Algorithm 1 Detect UI Tarpit

```
function DetectUITarpit(xml_1, xml_2, threshold)
    similarity ← CompareXML(xml_1, xml_2)
    if similarity > 90 then
        sim_count ← sim_count + 1
        if sim_count ≥ threshold then
            sim_count ← 0
            return True
        end if
    end if
    return False
end function
```

Algorithm 2 Compare XML

```
function CompareXML(xml_1, xml_2)
    tree_1, tree_2 ← Simplify xml_1, xml_2 and construct trees
    score, total ← CompareTree(tree_1, tree_2)
    return 100.0 if total = 0 else (score/total) × 100
end function
```

Algorithm 3 Main Exploration Loop

```
function START(input_manager)
    count ← 0
    while count < max_event_count do
        Update UI state and snapshots
        Start the APP if essential
        if LLM mode is active then
            event ← GENERATELLMEVENT
        else if DETECTUITARPIT(last_state, current_state) then
            Activate LLM Mode
            event ← GENERATELLMEVENT
        else
            event ← GENERATERANDOMEVENT
        end if
        EXECUTE(event)
        count ← count + 1
    end while
    Clean up and exit
end function
```

Algorithm 4 Generate LLM Event

```
function GENERATELLMEVENT
    if Continuing LLM Sequence then
        Build Next Action Prompt
        response ← CALLLLM
        response ← VALIDITEBYLLM
        act ← PARSEACTION(response)
    else
        Build Meaning Prompt
        r_1 ← CALLLLM
        Build Task Prompt
        r_2 ← CALLLLM
        Build First Action Prompt
        r_3 ← CALLLLM
        response ← VALIDITEBYLLM
        act ← PARSEACTION(response)
    end if
    Set LLM Mode to act.hasNext
    return WRAPASU2EVENT(act)
end function
```

Algorithm 5 Frequency-Aware Random Exploration Strategy

```
function GenerateEvent
    s ← current state
    if s ∉ input_table then
        possible_events ← GetPossibleInputs(s)
        Initialize input_table[s] with an empty events list
        for all event ∈ possible_events do
            Add event to input_table[s].events
            if event ∉ event_table then
                event_table[event] ← 0
            end if
        end for
    end if
    counts ← ∅
    for all event ∈ input_table[s].events do
        counts[event] ← event_table[event].tried
    end for
    weights ← GetWeights(input_table[s].events, counts)
    selected_event ← randomly select one event from the list using weights
    Increment event_table[selected_event].tried
    return selected_event
end function
```

		Time (%)			Event (%)		
		Max	Min	Avg	Max	Min	Avg
AnkiDroid	`llm`	6.56	-43.83	-18.69	17.63	-43.83	-10.45
	`new`	51.32	2.72	32.22	55.44	6.91	36.28
Feeder	`llm`	6.88	-22.11	-15.54	10.59	-17.11	-11.76
	`new`	29.81	1.02	15.87	33.25	4.99	22.54
Newpipe	`llm`	121.86	22.63	32.71	144.69	28.24	41.09
	`new`	140.29	29.24	44.10	140.29	42.33	51.83
Omninotes	`llm`	-13.73	-50.08	-39.31	-13.73	-47.34	-35.73
	`new`	16.12	-1.65	3.55	16.12	0.19	9.44
Wikipedia	`llm`	6.68	-39.63	-15.34	34.74	-30.24	-0.58
	`new`	14.57	-14.38	1.97	28.24	-10.54	17.27

Table 1: Improvement on branch coverage

		Time (%)			Event (%)		
		Max	Min	Avg	Max	Min	Avg
AnkiDroid	`llm`	6.56	-43.83	-18.69	17.63	-43.83	-10.45
	`new`	51.32	2.72	32.22	55.44	6.91	36.28
Feeder	`llm`	6.88	-22.11	-15.54	10.59	-17.11	-11.76
	`new`	29.81	1.02	15.87	33.25	4.99	22.54
Newpipe	`llm`	121.86	22.63	32.71	144.69	28.24	41.09
	`new`	140.29	29.24	44.10	140.29	42.33	51.83
Omninotes	`llm`	-13.73	-50.08	-39.31	-13.73	-47.34	-35.73
	`new`	16.12	-1.65	3.55	16.12	0.19	9.44
Wikipedia	`llm`	6.68	-39.63	-15.34	34.74	-30.24	-0.58
	`new`	14.57	-14.38	1.97	28.24	-10.54	17.27

Table 2: Improvement on branch coverage

App	**Policy**	**Time (%)**			**Event (%)**		
		Max	Min	Avg	Max	Min	Avg
AnkiDroid	`llm`	6.56	-43.83	-18.69	17.63	-43.83	-10.45
	`new`	51.32	2.72	32.22	55.44	6.91	36.28
Feeder	`llm`	6.88	-22.11	-15.54	10.59	-17.11	-11.76
	`new`	29.81	1.02	15.87	33.25	4.99	22.54
Newpipe	`llm`	121.86	22.63	32.71	144.69	28.24	41.09
	`new`	140.29	29.24	44.10	140.29	42.33	51.83
Omninotes	`llm`	-13.73	-50.08	-39.31	-13.73	-47.34	-35.73
	`new`	16.12	-1.65	3.55	16.12	0.19	9.44
Wikipedia	`llm`	6.68	-39.63	-15.34	34.74	-30.24	-0.58
	`new`	14.57	-14.38	1.97	28.24	-10.54	17.27

Table 3: Improvement on branch coverage