

1. Introduction

- Details: Open Chrome with Guest Profile (not Guest Mode) by clicking the avatar of the profile (not the name) and login to Google Drive with my account provided: (After logging in successfully, ignore what Chrome profile you are using and continue the task). 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 that downloaded file in Downloads folder, then append the text ' How do you do that?' to the end of the document and save the changes (If you successfully append the text, that means you already downloaded the file, no need to check again). Next, open a new Chrome tab in the Chrome you have already opened and login to Gmail. Compose a new email for clonemamoiso3@gmail.com (click the empty space to the right of 'To' to type text, 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' that was correctly modified attached.
- Difficulty: Hard
- Details in step:

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

Step 2: Click "Sign in", type clonemamoiso5@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 File Explorer, then open the downloaded file 'test case 5' in Download folder.

Step 9: Append the text ' How do you do that?' to the end of the document and save the changes.

Step 10: Back to Edge (if Edge is not in fullscreen mode, make it become full screen), open a new tab and navigate to website gmail.com.

Step 11: Compose a new email.

Step 12: Type 'Test case 5' in the field Subject.

Step 13: Click the empty space to the right of 'Recipient' to type receiver.

Step 14: 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 15: Open Downloads folder in the file attachment window and choose 'test case 5.docx' file you have already downloaded to attach.

Step 16: 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: 9
- Number of steps: 48
- Time taken: 1140.78 seconds (~19 minutes)
- Average 23.8 seconds per step.
- Token cost:

	Prompt	Completion	Total
gemini-2.5-flash	726660	7359	755136
bytedance/ui-tars-1.5-7b	169436	776	170212

3. Behavior

a) Trajectory:

- The agent begins with identifying that Visual Studio Code is open and Google Chrome is available on the taskbar. It opens Chrome, selects the red Guest profile avatar, and searches for Google Drive. From the search results, it clicks the correct Google Drive login link and proceeds to the login page. It enters the provided account and login. After reaching Google Drive, it dismisses a popup, opens the "Test case 5" folder, and downloads the "test case 5" file.
- Next, the agent tries to open the downloaded file via Windows search but mistakenly triggers a web search. It corrects this by opening File Explorer, navigating to Downloads, and double-clicking test case 5.docx to open it in Word. Upon seeing

the Protected View banner, it enables editing, types “How do you do that?” but inserts it at the beginning by mistake. The agent then repositions the cursor unsuccessfully, deletes all text, and retypes the correct content: “Test case number five. How do you do that?” It saves the document with Ctrl+S, ensuring the update is complete.

- After saving, the agent attempts to switch back to Chrome but encounters repeated activation failures, including accidentally opening SQL Server Management Studio. It closes the unintended application and retries by clicking the Chrome icon directly. Chrome reopens at the profile selection screen, so the agent reselects the Guest profile, restoring access to Google Drive. It then opens a new tab, searches for Gmail, clicks the login result, waits for the page to load, and dismisses a notifications popup.
- In Gmail, the agent clicks Compose, enters the receiver email address, sets the subject to “Test case 5,” and clicks the attachment icon. From the Downloads folder, it selects and attaches test case 5.docx, confirms the upload, and sends the email. After verifying that the message was successfully sent, the agent concludes the process by executing agent.done().

b) Problem:

- This test case encounters the same issue as Test Case 5. Additionally, after entering the email subject, the agent forgets that the file has already been downloaded and unnecessarily navigates back to Google Drive to download it again. In other attempts, it forgets that it has already modified the downloaded file and attempts to reopen it for verification.
- After several attempts, I switched the main generation model to Gemini, which was then able to complete the task. However, despite completing the workflow, it still inserted the appended text in the wrong location within the file and had trouble returning to Chrome after successfully modifying the “test case 5” file.