I. Introduction

- Description: This test case requires the agent to open Microsoft Edge, create a new document in Google Docs. Then it accesses the link to a research paper, read and summarize the paper according to the user's questions, write the summaries into the document, and finally download the file.
- Input: Link of a research paper and questions for agent to summarize the key idea.
- Output: A downloaded document file providing report that answers all the questions of the users.
- Succeed condition: The file must be downloaded locally. The summary should be concise but still contains all the main points and must not have any wrong information related to the original paper.

II. Experiment Setup

- Agent S3 agentic framework was used to execute this test case.
- Most of the time Gemini 2.5 flash was used as main generation model.
- GPT 5 and Gemini 3 Pro Preview were also used in some attempts to serve as a comparison with Gemini 2.5 flash.
- Grounding model: UI TARS 1.5 7B.

III. Instruction versions

1. V1 – Page by page + Full page

1.1. Description:

- This is the first version of instruction only for testing whether the agent can complete this task.
- For paper reading, the idea is to have the agent process the document page by page and generate a summary for each page. Then the agent will combine and summarize these page level summaries into a final report that answers the user's questions.

1.2. Instruction:

- Follow the instruction below carefully. Analyze the screenshot thoroughly and move the mouse to the exact position when needed. Do not interact with anything unrelated to the task. Your task is open Microsoft Edge, then go to website docs.google.com to open Google Doc and create a Blank document (option on the most left) named "Summarize." Next, open a new tab and navigate to

https://arxiv.org/pdf/2506.16020 to access the paper titled “VS-Singer: Vision-Guided Stereo Singing Voice Synthesis with Consistency Schrödinger Bridge.” Read the paper page by page. On the top toolbar of the PDF viewer, find the box showing the current page number (for example, “1 of 5”). Click inside that box (for example, click '1', not '5', the current page number color will change to white with blue background around), type the page number you want to go to, and then press Enter on the keyboard to jump to that page. Do not use any other buttons in the toolbar or scroll manually. After reading each page, switch to the Google Docs tab in the same browser and take notes summarizing the key points from that page. Once you finish noting, return to the paper tab in the same browser and continue reading the next page. Repeat this process until you have gone through the entire document. You must read and summarize all pages before doing the next step. Go back to Google Docs tab in the same browser, review your notes, and refine them into a complete summary following this format: What is the research/project about? What are its main contributions? What is its methodology? What are the results? What are the current challenges or limitations? What are the future directions? Finally, download that document file.

1.3. Problem:

- Although the agent was able to complete the task, it often requires multiple attempts before achieving a successful run. The failed attempts were primarily due to the agent mistakenly clicking the ‘+’ icon to open a new browser tab instead of the ‘+’ for creating a new blank document, difficulties in navigating pages using the page number input box, mistakenly switching back to the Google Docs applications instead of the Google Docs tab in the same browser and producing summaries that were suboptimal, containing inaccuracies or information does not present in the original paper.

2. V2 – Step + Cross-page coherence

2.1. Description:

- In this version, I divided the instruction into small steps to make planning and performing action easier for the agent.
- In Step 1, an additional action is included to maximize the Edge window if it is not already maximized. This ensures clearer element recognition, reducing the likelihood of the agent misclicking due to incorrect coordinated detection.

- In Step 5, I added instructions indicating that if the content of a page continues onto the next page, the agent must proceed to the following page and read the entire continuation before summarizing. I also introduced additional 'Important summary rules' constraints to ensure the quality and accuracy of the agent's summaries.
- To solve the problem of switching to Google Docs application instead of Google Docs tab, I made it use Ctr + Tab instead of letting it figure out for itself.

2.2. Instruction:

- Follow the instruction below carefully. Analyze the screenshot thoroughly and move the mouse to the exact position when needed. Do not interact with anything unrelated to the task. Do not switch to other applications when doing this task.
- Step 1. Open Microsoft Edge. If Edge is not maximized, move the mouse to the top-right corner of the Edge window and click the square icon (the one to the left of the X close button). This makes the Edge window fill the whole screen, but the title bar and taskbar remain visible (not the F11 fullscreen mode).
- Step 2. Go to the website https://docs.google.com. When the page is fully loaded, create a Blank document (the option on the far left). Name this new document "Summarize."
- Step 3. Open a new tab in the same Edge window. Go to https://arxiv.org/pdf/2506.16020 to access the paper titled "VS-Singer: Vision-Guided Stereo Singing Voice Synthesis with Consistency Schrödinger Bridge."
- Step 4. On the PDF viewer toolbar, find the box showing the current page number (for example, "1 of 5"). Click inside that box (for example, click '1', not '5', the current page number color will change to white with blue background around), type the page number you want to go to, and then press Enter on the keyboard to jump to that page. Do not use other toolbar buttons or scroll manually. Use only this page number box to navigate pages.
- Step 5. Start reading the paper page by page. For each page: Carefully read and analyze the content on the current page. Check if the content continues onto the next page (for example, when a sentence or paragraph is cut off). If yes, read the next page immediately after to complete that idea or paragraph before summarizing. After finishing that continuous section (which may span multiple pages), switch to the Google Docs tab (which is a tab in the same browser, not the application, use Ctr + Tab to switch between tabs). Write a summary of the key points from what you just read. Important summary rules: Summarize faithfully using only information present in the paper. Do not paraphrase excessively or reword in a way

that changes meaning. Do not add personal opinions, comments, or inferred details beyond what the paper states. Maintain logical continuity between pages if one idea spans multiple pages (connect sentences naturally so the meaning remains intact).

- Step 6. After finishing your notes for that section, return to the paper tab using the same Edge window. Use the page number box again to go to the next unread page. Repeat the process in Step 5 until you have read and summarized all pages of the paper.
- Step 7. Once you have finished reading the entire paper and taking notes: Go back to the Google Docs tab. Review all your notes carefully. Combine and refine them into a complete structured summary following this format: What is the research/project about? What are its main contributions? What is its methodology? What are the results? What are the current challenges or limitations? What are the future directions?
- Step 8. After completing the full summary, download the document file from Google Docs.

2.3. Problem:

- The action of clicking the square button next to the X button to maximize the Edge window is unreliable, as the agent frequently clicks the wrong location. Because of this, I switched to using the Win + Up shortcut instead. However, this introduced a new issue: when the Edge window is not maximized, the agent can successfully maximize it, but when it is already maximized, the agent sometimes incorrectly detects it as not maximized and still attempts to maximize it.
- Some issues from Version 1 also remain, such as difficulties navigating pages using the page number input box and suboptimal summary quality.

3. V3 – More coherence between pages

3.1. Description:

- I added a summary structure for each page in Step 4 so the agent can better understand and clearly present the continuity between pages.

3.2. Instruction:

- Follow the instruction below carefully. Analyze the screenshot thoroughly and move the mouse to the exact position when needed. Do not interact with anything unrelated to the task. Do not switch to other applications when doing this task.
- Step 1. Open Microsoft Edge. If Edge is not maximized, use Win + Up else continue the task. Edge is maximized if Edge window fills the whole screen, but the title bar and taskbar remain visible (not the F11 fullscreen mode).
- Step 2. Go to the website https://docs.google.com. When the page is fully loaded, create a Blank document (the option on the far left). Name this new document "Summarize."
- Step 3. Open a new tab in the same Edge window. Go to https://arxiv.org/pdf/2506.16020 to access the paper titled "VS-Singer: Vision-Guided Stereo Singing Voice Synthesis with Consistency Schrödinger Bridge."
- Step 4. Start reading the paper page by page from page 1. For each page: Read and analyze the content shown on the current page. If the last sentence or paragraph on the page continues to the next page, go to the next page using the page number box, read only the continuation part until that section or paragraph ends, Then stop reading further new sections on that next page. When a section or subsection spans multiple pages, keep the summary connected across pages (Example: If page 1 ends halfway through "Introduction" and page 2 continues it, summarize it as: Page 1 Summary: I. Introduction ... Page 2 Summary: I. Introduction (continue) ... If page 2 also starts a new section such as "Related Work," include it as: Page 2 Summary: I. Introduction (continue) ... II. Related Work 1. Subsection 1 - main content 2. Subsection 2 - main content). After reading the necessary part, switch to the Google Docs tab (in the same Edge window). Write a summary for that page in this structured format: Page [number] Summary: Sections -> Subsections -> Key ideas or findings of each Subsections (use bullet points or short sentences) - Include "(continue)" tag if it connects to the previous page. Make sure each page summary clearly shows section boundaries and continuation markers like "(continue)" so the summaries stay connected across the document. Important summary rules: Summarize faithfully using only information present in the paper. Do not paraphrase excessively or reword in a way that changes meaning. Do not add personal opinions, comments, or inferred details beyond what the paper states. Maintain logical continuity between pages if one idea spans multiple pages (connect sentences naturally so the meaning remains intact). Use Ctrl + End to move to the end of the document before typing the summary, don't click anywhere. How to change page: On the PDF viewer toolbar, find the box showing the current page number (for example, "1 of 5"). Click inside that box (for example, click '1', not '5', the current

page number color will change to white with blue background around), type the page number you want to go to, and then press Enter on the keyboard to jump to that page. Do not use other toolbar buttons or scroll manually. Use only this page number box to navigate pages.

- Step 5. After finishing your notes for that section, return to the paper tab using the same Edge window. Continue reading the page if not finish or use the page number box again to go to the next unread page. Repeat the process in Step 4 until you have read and summarized all pages of the paper.
- Step 6. Once you have finished reading the entire paper and taking notes: Go back to the Google Docs tab. Review all your notes carefully. Combine and refine them into a complete structured summary following this format: What is the research/project about? What are its main contributions? What is its methodology? What are the results? What are the current challenges or limitations? What are the future directions?
- Step 7. After completing the full summary, download the document file from Google Docs.

3.3. Problem:

- I observed that the reason the agent continues to trigger the maximize action even when the Edge window is already maximized is related to Edge's UI design. The browser includes curved inner borders, and in certain cases the agent misinterprets these curved edges as an indication that the window is not maximized.
- Based on the chat log and page summary examples, although the agent could recognize continuity in its plan and the content by marking certain parts in the next page summary as "continued," the logical flow and connection with the sentences or paragraphs from the previous page were still not clearly maintained or were sometimes omitted.
- When the page changing step failed, the agent sometimes hallucinated the next page on its own, resulting in incorrect summaries.
- Despite explicit summarization guidelines in the instructions, the agent's output still contained inaccuracies, unspecified or unverifiable claims and spelling errors.

4. V4 – Section by section + Small steps

4.1. Description:

- In this version, I tried another approach by having the agent read the paper and summarize it section by section instead of page by page.
- The concept is for the agent to read the section on the current page, then verify on the next page whether the section has ended. If the section continues, the agent should proceed with reading until a new section begins.
- I broke down the original Step 4 into smaller conditional steps, if condition A is met, proceed to the next step; otherwise, repeat the current step or the previous one. This reduces the cognitive load on the agent, preventing it from having to handle too many tasks at once.
- I realized that the page changing issue could be solved more easily by having the agent use the right arrow key to move to the next page instead of clicking the page number box.
- Regarding the issue where the agent mistakenly clicked a '+' icon elsewhere instead of the one for creating a new blank document, I added a more detailed description of the "Create new blank document" option. This has significantly reduced the occurrence of such errors.
- In previous versions, I had the agent check for continuity at the end of each page before moving on. However, it often made incorrect judgments, so I changed the workflow: the agent must now always proceed to the next page to verify continuity directly.
- Switching to the new page chaging method and simplifying step 4 into smaller, clearer steps helped the agent operate more smoothly. It can recognize the continuity between sections across consecutive pages, identifying where a section starts or ends in which column and page. In this version, I used a paper with 10 pages, which is twice the length of the papers used in previous versions.
- **In this version, the agent is now able to execute and complete the test case smoothly, with far fewer minor errors.**

4.2. Instruction:

- Follow the instruction below carefully. Analyze the screenshot thoroughly and move the mouse to the exact position when needed. Do not interact with anything unrelated to the task. Do not switch to other applications when doing this task.
- Step 1. Open Microsoft Edge. If Edge is not maximized, use Win + Up else continue the task. Edge is maximized if Edge window fills the whole screen, but the title bar and taskbar remain visible (not the F11 fullscreen mode).

- Step 2. Go to the website https://docs.google.com. When the page is fully loaded, on the main Docs screen, look for the section labeled “Start a new document”, under that text, locate the large white rectangle with a multicolor plus icon (blue, red, yellow, green) in the center. Click inside that white rectangle (the “Blank document” option). Name this new document “Summarize.”
- Step 3. Open a new tab in the same Edge window. Go to https://arxiv.org/pdf/2106.07447 to access the paper. Carefully examine the paper’s format. If the paper is organized into two columns, you should read the text from the left column first, then proceed to the right column, following a top-to-bottom order. The end of a page corresponds to the bottom of the right column, while the beginning of a page corresponds to the top of the left column.
- Step 4. Start reading the paper section by section. Begin reading all content under this section, including any subsections. The content of each page is fully displayed, do not scroll down. Do not switch back to Google Docs tab to summarize before doing the steps below.
- Step 5. When you reach the end of the current page, you must go to the next page to check whether the same section continues. How to change page: Use right arrow key to go to the next page. Do not use other toolbar buttons or scroll manually. If you can't change the page yet, keep trying until it works, do not hallucinate the image or content of the next page.
- Step 6. Keep reading the content to the point a new section title appears (for example, “5. Conclusion”, “V. Results”, etc., not Fig. or Table.), read only up to the line where the previous section actually ends, then stop immediately before the new section begins and process to step 7; if you don't see any new section title on this page, repeat step 5.
- Step 7. Only after confirming that a new section begins may you stop reading and proceed to summarize the previous section. When a section is fully read, switch to the Google Docs tab (which is a tab in the same browser, not the application, use Ctr + Tab to switch between tabs). Use Ctrl + End to move to the end of the document before typing the summary, don't click anywhere. Write a summary of the key points from what you just read for each subsection of that section (Just type the text to the document, no need to check for truncated or missing text after typing). Important summary rules: Summarize faithfully using only information present in the paper. Do not paraphrase excessively or reword in a way that changes meaning. Do not add personal opinions, comments, or inferred details beyond what the paper states.

- Step 8. After finishing your notes for that section, return to the paper tab using the same Edge window. Continue reading the next section. Repeat the process from step 4 to 7 until you have read and summarized all sections of the paper.
- Step 9. Once you have finished reading the entire paper and taking notes: Go back to the Google Docs tab. Review all your notes carefully. Combine and refine them into a complete structured summary following this format: What is the research/project about? What is its methodology? What are the results?
- Step 10. After completing the full summary, download the document file from Google Docs.

4.3. Problem:

- Minor clicking issues still occurred, and the quality of section title recognition and summarization remained suboptimal, mainly due to the main generation model's limitations in extracting information from screenshots, producing accurate and quality summaries.
- In several attempts, GPT-5 identified the text as difficult to read and tried to zoom in further on its own, which resulted in a loop and eventual failure.

5. V5 – Section by section + Half page

5.1. Description:

- To address the issue of poor summary quality that may be caused by the small text size making it difficult for the agent to read information accurately (as evidenced by frequent mistakes in summarizing parameters such as batch size, GPU, and steps in the experiment setup), I tested enlarging a page of the paper.
- Instead of displaying the entire page at once, I zoomed in so that only half of the page is shown at a time, making it easier for the agent to read. I had the agent navigate between the two halves of a page using Space and Shift + Space keyboard. For papers with a two-column format, I instructed it to read in the following order (left to right column, from top to bottom): upper left, lower left, upper right, then lower right.

5.2. Instruction:

- Step 3. Carefully examine the paper's format. If the paper is in two-column layout, the reading order of every page follows a strict sequence: Upper Left → Lower Left

→ Upper Right → Lower Right (top-to-bottom, left-to-right). You must never look at both columns at the same time. You must never make conclusions based on text in a column you have not read yet. Reading direction rules: You only read the current column of the currently half-page. You do not examine the other column. You do not extract section titles from columns you have not reached. You do not infer the structure from future columns. The end of a page corresponds to finishing the lower-right column. The beginning of a page corresponds to the upper-left column.

- Step 4. Start reading the paper section by section (Ignore Abstract). Read only the current column of the current half-page, exactly in this order: Determine whether you are in upper half or lower half. Each page initially shows only the upper half. Press Space once to reveal the lower half. Press Shift + Space once to return to the upper half. If you see a large white space between the taskbar and the text, it means you are in the lower half. Otherwise you are in the upper half. Read only one column at a time. For each half-page: Upper half → Upper Left column. Read only the upper-left column. Do not inspect or process the upper-right column yet. Press Space to move to lower half → Lower Left column. Read only the lower-left column. Ignore the lower-right column. Press Shift + Space to return to upper half → Upper Right column. You will see the previously read upper-left column again, ignore it. Read only the upper-right column. Press Space to go to lower half → Lower Right column. Read only the lower-right column. After finishing the lower-right column, you are at the end of the current page, proceed to Step 5. When reading each column: If you encounter a new section title (e.g., “5. Conclusion”, “V. Results”, “3 Method”), AND it appears in the column you are currently reading, immediately stop reading and proceed to Step 6. You must ignore any section titles located in columns you have not reached yet. You must not detect section titles in: The right column while you are reading the left column.
- Step 5. When you reach the end of the current page, you must go to the next page to check whether the same section continues. How to change page: Use right arrow key to go to the next page. Do not use other toolbar buttons or scroll manually. If you can't change the page yet, keep trying until it works, do not hallucinate the image or content of the next page.
- Step 6. Keep reading the content until you reach the exact point where a new section title begins (for example, “5. Conclusion”, “V. Results”, etc., not Fig. or Table.). When you encounter a new section title, you must read all text in the same current column above that title on the same page as part of the current section (do not

conclude that the section ended on the previous page just because the new section title appears on the next page), then process to step 7. You must always check the beginning of the next page to confirm whether the current section continues. If you fully read the current page and don't see any new section title on this page, repeat step 5.

- Step 7. Only after fully reading the section may you stop reading and proceed to summarize the previous section. Switch to the Google Docs tab (which is a tab in the same browser, not the application, use Ctr + Tab to switch between tabs). Write a summary of the key points from what you just read for each subsection of that section. Use Ctrl + End to move to the end of the document before typing the summary, don't click anywhere. Just type the text to the document, no need to check for truncated or missing text after typing. Important summary rules: Summarize faithfully using only information present in the paper. Do not paraphrase excessively or reword in a way that changes meaning. Do not add personal opinions, comments, or inferred details beyond what the paper states.
- Step 8. After finishing your notes for that section, return to the paper tab using the same Edge window. Continue reading the next section. Repeat the process from step 4 to 7 until you have read and summarized all sections of the paper.

5.3. Problem:

- Even with large text after zooming to show only half a page at a time, Gemini 2.5 Flash still produced incorrect summaries or added information that is not present in the original text. GPT-5 summarized more accurately but repeatedly misidentified column positions, preventing it from completing the entire paper. Gemini 3 Pro also delivered stronger summaries, but I ran out of credits before it could finish the full paper.
- Both GPT-5 and Gemini 2.5 Flash struggle to correctly identify which column belongs to the current half page view order. For example, when the Abstract finishes in the upper left column, the agent may mistakenly treat the upper right quadrant as a continuation even though the instructions explicitly restrict it to read and process only on the currently reading order column. This misrecognition leads to repeated page switching attempts, looping behavior, and incorrect summaries where content from other sections is incorrectly attributed to the current section.
- Gemini 3 Pro Preview demonstrated strong reasoning when identifying the positions of columns within each half page and followed the instructions correctly. However,

it consistently saved the text of each section into knowlegde, and its token cost was relatively high.

IV. Conclusion

- Using the full page, section by section reading approach and after several rounds of instruction enhancement, the agent was able to complete this test case with fewer attempts. However, the summary quality remains imperfect and still depends heavily on the main generation model's ability to extract information from screenshots and produce accurate summaries.
- With the half page reading approach, although the text appears larger and Gemini 2.5 Flash shows slight improvement in summary quality, with fewer inaccuracies or details do not present in the original paper, errors still persist. While the agent can complete the task, Gemini 2.5 Flash and GPT-5 both continue to hallucinate information across columns within the same half page, despite explicit constraints in the instructions.
- Currently, Gemini 3 Pro Preview demonstrates the best performance to planning as instructions and the highest summary quality, though at a significantly higher cost.