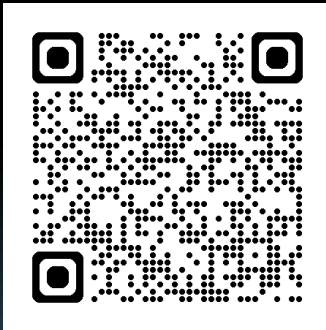


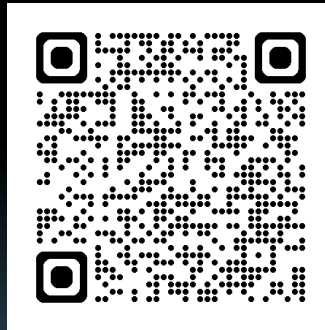


Computer 2.0: How AI is Redefining the Desktop Experience

[LinkedIn](#)



[GitHub](#)



Email: cameron@cameronrohn.com



Introduction

Background

- Director, Managed Cloud and Infrastructure (2012 – 2024)
- vCIO and Technical Account Management
- Financial Institutions, NFP, Manufacturing

AI/ML Journey so far...

Privacy→NVIDIA ChatRTX
→Python→Langchain→Data
Pipelines→LangGraph→Graph
DBs

Current Capabilities

Which commands can you run? Which can you not run?

I have access to 3 main functions to interact with the system:

`computer`

Control mouse movements and clicks (left, right, middle, double-click, drag)- Type text and press keyboard keys/combinations- Take screenshots- Get cursor position This is primarily for GUI interactions.

`bash`

Run bash shell commands, Install packages via apt and pip, Access a mirror of common Linux packages, State persists between commands, Run background processes, inspect files and run typical Linux commands.

`str_replace_editor`

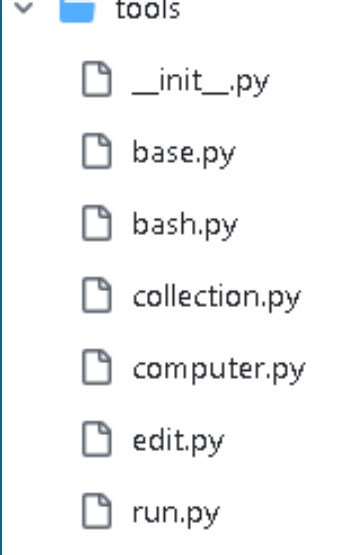
View files and directories (`view` command)- Create new files (`create` command)- Replace text in files (`str_replace` command)- Insert text at specific lines (`insert` command)- Undo edits (`undo_edit` command)

```
import anthropic

client = anthropic.Anthropic()

response = client.beta.messages.create(
    model="claude-3-5-sonnet-20241022",
    max_tokens=1024,
    tools=[
        {
            "type": "computer_20241022",
            "name": "computer",
            "display_width_px": 1024,
            "display_height_px": 768,
            "display_number": 1,
        },
        {
            "type": "text_editor_20241022",
            "name": "str_replace_editor"
        },
        {
            "type": "bash_20241022",
            "name": "bash"
        }
    ],
    messages=[{"role": "user", "content": "Save a picture of a cat to my desktop."}],
    betas=["computer-use-2024-10-22"],
)

print(response)
```



Immediate Applications

Basic Automation

- Form filling and data entry
- Document formatting
- File organization
- Basic software installation

User Assistance

- GUI navigation help
- Software tutorial guidance
- Error troubleshooting
- Settings optimization

Near-Term and Future Evolution

Intermediate Applications

- Complex workflow automation
- Multi-application coordination- Visual data analysis and reporting
- Real-time software training
- Automated testing and QA

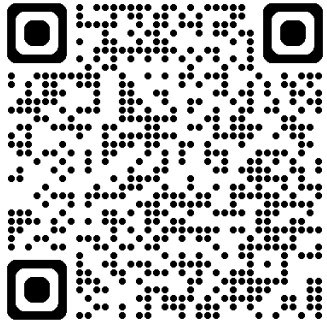
Advanced Integration

- Contextual desktop organization
- Predictive task completion
- Cross-platform workflow optimization
- Intelligent software configuration

Demo - Windows Adaptation



Cam10001110101



computer-use-windows

Public

Integrating Windows + Python + Anthropic's Computer Use API for the development of agentic desktop applications.

Python

☆ 6

🔗 3

<https://github.com/Cam10001110101/computer-use-windows>

10-0011-10-10-1

Additional Information

Limitations

- I cannot directly access **external resources or systems** except through the provided functions
- For the computer GUI function, I don't have direct terminal or applications menu access - I must use **desktop icons**
- Screen resolution is **fixed**
- When using the web browser, I should **ignore setup wizards** and go straight to using the address bar
- For large command outputs, I need to redirect to files and use filtering
- The editor tool will truncate very long outputs

```
SYSTEM_PROMPT = f'""<SYSTEM_CAPABILITY>
```

- You are utilising an **Ubuntu virtual machine** using {platform.machine()} architecture with internet access.
- You can feel free to **install Ubuntu applications** with your bash tool. Use curl instead of wget.
- To open firefox, please just click on the firefox icon. Note, firefox-esr is what is installed on your system.
- Using bash tool you can start GUI applications, but you need to set export DISPLAY=:1 and use a **subshell**. For example "(DISPLAY=:1 xterm &)". GUI apps run with bash tool will appear within your desktop environment, but they may take some time to appear. Take a screenshot to confirm it did.
- When using your bash tool with commands that are expected to output very large quantities of text, redirect into a tmp file and use str_replace_editor or `grep -n -B <lines before> -A <lines after> <query> <filename>` to confirm output.
- When viewing a page it can be helpful to zoom out so that you can see everything on the page. Either that, or make sure you scroll down to see everything before deciding something isn't available.
- When using your computer function calls, they take a while to run and send back to you. Where possible/feasible, try to chain multiple of these calls all into one function calls request.
- The current date is {datetime.today().strftime('%A, %B %d, %Y')}.

```
</SYSTEM_CAPABILITY>
```

* When using Firefox, if a startup wizard appears, IGNORE IT. Do not even click "skip this step". Instead, click on the address bar where it says "Search or enter address", and enter the appropriate search term or URL there. * If the item you are looking at is a **pdf**, if after taking a single screenshot of the pdf it seems that you want to read the entire document instead of trying to continue to read the pdf from your screenshots + navigation, determine the URL, use curl to download the pdf, install and use pdftotext to convert it to a text file, and then read that text file directly with your StrReplaceEditTool. ""

Notable Related Projects



<https://arxiv.org/pdf/2408.00203>



Texts:

'Text Box ID 0: 4:47',
'Text Box ID 1: Freeform',
'Text Box ID 2: Twitter',
'Text Box ID 3: Journal',
'Text Box ID 4: Ticketmaster',

Icons:

'Icon Box ID 24: a delete or remove function.',
'Icon Box ID 25: a note-taking application.',
'Icon Box ID 26: a weather or weather-related application.',
'Icon Box ID 27: a phone call or messaging application.',
'Icon Box ID 28: a tomato with a long shadow.'

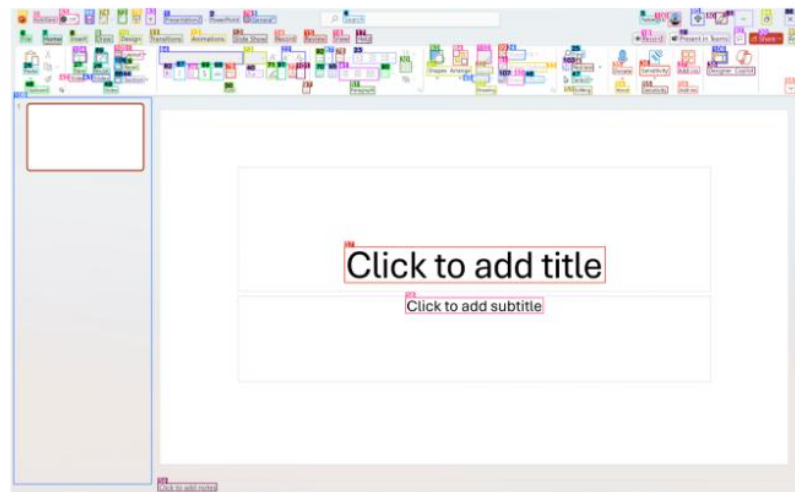


Texts:

'Text Box ID 0: 14:50',
'Text Box ID 1: 6z',
'Text Box ID 2: Settings',
'Text Box ID 3: WLAN',
'Text Box ID 4: Edit',

Icons:

'Icon Box ID 17: a toggle switch in the "on" position.',
'Icon Box ID 18: a toggle switch in the off position.',
'Icon Box ID 19: an edit function in a text editor.',
'Icon Box ID 20: an information or help function.',
'Icon Box ID 21: an information or help function.'



Texts:

'Text Box ID 0: AutoSave',
'Text Box ID 1: Presentation2',
'Text Box ID 2: PowerPoint',
'Text Box ID 3: General',
'Text Box ID 4: Search',

Icons:

'Icon Box ID 62: a chat or messaging feature.',
'Icon Box ID 63: a toggle switch in the "off" position.',
'Icon Box ID 64: a loading or progress bar.',
'Icon Box ID 65: a menu or list of options.',
'Icon Box ID 66: a floppy disk drive.'