

1. Introduction

- Details: Open Chrome with Guest Profile (not Guest Mode) and login to Google Drive with my account provided. Then find 'test case 5' file in folder Test case 5 through menu My Drive (ignore menu Suggestion files) and then download it. After that, open a new tab and login to Gmail. Compose a new email for clonemamaiso3@gmail.com (just type it to the 'To' field, don't click 'To'). Type 'Test case 5' in the field Subject. Click the icon next to the right of 'Aa' icon in the tool bar at the bottom of the New message window to attach 'test case 5' file you already downloaded from Google Drive in the Downloads folder. Finally, send that email.
- Task completion conditions: The email is successfully sent to provided receiver with the file 'test case 5' attached.
- Difficulty: Hard.
- Details in step:

Step 1: Open Edge Browser and navigate to website drive.google.com.

Step 2: Click "Sign in", type clonemamaiso5@gmail.com in "Email or phone" field then click "Next".

Step 3: Type PD12345678! in "Enter your password" field then click "Next".

Step 4: Click Not Now if there is a pop up 'Save your password?'.

Step 5: Go to My Drive in the left sidebar, look for folder 'Test case 5' through 'Search in Drive' and go to that folder.

Step 6: Look for the file 'test case 5'. If you don't see it on the screen, try scrolling down to find it.

Step 7: Right click on 'test case 5' to open Context menu and choose Download.

Step 8: Open a new tab and navigate to website gmail.com.

Step 9: Compose a new email.

Step 10: Type 'Test case 5' in the field Subject.

Step 11: Click the empty space to the right of 'Recipient' to type 'clonemamaiso3@gmail.com'.

Step 12: Click the icon next to the right of 'Aa' icon in the tool bar at the bottom of the New message window to open file attachment window.

Step 13: Open Downloads folder in the file attachment window and choose ‘test case 5.docx’ file you have already downloaded to attach.

Step 14: Make sure all required fields in the New message window are filled out, then click Send to send the email.

2. Result

- Number of attempts until success: 7
- Number of steps: 30
- Time taken: 1952.09 seconds (~32 minutes 32 seconds)
- Average 65 seconds per step.
- Token cost:

	Prompt	Completion	Total
Openai/gpt-5	588251	25194	613445
bytedance/ui-tars-1.5-7b	109816	490	110306

3. Behavior

a) Trajectory:

- The agent begins by detecting that Chrome is closed and deciding to open it to access Google services. After launching Chrome, it selects a saved Guest profile with a pink avatar instead of standard Guest mode. Once in the correct profile, the agent focuses the address bar, types “drive.google.com,” and presses Enter to navigate to Google Drive. When the navigation unexpectedly opens the Google homepage instead, the agent adapts by clicking the Google Apps grid and selecting Drive from the menu.
- Upon reaching the Google Drive sign-in page, the agent enters the provided email, clicks “Next”, types the password and clicks “Next” again to complete authentication. Once logged in, it navigates to “My Drive,” opens the “Test case 5” folder, locates the file “test case 5,” right-clicks to open the context menu, and selects “Download.” The agent notes that the download has likely begun successfully and transitions to the next part of the task which is sending the downloaded file via Gmail.
- To proceed, it opens a new tab using Ctrl+T, types “mail.google.com” and presses Enter. When waiting for Gmail to load, it refocuses the address bar with Ctrl+L. The agent clicks “Compose,” focuses the “To” field and enters the email address

provided. It then clicks the Subject field, types “Test case 5” and proceeds to attach the file.

- Within the file picker, the agent navigates to the Downloads folder, identifies test case 5.docx, and double-clicks it to attach. Once attached, it uses Ctrl+Enter to send email. A “Message sent” notification confirms successful delivery with the attachment, after which the agent ends the process by calling agent.done().

b) Problem:

- Some of the attempts failed due to the verification when logging into Google Drive.
- The agent always fails the first-time password entering. After downloading the file, the model will reflect the user instruction about Guest mode of Chrome. Sometimes it will focus on the current trajectory and doesn't dwell on past steps but sometimes it will check the profile and encounter some looping steps before closing the current Chrome browser. Sometimes the model still clicks To in compose window instead of typing text directly that cause loops.