



Department of
Computer Science

香港城市大學
City University of Hong Kong

PUO-Bench: A Panel Understanding and Operation Benchmark with A Privacy-Preserving Framework

Wei Lin^{12*}, Yiwei Zhou^{13*}, Junkai Zhang^{14*}, Rui Shao^{15*}, Zhiyuan Zhao¹,
Junyu Gao^{15†}, Antoni B. Chan², Xuelong Li^{1†}

¹The Institute of Artificial Intelligence, China Telecom (TeleAI)

²Dept. of Computer Science, City University of Hong Kong, ³Beijing Institute of Technology,

⁴Beijing University of Technology, ⁵Northwest Polytechnical University

Background of PUO

The diagram illustrates the background of PUO (Physical User-Oriented) agents through two examples of agent interactions with different environments.

Example 1: Digital Environment (News Page)

- Agent Question:** "What is the function of the element at [0.034, 0.132, 0.098, 0.208]?"
- Agent Action:** "Clicking the button 'BACK' will return to the previous page."
- Agent Question:** "How to list all video related news?"
- Agent Action:** "Click the button with text of 'VIDEO', at [0.863, 0.179]."

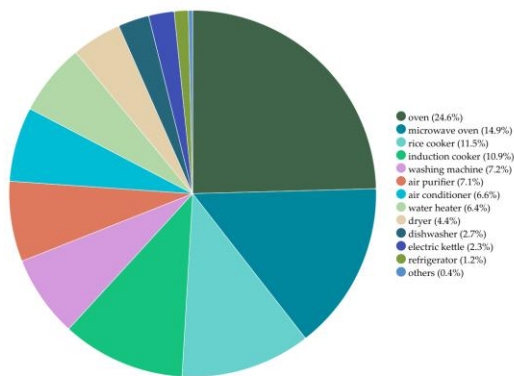
Example 2: Physical Environment (Kitchen)

- Agent Question:** "What is the function of the module at [0.436, 0.435, 0.465, 0.456]?"
- Agent Action:** "Clicking the button text with 'AIR FRY' to active air fry mode."
- Agent Question:** "How should I start the cooking?"
- Agent Action:** "After completing the setup, press the start button, at [0.463, 0.463]"

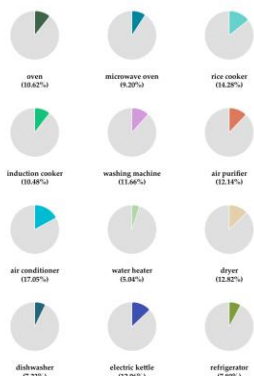
- The generalization of UI agents remains largely confined to interfaces within the digital world
- Assist users in interacting with physical objects and environments featuring panel interfaces
 - operate unfamiliar appliances
 - Interact with complex industrial equipment

PUO Dataset

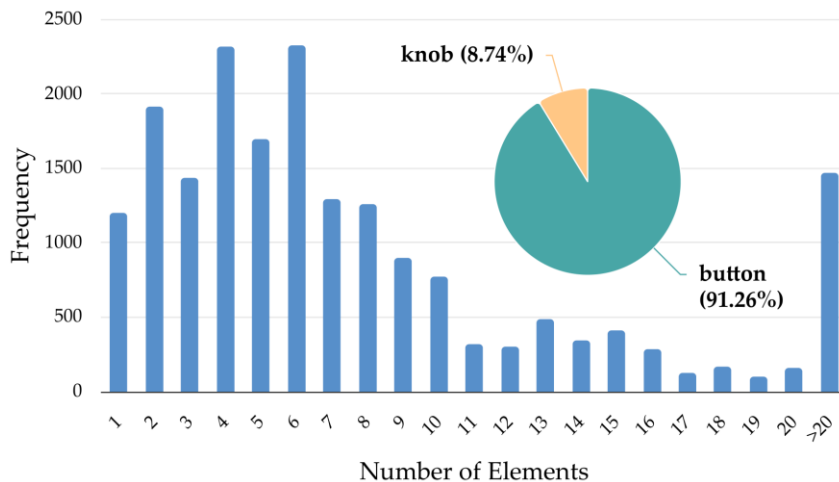
19k panel images & 430k instruction-following QA pairs



(a) The percentage of each category in the dataset.



(b) Average panel area ratio of each category.



PUO Dataset

19k panel images & 430k instruction-following QA pairs

Keyword	Explanation
Type	The category of the UI element, such as “button” or “knob,” which defines its interaction style.
Position	Coordinates of the top-left and bottom-right corners of the UI element’s bounding box.
Text	Surrounding textual information that provides context or describes the purpose of the UI element in detail.
Icon	Visual symbol or graphic associated with the UI element, offering additional context about its role or category.
Function	Specific operational role or task performed by the UI element on the appliance it belongs to.



Type: knob
Position: (188, 428, 600, 858)
Text: Temperature °F
min warm 250 300 350 400 450 broil toast airfry
Icon:
Function: This is a knob that may function to adjust the oven's temperature and modes; when adjusted, it can select different temperatures (e.g., 250°F to 450°F) or functions (e.g., warm, broil, toast, airfry)

Type: button
Position: (427, 639, 514, 722)
Text: Temp
Icon: A triangle pointing upwards
Function: This is a button that may function to adjust the temperature, pressing it may increase the set temperature of the air conditioner.



PUO QA Benchmark

(a) Panel Description

- **Button 'Micro Power'** - This button may function to adjust the microwave power level.
- **Button 'Express'** - This button may function to provide quick cooking settings.
- **Button 'Defrost/Clock'** - This button may function to defrost food and set the clock.
- **Button 'Auto Cooking'** - This button may function to automatically set cooking parameters based on food type.
- **Knob 'Time' and 'Weight'** - This knob may function to set time and weight for cooking or defrosting.
- **Button 'Pause/Cancel'** - This button may function to pause or cancel the current operation.
- **Button 'Instant/Start'** - This button may function to start the cooking process immediately.

(b) Element Grounding

- **Question:** I want to *thaw some frozen meat*. Which control should I use for that?
- **Answer:** You should use the *button labelled "Defrost/Clock"* to activate the defrosting function. Therefore, click *button at [0.743, 0.466]*.



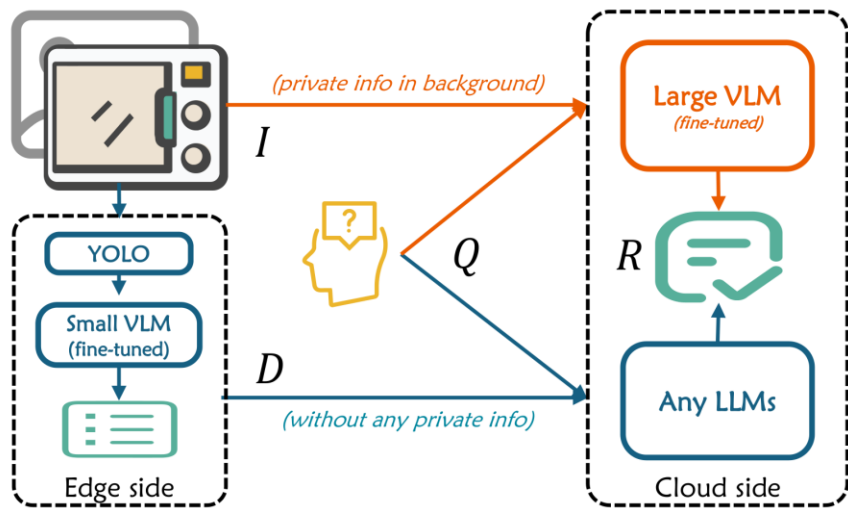
(c) Function Inference

- **Question:** If the microwave is heating food, what happens when pressing the *button located at [0.718, 0.708, 0.773, 0.767]*?
- **Answer:** When the microwave oven is heating food and you press this button, it will *pause the current heating operation*. This allows you to temporarily stop the process without cancelling the settings or turning off the oven.

(d) Multi-Step Operation Planning

- **Question:** You have some frozen food and you want to *defrost it first, then cook it using the Auto Cooking mode*. How would you go about setting this on your microwave oven?
- **Answer:** To defrost the frozen food and then cook it using the Auto Cooking mode on your microwave oven, follow these steps:
 1. *Press the 'Defrost/Clock' button (labelled with 'Defrost/Clock')* to activate the defrosting operation. This will set the microwave to defrost mode.
 2. *Using the 'Time - Weight +' knob* to adjust according to the weight of the food and ensure proper defrosting.
 3. *Press the 'Pause/Cancel' button (labelled with 'Pause/Cancel')* to stop the defrost operation once defrosting is complete.
 4. *Press the 'Auto Cooking' button (labelled with 'Auto Cooking')* to select the automatic cooking mode.
 5. *Press the 'Instant/Start' button (labelled with 'Instant/Start')* to start the automatic cooking mode.

Methods for PUO Tasks



End-to-End VLMs

$$R = \mathcal{V}(I, Q; \Theta_v + \Delta W)$$

Privacy-Preserving Framework

$$B = \{(c_i, b_i)\}_{i=1}^N$$
$$P = \mathcal{E}(I, B; \Theta_e),$$
$$R = \mathcal{L}(Q, P; \Theta_l)$$

Experiments: panel description

$$C_{\text{sim}} = 2\langle S, M \rangle / (N_{\text{pred}} + N_{\text{gt}})$$

- $S \in \mathbb{R}^{N_{\text{pred}} \times N_{\text{gt}}}$ with $S_{ij} = \text{Sim}(e_i, e'_j)$
- $M \in \{0, 1\}^{N_{\text{pred}} \times N_{\text{gt}}}$ is the optimal matching when performing Hungarian algorithm to S

$$\text{LJS-F1} = 2N_{\text{match}} / (N_{\text{pred}} + N_{\text{gt}})$$

- N_{match} is the LLM-judged count of semantically matched elements

$$\text{LJS-Acc} = \frac{1}{N} \sum_{i=1}^N \text{LLM}_{pt}(R_i, R'_i).$$

Table 2: Comparison on PUO benchmarks.

Method	FT	Description			Grounding		Function		Planning
		C _{CIDEr}	C _{BERT-S}	LJS-F1	Coord	Acc	BERT-S	LJS-Acc	LJS-Acc
GPT-4o	✗	35.40	0.769	0.545	0.992	0.063	0.900	0.297	0.294
Claude-3.7	✗	30.31	0.765	0.538	0.985	0.072	0.894	0.274	0.192
Qwen-VL-72B	✗	0.094	0.117	0.464	0.941	0.082	0.894	0.265	0.122
Yi-VL-6B [61]	✗	5.34	0.455	0.068	0.000	0.000	0.091	0.167	0.030
	✓	163.45	0.802	0.527	0.992	0.106	0.910	0.290	0.393
LLAVA-7B [42]	✗	13.09	0.627	0.146	0.000	0.000	0.869	0.124	0.049
	✓	248.93	0.879	0.749	0.995	0.498	<u>0.918</u>	0.463	0.490
Qwen-VL-7B [8]	✗	23.86	0.658	0.408	0.000	0.000	0.876	0.349	0.239
	✓	<u>238.23</u>	0.884	<u>0.748</u>	<u>0.995</u>	0.625	0.918	0.528	<u>0.444</u>
PPF (w/ GPT-4o)	✓	192.90	0.898	0.706	0.990	<u>0.673</u>	0.903	0.630	0.420
PPF (w/ Claude)	✓	189.72	0.898	0.706	0.990	0.676	0.906	<u>0.532</u>	0.335
PPF (w/ Qwen)	✓	191.01	<u>0.898</u>	0.709	0.997	0.668	0.909	0.510	0.315

Table 3: Ablation study on with and without detection results.

Method	YOLO	Description			Grounding		Function		Planning
		C _{CIDEr}	C _{BERT-S}	LJS-F1	Coord	Acc	BERT-S	LJS-Acc	LJS-Acc
PPF (w/ GPT-4o)	✗	156.01	0.854	0.641	0.989	0.515	0.897	0.456	0.333
	✓	192.90	0.898	0.706	0.990	0.673	0.903	0.567	0.420
Qwen-VL-7B [8]	✗	238.23	0.884	0.748	0.995	0.625	0.918	0.528	0.444
	✓	186.86	0.875	0.711	0.998	0.705	0.956	0.551	0.416

Visualization: edge-side VLM

Detector Result



VLM Enhancement

📍 A knob from $(0.348, 0.643)$ to $(0.679, 0.858)$.

🗣️ It has *'START + 30 SEC TIME WEIGHT PROG LEVEL'* on it.

⚙️ This is a knob that may be used to select different functions of the microwave oven, such as starting heating, increasing by 30 seconds each time, defrosting, setting the time, defrosting by weight, or adjusting the power level. Rotating this knob can select different options.

📍 button from $(0.341, 0.542)$ to $(0.384, 0.585)$.

🗣️ It has *'SPEED'* on it and looks like *fan* icon.

⚙️ Its possible function is to adjust the wind speed of the air purifier. After pressing it, the wind speed can be adjusted.

Visualization

Panel Description

- Button TEMP (with upward arrow symbol) - This button is used to increase the set temperature.
- Button TEMP (with downward arrow symbol) - This button is used to decrease the set temperature.
- Button MODE - This button is used to switch between different modes (AUTO, COOL, DRY, HEAT).
- Button ON/OFF - This button is used to power the air conditioner on or off.
- Button FAN SPEED - This button is used to adjust the fan speed.
- Button SWING - This button is used to control the swing function of the air direction.
- Button SLEEP - This button is used to activate or deactivate the sleep mode.
- Button TIMER ON - This button is used to set the timer to turn the air conditioner on at a specified time.
- Button TIMER OFF - This button is used to set the timer to turn the air conditioner off at a specified time.
- Button RESET (with lock symbol) - This button is used to reset the settings to their default values or to lock the current settings.
- Button AIR DIRECTION - This button is used to adjust the direction of the air flow.
- Button LED DISPLAY - This button is used to control the LED display.
- Button TURBO - This button is used to activate the turbo mode for faster cooling or heating.



GPT-4o

Panel Description

- Button TEMP (↑) - This button adjusts the air-conditioning temperature. Pressing it allows relevant operations on the temperature setting.
- Button ON/OFF (⏻) - This button controls turning the air conditioner on or off. Pressing it executes this operation.
- Button AIR DIRECTION (↻) - This button adjusts the air outlet direction of the air conditioner. Pressing it changes the outlet angle.
- Button TEMP (↓) - This button adjusts the air-conditioning temperature. Pressing it allows adjustment of the set temperature.
- Button LED DISPLAY (⏻) - This button controls the LED display screen of the air conditioner. Pressing it allows operations such as turning the display on/off or switching modes.
- Button TIMER ON (⌚) - This button turns on the air conditioner timer. Pressing it allows setting a scheduled run time.
- Button TIMER OFF (⌚) - This button sets the air conditioner to turn off at a scheduled time. Pressing it allows setting the turn-off time.
- Button FAN SPEED (⏻) - This button adjusts the air conditioner fan speed. Pressing it switches between different speed settings.
- Button SLEEP (🌙) - This button turns on the sleep mode of the air conditioner. Pressing it activates sleep mode.
- Button TURBO (⚡) - This button turns on the strong mode of the air conditioner. Pressing it operates the air conditioner at higher power.
- Button MODE (⏻) - This button switches the air conditioner mode. Pressing it cycles through different modes such as cooling, heating, dehumidifying, air supply, etc.
- Button SWING (↻) - This button controls the air swing function of the air conditioner. Pressing it starts the air direction swinging up and down or left and right.

PPF(ours)

Multi-Step Operation Planning

Question: I want to set the induction cooker to 350 F and cook for 25 minutes. How should I set this up and start cooking?

Answer:

1. Press the "POWER" button with the "On/Off" icon to turn on the device.
2. Press the "TEMP (F)" button with the thermometer icon to select the temperature mode.
3. Use the "+" or "-" buttons near the center of the panel to adjust the temperature to display "350".
4. Press the "TIMER (min)" button with the clock icon to select the timer mode.
5. Use the "+" or "-" buttons near the center of the panel to set the time to "25" minutes.
6. Press the "HEATING" button with the flame icon to start the cooking process.



GPT-4o

Multi-Step Operation Planning

Question: I want to set the induction cooker to 350 F and cook for 25 minutes. How should I set this up and start cooking?

Answer:

1. Press the "POWER ON/OFF" button (labeled with a power symbol) to turn on the induction cooker.
2. Press the 'a button with a thermometer icon' (for adjusting temperature) until the temperature setting function is selected.
3. Use the 'A plus sign (+) in a square button' (with a plus sign) to increase the temperature to 350 F.
4. Press the "Timer (min)" button (labeled "Timer (min)") to set the cooking duration.
5. Use the 'A plus sign (+) in a square button' (with a plus sign) to increase the timer to 25 minutes.
6. Finally, press the "HEATING" button (labeled with a flame icon) to start the heating and begin cooking.

PPF(ours)

Function Inference

Question: There is a button located at [0.79, 0.625, 0.831, 0.655]. Can you explain the dual functionality of this button?

Answer:

The button is labeled "STOP (Cancel)" and serves two functions: it stops the microwave during operation and cancels any current settings or operations.



GPT-4o

Function Inference

Question: There is a button located at [0.79, 0.625, 0.831, 0.655]. Can you explain the dual functionality of this button?

Answer:

This button located at [0.789, 0.625, 0.831, 0.656] has dual functionality:

1. **START:** When pressed, it starts the microwave oven's operation.
2. **+30Sec.:** Additionally, it can set a 30-second heating time, starting the microwave for that duration.

PPF(ours)

Element Grounding

Question: I want the rice cooker to start cooking in the morning while I'm away, so I can have fresh rice ready when I get home from work. Can I set a delayed start?

Answer:

You should press the button labeled "Delay Start" to set a delayed start for cooking. Therefore, click [0.33, 0.82]



GPT-4o

Element Grounding

Question: I want the rice cooker to start cooking in the morning while I'm away, so I can have fresh rice ready when I get home from work. Can I set a delayed start?

Answer:

You should press the "Delay Start" button to set the delayed start time for cooking rice. Therefore, click button at [0.483, 0.613].

PPF(ours)



Department of
Computer Science

香港城市大學
City University of Hong Kong

Thanks

**Wei Lin^{12*}, Yiwei Zhou^{13*}, Junkai Zhang^{14*}, Rui Shao^{15*}, Zhiyuan Zhao¹,
Junyu Gao^{15†}, Antoni B. Chan², Xuelong Li^{1†}**

¹The Institute of Artificial Intelligence, China Telecom (TeleAI)

²Dept. of Computer Science, City University of Hong Kong, ³Beijing Institute of Technology,

⁴Beijing University of Technology, ⁵Northwest Polytechnical University