



USER INTENT (I_{USER})

Please use **IMDB** to search for "Oppenheimer" and tell me who is the first actor. Please use **Wiki** to search for the actor and tell me what sci-fi action movie he starred in in 2005. Please use **Douban** to search for the movie and tell me who is its director.



Task Parser, \mathcal{P}_θ

Toolset



IMDb Atomic Tasks 1

D_1 : Use **IMDB** to search for **Oppenheimer** and identify the first actor listed.

C_1 : None a_1 : Jarvis
 ξ_1 : Device_1 σ_1 : Pending

W Atomic Tasks 2

D_2 : Search **Wikipedia** for that actor to identify their 2005 sci-fi action movie.

C_2 : None a_2 : PC Agent
 ξ_2 : Device_2 σ_2 : Pending



Atomic Tasks 3

D_3 : Use **Douban** to search for the movie found in step 2 and identify its director.

C_3 : None a_3 : Jarvis
 ξ_3 : Device_1 σ_3 : Pending



Task Topology (\mathcal{T})

Atomic Tasks 1

Result: Cillian Murphy
Status: Succeeded ●

Atomic Tasks 2

Result: Batman Begins
Status: Succeeded ●

Atomic Tasks 3

Result: Cillian Murphy
Status: Executing ●



Workflow Manager

Atomic Tasks 1

Use **IMDB** to search for **Oppenheimer** and identify the first actor listed.

C_1 : None a_1 : Jarvis
 ξ_1 : Android_1 Result: Cillian Murphy

Atomic Tasks 2

Use **Wikipedia** to search for **Cillian Murphy** and find out the sci-fi action movie starred in during 2005.

C_2 : C. Murphy a_2 : PC-Agent
 ξ_2 : PC_1 Result: Batman Begins

Agents (\mathcal{A})

PC-Agent

Jarivs

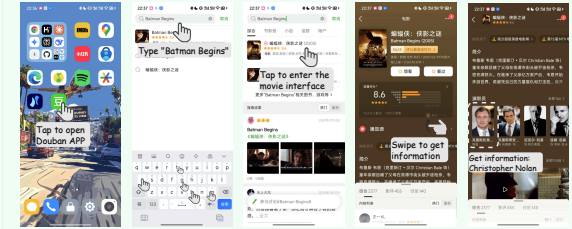
Mobile-Agent-e

Context Evolution Mechanism

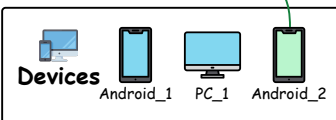
Update Agents Experience

Atomic Tasks 3 a_3 : Jarvis

Use **Douban** to search for the **Batman Begins** and identify its director.



C_3 : Batman...
 ξ_3 : Android_2 Result: Cillian Murphy



Multi-level Evaluation Framework

Level1: Weighted Pathway Success Rate (WPSR)

$$WPSR = \sum_{j,i} w_{j,i} \cdot S_{task}(j, i)$$

Task completion ability is assessed by a difficulty score ($D_{j,i}$) derived from the path length and number of applications.

Level2: Fine-grained traversal Mean Atomic Tasks Completion Ratio (MATCR)

$$MATCR = \frac{1}{N} \sum_{j=1}^N \frac{k_j}{n_j}$$

Measures the average ratio of consecutively completed atomic tasks.

Positional-Weighted Atomic Tasks Success Rate (p-ATSR)

$$p-ATSR = \frac{\sum_{j=1}^N \sum_{i=1}^{n_j} p(i) \cdot S_{atomic}(j, i)}{\sum_{j=1}^N \sum_{i=1}^{n_j} p(i)}$$

Measures the average ratio of consecutively completed atomic tasks.

Level3: Error Attribution

Knowledge Deficit (KD)
Lack of necessary knowledge.

Perceptual Error (PE)
Failure to extract correct information from the UI.

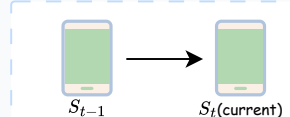
Operational Error (OE)
Imprecise execution of a correct action.



Hybrid Visual-Structural Perception

Visual Perception ($\mathcal{V}(I_t)$)

Structural Perception ($\mathcal{S}(X_t)$)



[1] ViewPager {id='com.google...'}
[2] TextView {text='Wed, ...', id='com.google.android.apps...'}
[3] ...

\mathcal{J} : Use **Douban** to search for the **Batman Begins** and identify its director.

h_{t-1} :

Step1: <think>...</think><action>...</action>
Step2: <think>...</think><action>...</action>
...
Step $_{t-1}$: <think>...</think><action>...</action>

$$\mathcal{P}_t = \mathcal{J} \oplus h_{t-1} \oplus o_t \Rightarrow (th_t, a_t) \sim \pi_\theta(\cdot | m_t)$$

th_t : <think> In my previous steps, I opened the Douban app, successfully searched for "Batman Begins," and accessed its details page. On this page, I couldn't find any information about the film's director; I need to scroll down to continue browsing. </think>
 a_t : <action> Swipe("DOWN", "MEDIUM") </action>

Atomic Action Space:
 $\mathcal{A} = \{TAP(u), INPUT(u, txt), SWIPE(d), DRAG(u_s, u_e), SYS(k)\}$

Jarvis Kernel

atomic action

