9 分钟，总也整不对。

实验 ui-tars-desktop (local computer)

🚀 实验详情：

- 平台：win11
- Agent: ui-tars-desktop - local computer
- 大模型: doubaot-1-5-thinking-vision-pro-250428
- 模型平台：火山引擎
- 任务：“在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条”
- 结果：成功
- 用时：88s

(要切换回国内网络环境。要使用单一浏览器)

回顾所有能用的 agent

还是 ui-tars-desktop 最靠谱。没有能比的。

问大模型：除了 ui-tars-desktop, microsoft ufo, open-interpreter, trycua/cua, agent s3, 还有没有其他开源 GUI_Agent / Computer_Use_Agent。无新发现。

GUI Agent 部署

规划器与执行器的解耦

GUI Agent 的两大组件 规划器与执行器

规划器 生成 JSON 格式的抽象动作描述。

执行器 负责动作的具体执行。

解耦的必要性：运行环境的不同

规划器 依赖 AgentScope/LangChain/HuggingFace 等庞大 AI 生态，环境配置复杂。由于终端多样性（arm64 机器/WinXP 系统等），难以确保成功安装和运行规划器。

执行器 依赖 Windows API 或 Android ADB。

跨平台适配成本 规划器逻辑（纯思维）是通用的，而执行器（实际操作）必须针对 Win/Mac/Linux 分别开发。解耦使得一套“大脑”可以控制多种不同系统的“手”。

最佳部署方式

规划器 若有一台服务器专门部署大模型，那么在这台服务器上部署规划器是最优选择。且省去了规划器与大模型的通信成本。

执行器 仅负责采集状态（截图/DOM）上传与接收指令落地。

cua

项目介绍

用途 面向计算机使用代理的开源基础设施。包括 Agent SDK（规划器）和 Computer SDK（执行器）两大组件。

地址 <https://github.com/trycua/cua>

stars 11.5k

最近更新时间 20225 年 12 月 13 日

平台 macOS, Linux, Windows

执行器的使用

任务描述 按下 win 键，并在开始菜单中输入 helloworld

控制端 负责发送指令

受控端 负责执行指令。可为 windows/linux/macOS

控制端

```
# pip install pip install cua-computer
import asyncio
from computer import Computer
async def main():
    computer = Computer(os_type="windows", use_host_computer_server=True)
    try:
        await computer.run()
        await computer.interface.press("win")
        await computer.interface.type_text("Hello!")
    finally:
        await computer.stop()
if __name__ == "__main__":
    asyncio.run(main())
```

受控端

```
pip install cua-computer-server
python -m computer_server
-----
2025-12-13 08:55:30,892 - computer_server.cli - INFO - Starting CUA Computer API
server on 0.0.0.0:8000...
2025-12-13 08:55:30,892 - computer_server.cli - INFO - HTTP mode (no SSL certificates
provided)
INFO:     Started server process [19840]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
INFO:     127.0.0.1:64189 - "POST /cmd HTTP/1.1" 200 OK
INFO:     127.0.0.1:64191 - "POST /cmd HTTP/1.1" 200 OK
INFO:     127.0.0.1:64193 - "POST /cmd HTTP/1.1" 200 OK
```

规划器的使用

```
import asyncio, os
from agent import ComputerAgent
```

```

from computer import Computer
os.environ["CUA_API_KEY"] = "sk_cua-api01_5e...247"
async def main():
    computer = Computer(os_type="windows", use_host_computer_server=True)
    agent = ComputerAgent(model="cua/bytedance/ui-tars-2", tools=[computer],
                           max_trajectory_budget=5.0)
    messages = [{"role": "user", "content": "Take a screenshot and tell me what you see"}]
    await computer.run()
    async for result in agent.run(messages):
        if "output" in result:
            for item in result["output"]:
                if item.get("type") == "message": print(item["content"][0]["text"])
    await computer.stop()
if __name__ == "__main__":
    asyncio.run(main())

```

我想试试 ui-tars-desktop 能否在 ubuntu 上运行。

首先要下载代码并解压: <https://github.com/bytedance/UI-TARS-desktop/archive/refs/tags/v0.3.0.zip>

```

npm install -g pnpm
cd ~/projects/UI-TARS-desktop-0.3.0
export PUPPETEER_SKIP_DOWNLOAD="true"
pnpm install --registry=https://registry.npmmirror.com # 国内加速非常有效
pnpm run dev:ui-tars

```

报错: Error: Electron uninstall 因为安装 electron 过程卡住了, 我按了 ctrl c, 所以没有装好。

当我重新安装时, 卡在安装 electron 很久,

```

packages/ui-tars/electron-ipc/node_modules/electron: Running postinstall script,
failed in 36m 46s
.../electron-ipc/node_modules/electron postinstall$ node install.js
| ReadError: The server aborted pending request
|   at IncomingMessage.<anonymous> (/home/moonlet/Downloads/UI-TARS-desktop-0.3.0/
node_modules/got/dist/source/core/index.js:809:31)
|     at Object.onceWrapper (node:events:632:28)
|     at IncomingMessage.emit (node:events:530:35)
|     at origin.emit (/home/moonlet/Downloads/UI-TARS-desktop-0.3.0/node_modules/
@szmarczak/http-timer/dist/source/index.js:43:20)
|     at IncomingMessage._destroy (node:_http_incoming:221:10)
|     at _destroy (node:internal/streams/destroy:122:10)
|     at IncomingMessage.destroy (node:internal/streams/destroy:84:5)
|     at TLSSocket.socketCloseListener (node:_http_client:478:11)
|     at TLSSocket.emit (node:events:530:35)
|     at node:net:346:12
└ Failed in 36m 46s at /home/moonlet/Downloads/UI-TARS-desktop-0.3.0/packages/ui-
tars/electron-ipc/node_modules/electron
  ELIFECYCLE Command failed with exit code 1.

```

还是因为网络问题导致 electron 下载有问题。解决方案是采用淘宝镜像:

```
ELECTRON_MIRROR="https://npmmirror.com/mirrors/electron/" npm install
```

上述方案有效。

build 时还会碰到一个错误：

```
error [tsc] src/hooks/postMake.ts:52:21 - error TS2345: Argument of type 'Buffer' is not assignable to parameter of type 'BinaryLike'.
```

这是一个常见的 TypeScript 类型定义冲突错误。Node.js 的 Buffer 类型与 TypeScript 标准库中最新的 ArrayBuffer 类型定义不兼容。简单来说，TypeScript 检查器认为你传入的 Buffer 对象缺少了一些标准 ArrayBuffer 应有的属性（如 resizable, transfer 等），导致编译失败。

解决方案(豆包错误 gemini 正确)：

这个问题通常是由于 @types/node 版本过旧导致的。

```
pnpm add -D @types/node@latest -w --registry=https://registry.npmmirror.com
```

启动时又报错：

```
ui-tars-desktop:dev: [243646:1214/192556.602903:FATAL:setuid_sandbox_host.cc(163)]  
The SUID sandbox helper binary was found, but is not configured correctly. Rather than run without sandboxing I'm aborting now. You need to make sure that /home/zhbli/projects/UI-TARS-desktop-0.3.0/node_modules/electron/dist/chrome-sandbox is owned by root and has mode 4755.  
ui-tars-desktop:dev: fatal error: all goroutines are asleep - deadlock!
```

解决方案：

```
# 进入 electron 安装目录  
cd /home/zhbli/projects/UI-TARS-desktop-0.3.0/node_modules/electron/dist/  
  
# 设置所有者为 root  
sudo chown root:root chrome-sandbox  
  
# 设置 SUID 权限 (4755)  
sudo chmod 4755 chrome-sandbox
```

启动时又报错：

```
ui-tars-desktop:dev: > ui-tars-desktop@0.2.4 dev /home/zhbli/projects/UI-TARS-desktop-0.3.0/apps/ui-tars  
ui-tars-desktop:dev: > electron-vite dev  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: vite v6.4.1 building SSR bundle for development...  
✓ 2360 modules transformed.  
Generated an empty chunk: "app_private".  
../../node_modules/file-type/core.js (1419:16): Use of eval in "../../node_modules/file-type/core.js" is strongly discouraged as it poses security risks and may cause issues with minification.  
dist/main/app_private-Dyvby5gX.js 0.01 kB  
ui-tars-desktop:dev: dist/main/index-CtsU6c2f.js 0.68 kB  
ui-tars-desktop:dev: dist/main/BrowserWebSocketTransport-D0cjxDeU.js 0.94 kB  
ui-tars-desktop:dev: dist/main/fileFromPath-1g7i6PBK.js 4.66 kB  
ui-tars-desktop:dev: dist/main/source-map-support-B5wTKdZL.js 19.00 kB  
ui-tars-desktop:dev: dist/main/index-D7bHkL9b.js 20.76 kB  
ui-tars-desktop:dev: dist/main/systemPermissions-mpJTHjsV.js 102.67 kB  
ui-tars-desktop:dev: dist/main/index-CfzASTlo.js 282.32 kB  
ui-tars-desktop:dev: dist/main/bidi-PmQPKeLr.js 603.60 kB  
ui-tars-desktop:dev: dist/main/main.js 6,815.76 kB  
ui-tars-desktop:dev: ✓ built in 13.55s
```

```
ui-tars-desktop:dev:  
ui-tars-desktop:dev: build the electron main process successfully  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: -----  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: vite v6.4.1 building SSR bundle for development...  
✓ 1 modules transformed.  
dist/preload/index.js 2.02 kB  
ui-tars-desktop:dev: ✓ built in 184ms  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: build the electron preload files successfully  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: -----  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: dev server running for the electron renderer process at:  
ui-tars-desktop:dev:  
ui-tars-desktop:dev:   → Local: http://localhost:5173/  
ui-tars-desktop:dev:   → Network: use --host to expose  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: start electron app...  
ui-tars-desktop:dev:  
ui-tars-desktop:dev: App threw an error during load  
ui-tars-desktop:dev: Error: Could not resolve "bufferutil" imported by "ws". Is it  
installed?  
ui-tars-desktop:dev:     at Module.<anonymous> (/home/zhbli/projects/UI-TARS-  
desktop-0.3.0/apps/ui-tars/dist/main/main.js:145656:7)  
ui-tars-desktop:dev:     at Module._compile (node:internal/modules/cjs/  
loader:1484:14)  
ui-tars-desktop:dev:     at Module._extensions..js (node:internal/modules/cjs/  
loader:1564:10)  
ui-tars-desktop:dev:     at Module.load (node:internal/modules/cjs/loader:1295:32)  
ui-tars-desktop:dev:     at Module._load (node:internal/modules/cjs/loader:1111:12)  
ui-tars-desktop:dev:     at c._load (node:electron/js2c/node_init:2:16955)  
ui-tars-desktop:dev:     at cjsLoader (node:internal/modules/esm/translators:350:17)  
ui-tars-desktop:dev:     at ModuleWrap.<anonymous> (node:internal/modules/esm/  
translators:286:7)  
ui-tars-desktop:dev:     at ModuleJob.run (node:internal/modules/esm/  
module_job:234:25)  
ui-tars-desktop:dev:     at async ModuleLoader.import (node:internal/modules/esm/  
loader:473:24)  
ui-tars-desktop:dev: A JavaScript error occurred in the main process  
ui-tars-desktop:dev: Uncaught Exception:  
ui-tars-desktop:dev: Error: Could not resolve "bufferutil" imported by "ws". Is it  
installed?  
ui-tars-desktop:dev:     at Module.<anonymous> (/home/zhbli/projects/UI-TARS-  
desktop-0.3.0/apps/ui-tars/dist/main/main.js:145656:7)  
ui-tars-desktop:dev:     at Module._compile (node:internal/modules/cjs/  
loader:1484:14)  
ui-tars-desktop:dev:     at Module._extensions..js (node:internal/modules/cjs/  
loader:1564:10)  
ui-tars-desktop:dev:     at Module.load (node:internal/modules/cjs/loader:1295:32)  
ui-tars-desktop:dev:     at Module._load (node:internal/modules/cjs/loader:1111:12)  
ui-tars-desktop:dev:     at c._load (node:electron/js2c/node_init:2:16955)  
ui-tars-desktop:dev:     at cjsLoader (node:internal/modules/esm/translators:350:17)  
ui-tars-desktop:dev:     at ModuleWrap.<anonymous> (node:internal/modules/esm/  
translators:286:7)
```

```
ui-tars-desktop:dev:      at ModuleJob.run (node:internal/modules/esm/
module_job:234:25)
ui-tars-desktop:dev:      at async ModuleLoader.import (node:internal/modules/esm/
loader:473:24)
ui-tars-desktop:dev: [248055:1214/193526.291558:ERROR:viz_main_impl.cc(185)] Exiting
GPU process due to errors during initialization
```

解决方案：

```
pnpm install bufferutil utf-8-validate --save --registry=https://registry.npmmirror.
com
```

运行时控制台报错：

```
ERROR:x11_software_bitmap_presenter.cc(150)] XGetWindowAttributes failed for window
25165892
```

解决方案：

```
cd /home/zhbli/projects/UI-TARS-desktop-0.3.0
LIBGL_ALWAYS_SOFTWARE=1 pnpm run dev:ui-tars
```

成功打开界面。但总是弹出“屏幕共享”界面

解决方案：采用 xorg 桌面。

虽然很不愿意承认，但是结果错误。根本无法按照指令打开网页。

2025 年 12 月 13 日

我想弄懂 GUI Agent 的规划器和执行器的分离

开一个 PPT 并弄懂：完成。

🚀✨后来我发现，不论平台不平台，分离不分离，找到一个能跑的，就已经挺不错了。

我想弄懂 cua

- 研究 github
 - ▶ 已完成。
- 跑通 cua 并用自然语言执行任务🚀✨
 - ▶ 已完成。
 - ▶ 已经整理到 PPT 中。

安装 CUA

```
conda create -n cua python=3.12
conda activate cua
pip install cua-agent[all]
pip install cua-computer-server
```

执行：

```
cd C:\Users\admin\Documents\GitHub\cua_example
conda activate cua
python agent_win.py
```

CUA 模型平台没钱了。尴尬。

我想测试速度与精度，通过 OSWorld

- 动机：真正用的时候，肯定重点关注速度与精度。我的心里要有数。

- cua 应该提供了。看看怎么用。
- HUD 是一个测评工具。其付费逻辑是怎样的？

我想看看 ui-tars-desktop 能不能跑

分析：我在做的事情本质上就是运行软件。这和运行 word 没什么区别。但是 word 直接双击图标就行了。但是我要运行的软件没有图标。因为我还没有安装。所以要安装软件。但是 word/wps 的安装很容易。百度一搜，下载 exe 双击就能安装。但是我要安装的软件没有那么容易安装。比如我正在安装 ui-tars-desktop，

```
npm install -g pnpm
pnpm install
```

就遇到了如下问题：

```
PS C:\Users\admin\Downloads\UI-TARS-desktop-0.3.0\UI-TARS-desktop-0.3.0> pnpm install
pnpm : 无法将“pnpm”项识别为 cmdlet、函数、脚本文件或可运行程序的名称。请检查名称的拼写，如果包括路径，请确保路径正确
，然后再试一次。
所在位置 行:1 字符: 1
+ pnpm install
+ ~~~~
+ CategoryInfo          : ObjectNotFound: (pnpm:String) [],
CommandNotFoundException
+ FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\admin\Downloads\UI-TARS-desktop-0.3.0\UI-TARS-desktop-0.3.0> npm install
-g pnpm

added 1 package in 11s

1 package is looking for funding
  run `npm fund` for details
PS C:\Users\admin\Downloads\UI-TARS-desktop-0.3.0\UI-TARS-desktop-0.3.0> pnpm install
Scope: all 32 workspace projects
Lockfile is up to date, resolution step is skipped
Packages: +2616
+++++
+++++
Downloading turbo-windows-64@2.4.4: 15.24 MB/15.24 MB, done
Downloading @dqbd/tiktoken@1.0.20: 10.62 MB/10.62 MB, done
Downloading electron-winstaller@5.4.0: 9.14 MB/9.14 MB, done
Downloading @img/sharp-win32-x64@0.33.3: 8.40 MB/8.40 MB, done
Packages are hard linked from the content-addressable store to the virtual store.
Content-addressable store is at: C:\Users\admin\AppData\Local\pnpm\store\v3
Virtual store is at:           node_modules/.pnpm
Downloading gifwrap@0.10.1: 5.89 MB/5.89 MB, done
Downloading js-tiktoken@1.0.19: 10.24 MB/10.24 MB, done
Downloading typescript@4.9.4: 11.62 MB/11.62 MB, done
Downloading @rspack/binding-win32-x64-msvc@1.4.0-beta.0: 21.43 MB/21.43 MB, done
Downloading react-icons@5.4.0: 22.22 MB/22.22 MB, done
Downloading @rspack/binding-win32-x64-msvc@1.2.8: 18.47 MB/18.47 MB, done
Progress: resolved 0, reused 0, downloaded 1987, added 2616, done
node_modules/msw: Running postinstall script, done in 88ms
node_modules/@parcel/watcher: Running install script, done in 103ms
node_modules/tsx/node_modules/esbuild: Running postinstall script, done in 959ms
node_modules/esbuild: Running postinstall script, done in 361ms
```

```
node_modules/bufferutil: Running install script, done in 1.6s
node_modules/utf-8-validate: Running install script, done in 1.2s
node_modules/sharp: Running install script, done in 83ms
node_modules/electron-winstaller: Running install script, done in 64ms
node_modules/electron: Running postinstall script...
node_modules/core-js: Running postinstall script, done in 64ms
node_modules/puppeteer: Running postinstall script, failed in 482ms
node_modules/puppeteer postinstall$ node install.mjs
| **INFO** Skipping Firefox download as instructed.
| Error: ERROR: Failed to set up chrome-headless-shell v134.0.6998.35! Set
"PUPPETEER_SKIP_DOWNLOAD" env variable to ...
|   at downloadBrowser (file:///C:/Users/admin/Downloads/UI-TARS-desktop-0.3.0/UI-
TARS-desktop-0.3.0/node_modules/p...
|     at process.processTicksAndRejections (node:internal/process/task_queues:105:5)
|     at async Promise.all (index 1)
|     at async downloadBrowsers (file:///C:/Users/admin/Downloads/UI-TARS-
desktop-0.3.0/UI-TARS-desktop-0.3.0/node_mo...
|   [cause]: Error: read ECONNRESET
|     at TLSWrap.onStreamRead (node:internal/stream_base_commons:216:20) {
|       errno: -4077,
|       code: 'ECONNRESET',
|       syscall: 'read'
|     }
|   }
|   Failed in 482ms at C:\Users\admin\Downloads\UI-TARS-desktop-0.3.0\UI-TARS-
desktop-0.3.0\node_modules\puppeteer
node_modules/secp256k1: Running install script, done in 172ms
ELIFECYCLE Command failed with exit code 1.
node_modules/@computer-use/mac-screen-capture-permissions: Running install script,
done in 66ms
node_modules/electron-vite/node_modules/esbuild: Running postinstall script...
packages/common/configs/node_modules/core-js: Running postinstall script, done in
63ms
packages/ui-tars/electron-ipc/node_modules/electron: Running postinstall script...
PS C:\Users\admin\Downloads\UI-TARS-desktop-0.3.0\UI-TARS-desktop-0.3.0>
```

Puppeteer 无法连接到 Google 服务器下载配套的 Chrome 浏览器。

尝试：跳过 Puppeteer 的浏览器下载。通常情况下，UI-TARS 这种桌面应用运行时会使用你系统里已经安装好的 Chrome 或 Edge，缺省这个下载通常不会影响核心功能。

```
$env:PUPPETEER_SKIP_DOWNLOAD="true"
pnpm install
```

🚀✨ 安装成功。启动：

```
cd C:\Users\admin\Downloads\UI-TARS-desktop-0.3.0\UI-TARS-desktop-0.3.0
pnpm run dev:ui-tars
```

需要填入三个信息：

```
VLM Base URL
VLM API Key
VLM Model Name
```

VLM Base URL = <https://ark.cn-beijing.volces.com/api/v3> (参见 <https://www.volcengine.com/experience/ark?mode=chat&modelId=doubao-seed-1-6-251015&>

csid=excs-202512131958-%5B2rgkgxRgrXLM3ScMdE8SY%5D - API 接入 - 快速接入测试)

VLM API Key = 72a2c8ba-4d88-47fd-9caa-529f54482133 (与 agent-tars-cli 相同)

VLM Model Name = doubao-1-5-thinking-vision-pro-250428 (与 agent-tars-cli 相同)



实验详情:

- 平台: win11
 - Agent: ui-tars-desktop - local browser
 - 大模型: doubao-1-5-thinking-vision-pro-250428
 - 模型平台: 火山引擎
 - 任务: “在浏览器中, 打开百度百科, 搜索词条《武林外传》”
 - 结果: 成功
 - 用时: 29s

我想试试 agent-tars-cli 能不能跑通

在 windows 上运行: npx @agent-tars/cli@latest

一直卡着：

试试在 4090 上运行。但是远程连接需要密码:abc12345。安装成功。

试试在 windows 上卡着该如何解决。

```
npm config set registry https://registry.npmmirror.com  
npx @agent-tars/cli@latest
```

成功。

开始在浏览器中运行。报错：

Error in agent execution: Error: The OPENAI_API_KEY environment variable is missing or empty; either provide it, or instantiate the OpenAI client with an apiKey option, like new OpenAI({ apiKey: 'My API Key' }).

添加模型方法: <https://agent-tars.com/guide/get-started/quick-start>

```
agent-tars \
--provider volcengine \
```

```
--model doubao-1-5-thinking-vision-pro-250428 \
--apiKey {apiKey}
```

获取火山引擎 API Key <https://console.volcengine.com/ark/region:ark+cn-beijing/apiKey?apikey=%7B%7D>

72a2c8ba-4d88-47fd-9caa-529f54482133

完整启动命令：

```
agent-tars --provider volcengine --model doubao-1-5-thinking-vision-pro-250428 --
apiKey 72a2c8ba-4d88-47fd-9caa-529f54482133
```

报错：

```
PS C:\Users\admin\Documents\GitHub\my_tyst> agent-tars --provider volcengine --model
doubao-1-5-thinking-vision-pro-250428 --apiKey 72a2c8ba-4d88-47fd-9caa-529f54482133
agent-tars : 无法将“agent-tars”项识别为 cmdlet、函数、脚本文件或可运行程序的名称。请检查名称的拼写，如果包括路径，请确保路径正确，然后再试一次。
```

所在位置 行:1 字符: 1

```
+ agent-tars --provider volcengine --model doubao-1-5-thinking-vision-p ...
+ ~~~~~
+ CategoryInfo          : ObjectNotFound: (agent-tars:String) [],
CommandNotFoundException
+ FullyQualifiedErrorId : CommandNotFoundException
```

新的完整启动命令：

```
npx @agent-tars/cli@latest --provider volcengine --model doubao-1-5-thinking-vision-
pro-250428 --apiKey 72a2c8ba-4d88-47fd-9caa-529f54482133
```

报错：

```
Error in agent execution: Error: 404 Your account %!s(int64=2101333747) has not
activated the model doubao-1-5-thinking-vision-pro-250428. Please activate the model
service in the Ark Console. Request id:
021765601742365c2dd50d19df53a1cf97de1fb7620a49af263db
```

原因：模型服务未开通 / 授权。

在 <https://console.volcengine.com/ark/region:ark+cn-beijing/experience/chat?modelId=doubao-1-5-thinking-vision-pro-250428&tab=Chat&csid=excs-202512131257-%5BRZumsSqNvcuRZCK33-y5P%5D> 中成功开通模型。

BUG：开通成功后，偶尔还会报刚才的错误。这时重发一遍信息就行了。

成功运行 agent-tars-cli

我想试试 UFO 能不能跑通？还试吗？

当你想买车的时候，你就只去逛一家么？一定要货比三家。

先试 UFO2 再试 UFO3. 因为 UFO3 依赖 UFO2

在 win11 笔记本上安装 ufo2

```
conda create -n ufo python=3.10
conda activate ufo
git clone https://github.com/microsoft/UFO.git
cd UFO
pip install -r requirements.txt
```

配置千问模型的方法: <https://microsoft.github.io/UFO/configuration/models/qwen/>
dashscope api: sk-efe1c9004f7e4de0a8ade26120301c6d

运行: python -m ufo -task test -r “在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条”

🚀 实验详情:

- 平台: win11
- Agent: Microsoft UFO2
- 大模型: qwen-vl-max
- 模型平台: 阿里云百炼
- 任务: “在右侧已打开的浏览器中，进入百度百科并搜索电视剧《武林外传》词条”
- 结果: 成功
- 用时: 240s

我想试试 Agent-S 能不能跑通

必须用 UI-TARS-1.5-7B

- 看看豆包有没有
- 看看 HF 怎么用的

2025 年 12 月 12 日

今日目标

我今天的目标是: 步数 14000 步, 20: 30 前往宿舍走, 不做梦。

我今天的目标是: 步数 14000 步 (2550 步), 20: 30 前往宿舍走, 不做梦。

我今天的目标是: 步数 14000 步 (4182 步), 20: 30 前往宿舍走, 不做梦。

我今天的目标是: 步数 14000 步 (14136 步), 20: 30 前往宿舍走, 不做梦。

我想读 fara 论文，并做一个简介的图

借助豆包阅读 进行中

我想通过 typst 文档了解作图

找到文档 <https://typst-doc-cn.github.io/tutorial/>。无收获。

学习最基本绘图 完成。

居中文字

我想绘制 UFO 图

弄懂 UFO 逻辑 很复杂，无法在一张图中展现。

我想整理 Surfer 2 PPT

读论文 已完成

我想研究 GLM 的 GUI Agent 能力

理解代码 只能输出动作对应的 json。

论文提到 GUI Agent 得分了吗？ 看论文 2025 年 8 月 15 日。在 OSWorld 上，GLM-4.5V 106B (A12B) thinking 得 35.8 分

qwen3vl 在 GUI Agent 得分了吗？ 看论文 2025 年 11 月 27 日。在 OSWorld 上，Qwen3-VL 235B-A22B thinking 得 38.1 分。最好的是 44.4 分 (Claude Opus 4.1 non-thinking)

我想研究机器人进展

首先定位研究方向 问 gemini。这是 VLA 任务。

读 VLA 论文 RT-2

直觉 1 “看一眼 -> 闭眼走几步 -> 再看一眼” 的模式

结论 我失去了兴趣。没有外在的压力，我根本不想做这些费脑力的事情。

我想找点事做度过时间

做澳大利亚 PPT 已经终止。因为有新事情了。

我想横向对比不同框架的可用性

让 gemini 做 进行中

我想知道 UI-Tars 到底在什么平台上运行 已完成

我想知道 GLM 到底支持什么平台 只支持手机

我想试试 cua 究竟能不能跑通 进行中

对于 win7，没有任何 Agent 能在上面跑 UFO 支持吗？

远程是怎么回事？客户端总得安装操作电脑的工具吧？ 进行中

2025 年 12 月 11 日

今日目标

我今天的目标是：步数 13000 步，20: 30 前往宿舍走，不做梦。我今天的目标是：步数 13000 步 (2315)，20: 30 前往宿舍走，不做梦。我今天的目标是：步数 13000 步 (3640)，20: 30 前往宿舍走，不做梦。我今天的目标是：步数 13000 步 (5224)，20: 30 前往宿舍走，不做梦。我今天的目标是：步数 13000 步 (7437)，20: 30 前往宿舍走，不做梦。我今天的目标是：步数 13000 步 (14789)，20: 30 前往宿舍走，不做梦。

整理 ReAct 到 PPT 中

借助豆包阅读整理 已完成

整理 UFO 到 PPT 中

理解论文 已完成

深入理解 open-interpreter

让 gemmini 找找有没有相关论文 没有。

如何感知中间过程是否正确执行？ 问 gemini。命令行能输出中间结果。也有截图功能。

从代码层面理解 借助豆包。但是我还没有明确具体能做什么任务。所以我需要调研 benchmark。已完成。

继续从代码层面理解 已经初步弄清楚。不懂的：能接受截图作为输入么？

能接受截图作为输入么？ 新的 os 模式支持。所有现有模式不支持。

深入理解 OSWorld

借助豆包理解论文 已完成

在处理 OSworld 任务时，是靠键鼠么？ 是

Typst 如何处理数学公式中下标有多字母的情况？

Latex 语法不适用。正确用法为 \$x_{(abc)}\$

OS-Copilot 论文阅读

借助豆包阅读 完成

The Dawn of GUI Agent: A Preliminary Case Study with Claude 3.5

Computer Use 论文阅读

借助豆包整理 PPT 完成

我想整理 Cradle PPT

借助豆包整理 完成

我想记录明天要做的事情

The Unreasonable Effectiveness of Scaling Agents for Computer Use 引用数=2

[1] Surfer 2: 未发布且闭源 2510.19949

[2] UI-TARS-2: 2509.02544。模型未发布

UI-TARS-2 Technical Report: Advancing GUI Agent with Multi-Turn Reinforcement Learning 引用数=33

[1] UltraCUA: 计划开源

[2] fara: 已开源(比 UI-TARS-1.5 性能好)

<https://github.com/trycua/cua>

<https://github.com/topics/computer-use>

什么是最好的查新方式?

看某篇论文的引用? 太多。不开源也没有用

上论文网站查关键字? 太多。不开源也没有用

问大模型? 不全不准不新。

google 搜索关键字+github: 找到的都是开源的。最有效。

open source GUI Agent / computer use models after 2025.07. give me the model link
(gemini 找出了 fara)

2025 年 12 月 10 日

typst 的 BUG

问题描述

在使用如下 PPT 主题的前提下,

```
#import "@preview/touying:0.5.2": *
#import themes.simple: *

#show: simple-theme.with(aspect-ratio: "16-9")
```

一个 == 段落使用多次 #set text 命令, 则后续 == 段落显示不出来。

原因

在 Typst 中, #set 规则如果不加限制, 会作用于当前上下文的剩余部分。

Touying 的工作原理是捕获 == 下面的所有内容。当你在这个内容流中多次使用全局的 #set text(...) 时, 由于没有闭合的作用域, 这些样式设置可能会干扰 Touying 内部用于分割幻灯片 (Slide) 的状态机或内容分块逻辑, 导致后续的标题无法被正确识别为新的幻灯片起点。

解决方案: 使用代码块

将需要变更样式的部分用 [...] 包裹起来, 把 #set 放在里面。这样样式只会影响这个块, 不会“泄漏”到外部, 也不会破坏幻灯片的分割逻辑。

```
// 使用代码块限制作用域
#[
  #set text(size: 12pt)
  这里是具体内容
]
```