

# GUI Agent 测评报告

## 测评结论与推荐

	Windows 11	Ubuntu 24.04
Web 任务	browser-use UI-TARS	browser-use
OS 任务	CUA UFO2	✗ 无可用推荐 (全失败)

**Web 任务** 仅浏览器/DOM 操作

**OS 任务** 涉及桌面/全局视觉)

## 测评详情

# 实验介绍

**实验目标** 在 Windows 11/Ubuntu24.04 平台上测试各开源 GUI Agent 的可用性、速度与精度。

**测试任务** “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”

**参测 Agent** CUA, Open Interpreter, browser use, Agent S3, UI-TARS Desktop, Microsoft UFO2

# CUA on Windows

项目地址	<a href="https://github.com/trycua/cua">https://github.com/trycua/cua</a>		
stars 数	11.5k	更新日期	2025 年 12 月 14 日
模型	glm-4.5v	模型平台	openrouter
操作系统	Win11	任务类型	基于视觉的 Computer Use
测试任务	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	耗时	150s

# CUA on Windows 运行方式

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

# CUA on Ubuntu

项目地址	<a href="https://github.com/trycua/cua">https://github.com/trycua/cua</a>		
stars 数	11.5k	更新日期	2025 年 12 月 14 日
模型	glm-4.5v	模型平台	openrouter
操作系统	Ubuntu24.04	任务类型	基于视觉的 Computer Use
测试任务	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	失败，耗时 250s。错误地将“武林外传”输入成“武林林传”。错误地判断为任务已成功完成。		

# UI-TARS-desktop on Windows (local computer)

项目地址	<a href="https://github.com/bytedance/UI-TARS-desktop">https://github.com/bytedance/UI-TARS-desktop</a>		
stars 数	19.9k	更新日期	2025 年 12 月 11 日
模型平台	火山引擎	模型	doubao-1-5-thinking-vision-pro-250428
操作系统	Win11	任务类型	基于视觉的 Computer Use
测试任务	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 1	成功	耗时	88s
结果 2	失败。未正确输入回车键进入网页，而 Agent 误以为是网络卡顿，一直处于等待状态。		



# UI-TARS-desktop on Windows (local browser)

项目地址	<a href="https://github.com/bytedance/UI-TARS-desktop">https://github.com/bytedance/UI-TARS-desktop</a>		
stars 数	19.9k	更新日期	2025 年 12 月 11 日（字节）
模型平台	火山引擎	模型	doubao-1-5-thinking-vision-pro-250428
操作系统	Win11	任务类型	基于 DOM 的 Browser Use
测试任务	在浏览器中，打开百度百科，搜索词条《武林外传》		
结果	成功	耗时	29s 至 48s

# UI-TARS-desktop on Windows 运行方式

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

# UI-TARS-desktop on Ubuntu (local computer)

项目地址	<a href="https://github.com/bytedance/UI-TARS-desktop">https://github.com/bytedance/UI-TARS-desktop</a>		
stars 数	19.9k	更新日期	2025 年 12 月 11 日
模型平台	火山引擎	模型	doubao-1-5-thinking-vision-pro-250428
操作系统	Ubuntu24.04	任务类型	基于视觉的 Computer Use
测试任务	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	失败	表现	未执行有效操作却错误判定成功完成任务

# UI-TARS-desktop on Ubuntu (local browser)

项目地址	<a href="https://github.com/bytedance/UI-TARS-desktop">https://github.com/bytedance/UI-TARS-desktop</a>		
stars 数	19.9k	更新日期	2025 年 12 月 11 日
模型平台	火山引擎	模型	doubao-1-5-thinking-vision-pro-250428
操作系统	Ubuntu24.04	任务类型	基于 DOM 的 Browser
测试任务	在浏览器中，打开百度百科，搜索词条《武林外传》		
结果	失败	表现	未执行有效操作却错误判定成功完成任务

# UI-TARS-desktop on Ubuntu 运行方式

```
cd ~/projects/UI-TARS-desktop-0.3.0
```

```
LIBGL_ALWAYS_SOFTWARE=1 pnpm run dev:ui-tars
```

# UFO2 on Windows

项目地址	<a href="https://github.com/microsoft/UFO">https://github.com/microsoft/UFO</a>		
stars 数	7.8k	更新日期	2025 年 12 月 17 日（微软）
模型平台	阿里云百炼	模型	qwen-vl-max
操作系统	Win11	任务类型	基于视觉的 Computer Use
测试任务	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	耗时	240s

# browser use on Ubuntu

项目地址	https://github.com/browser-use/browser-use		
stars 数	73.7k	更新日期	2015 年 12 月 15 日
模型平台	ModelScope	模型	Qwen/Qwen3-VL-235B-A22B-Instruct
操作系统	Ubuntu24.04	任务类型	基于 DOM 的 Browser Use
测试任务	介绍百度百科中搜索“武林外传”词条		
结果	成功	耗时	109s

# Agent S3 on Windows

项目地址	<a href="https://github.com/similar-ai/Agent-S">https://github.com/similar-ai/Agent-S</a>		
stars 数	8.8k	更新日期	2025 年 12 月 16 日
模型	gpt-5-2025-08-07	模型平台	openrouter
操作系统	win11	任务类型	基于视觉的 Computer Use
测试任务	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	失败，运行 9 分钟后中止。陷入死循环。Agent 反复尝试聚焦地址栏 (Ctrl+L) 并输入 URL，但在验证步骤中无法确认操作成功，导致任务无法推进。		



# Agent S3 on Windows 运行方式

```
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
```

# 放弃的 Agents

**Microsoft UF02 on Ubuntu** 仅限 Windows 平台运行

**OpenAI Operator/Claude Computer Use** 闭源

**Open Interpreter** OS Mode 强依赖 Anthropic API (闭源), 且非 OS Mode 缺乏视觉能力

**<https://github.com/suyiiyii/AutoGLM-GUI>** 仅限 Android 平台运行