

## For Part 1 of the Question: Setup & Agent Framework Scaffold

Goal: {goal}

Observation:

{obs}

What is the next best action?

Respond exactly in the format: CLICK("<element>") or OPEN\_APP("<app>") or STATUS("complete")

""

## For Part 2 of the Question: Prompting & Evaluation Strategy

**ZERO\_SHOT** = ""\n

Goal: {goal}

You are an autonomous agent controlling an Android UI.

Below is the current screen's list of visible elements:

{obs}

Instructions:

1. Pick exactly one action that moves you toward the goal.
2. Only click or open elements that appear in the list above.
3. When done, use STATUS("complete").
4. Respond with exactly: Action: [your\_action]

Available actions:

- CLICK("element\_text") - click on UI element by its visible text/description
- OPEN\_APP("app\_name") - open an application
- INPUT\_TEXT("text") - type text into a field
- SCROLL("direction") - scroll up/down/left/right
- STATUS("complete") - when task is finished

Example: Action: CLICK("Camera")

Your turn:

""

**IMPROVED\_FEW\_SHOT** = ""\n

You are an Android UI automation agent. Your goal is: {goal}

{history}

Here are examples of good agent behavior:

Example 1:

Goal: Take a photo

History: No previous actions taken.

Visible elements:

- Phone
- Messages
- Chrome
- Gmail
- Photos
- Search

Reasoning: I need to open the Camera app to take a photo, but I don't see it in the current elements. I should look for a camera-related app or use search.

Action: OPEN\_APP("Camera")

Example 2:

Goal: Take a photo

History: Step 1 - OPEN\_APP("Camera") - Successfully opened Camera app

Visible elements:

- Options
- Shutter
- MODE LIST
- FILMSTRIP
- Z-
- Z+

Reasoning: Camera is now open and I can see the Shutter button. I'll click it to take the photo.

Action: CLICK("Shutter")

Example 3:

Goal: Create a timer with 5 minutes

History: Step 1 - OPEN\_APP("Clock") - Opened Clock app

Visible elements:

- Timer
- Alarm
- World Clock
- Stopwatch

Reasoning: I'm in the Clock app and need to access the Timer function to set a 5-minute timer.

Action: CLICK("Timer")

Example 4:

Goal: Create a new contact for John Smith

History:

Step 1 - OPEN\_APP("Contacts") - Opened Contacts app

Step 2 - CLICK("Add") - Clicked add contact button

Visible elements:

- First name
- Last name
- Phone
- Email
- Save
- Cancel

Reasoning: I'm in the add contact form. I need to fill in the first name field with "John".

Action: INPUT\_TEXT("John")

Now your turn:

Goal: {goal}

{history}

Visible elements:

{obs}

Reasoning:

Action:

""""

**SELF\_REFLECTION** = """"\

Goal: {goal}

{history}

Current visible UI elements:

{obs}

Please follow these steps:

1. Analyze the goal and current situation
2. Consider what actions you've already taken (if any)
3. Propose the next action: CLICK("element"), OPEN\_APP("app"), INPUT\_TEXT("text"), SCROLL("direction"), or STATUS("complete")
4. Verify that your chosen element actually appears in the list above
5. Provide your reasoning

Format your response as:

ANALYSIS: [Your analysis of the current situation and progress toward goal]

ACTION: [Your proposed action]

VERIFICATION: [Confirm the element exists in the list - true/false]

REASONING: [Why this action moves toward the goal]

""""

### For Part 3 of the Question: Benchmarking & Report

**ZERO\_SHOT** = """"\

Goal: {goal}

You are an autonomous agent controlling an Android UI.

Below is the current screen's UI element tree:

{obs}

Instructions:

1. Pick exactly one action that moves you toward the goal.
2. Only interact with elements that appear in the tree above.
3. When done, use STATUS("complete").
4. Respond with exactly: Action: [your\_action]

Available actions:

- CLICK("element\_text") - click on UI element by its visible text/description
- CLICK("element\_N") - click on element by its index number
- OPEN\_APP("app\_name") - open an application
- INPUT\_TEXT("text") - type text into a field
- SCROLL("direction") - scroll up/down/left/right
- STATUS("complete") - when task is finished

Example: Action: CLICK("Camera")

Your turn:

""

**FEW\_SHOT** = ""\

Goal: {goal}

You are an autonomous agent controlling an Android UI. Here are examples of good action sequences:

EXAMPLE 1 - Take a photo:

Goal: Take one photo

Step 1: OPEN\_APP("Camera") - Opens camera app

Step 2: CLICK("element\_2") - Clicks shutter button to take photo

Step 3: STATUS("complete") - Task completed

EXAMPLE 2 - Create contact:

Goal: Create contact for John Smith, phone +1234567890

Step 1: CLICK("element\_2") - Opens contacts/phone app

Step 2: CLICK("element\_6") - Navigates to contacts section

Step 3: CLICK("element\_1") - Clicks add new contact button

Step 4: INPUT\_TEXT("John") - Enters first name

Step 5: INPUT\_TEXT("Smith") - Enters last name

Step 6: INPUT\_TEXT("+1234567890") - Enters phone number

Step 7: CLICK("element\_2") - Saves the contact

Step 8: STATUS("complete") - Task completed

{history}

Current UI element tree:

{obs}

Instructions:

1. Follow the pattern from examples above
2. Use exact element references from the tree
3. Take one action at a time toward the goal

Action: [your\_action]

""

**SELF\_REFLECTION** = ""\

Goal: {goal}

{history}

Current UI element tree:

{obs}

Please follow these steps:

1. Analyze the goal and current situation
2. Consider what actions you've already taken (if any)
3. Propose the next action: CLICK("element"), OPEN\_APP("app"), INPUT\_TEXT("text"), SCROLL("direction"), or STATUS("complete")
4. Verify that your chosen element actually appears in the tree above
5. Provide your reasoning

Format your response as:

ANALYSIS: [Your analysis of the current situation and progress toward goal]

ACTION: [Your proposed action]

VERIFICATION: [Confirm the element exists in the tree - true/false]

REASONING: [Why this action moves toward the goal]

""""