

ExFrame Test Plan (GUI)

Date: 2026-02-12

URL: http://localhost:3000

Method: Manual testing through the web interface

Run each test, mark PASS/FAIL, and note any issues.

1. Role Context (Critical)

The `role_context` in `domain.json` controls AI behavior. If this breaks, the AI ignores domain instructions.

Test 1.1 — Peter journal entry (no ****** prefix)

- **Domain:** peter
- **Input:** Pick up milk on the way home
- **Expected:** A timestamped echo like [2026-02-13 HH:MM:SS] Pick up milk on the way home
- **FAIL if:** The AI gives advice about milk, asks questions, or generates anything beyond a timestamp + echo
- **Result:** []

Test 1.2 — Peter question (****** prefix)

- **Domain:** peter
- **Input:** ** What did I say about a cake?
- **Expected:** A response referencing the cherry chocolate cake journal entry, drawn from the domain log
- **FAIL if:** The AI gives a generic answer about cakes not based on journal entries
- **Result:** []

Test 1.3 — Peter timestamp year

- **Domain:** peter
- **Input:** Dentist appointment next Tuesday
- **Expected:** Timestamp shows year 2026 (not 2024 or 2025)
- **FAIL if:** Wrong year in the timestamp
- **Result:** []

Test 1.4 — Domain without `role_context`

- **Domain:** cooking (or any domain without a `role_context` field in its `domain.json`)
 - **Input:** How do I make scrambled eggs?
 - **Expected:** A normal helpful response (falls back to "You are a helpful assistant")
 - **FAIL if:** Error, crash, or empty response
 - **Result:** []
-

2. Conversation Memory Modes

Test 2.1 — Question mode skips memory for plain entries

- **Domain:** peter
- **Input:** Bought new running shoes today
- **Expected:** Fast response (under 15 seconds). Timestamped echo only.
- **Note:** Check the Traces tab — should NOT show "Loaded conversation memory"
- **Result:** []

Test 2.2 — Question mode loads memory for ** queries

- **Domain:** peter
 - **Input:** ** What have I been up to recently?
 - **Expected:** Response references recent journal entries (shoes, milk, dentist, etc.)
 - **Note:** Check the Traces tab — should show conversation memory was loaded
 - **Result:** []
-

3. Persona Types

Test 3.1 — Poet persona (void data source)

- **Domain:** poetry_domain
- **Input:** Write me a short poem about rain
- **Expected:** A generated poem. No web search, no document references.
- **Result:** []

Test 3.2 — Librarian persona (library data source)

- **Domain:** exframe
- **Input:** How do personas work in ExFrame?
- **Expected:** Response synthesized from local documents. May reference ExFrame architecture.
- **FAIL if:** Response is clearly generic with no domain-specific knowledge
- **Result:** []

Test 3.3 — Researcher persona (internet data source)

- **Domain:** cooking
 - **Input:** What is a good recipe for banana bread?
 - **Expected:** Response with web search results. Should show source URLs at the bottom.
 - **FAIL if:** No sources shown, or response is clearly from training data only
 - **Result:** []
-

4. Pattern Override

Test 4.1 — Domain with patterns returns pattern-based answer

- **Domain:** cooking
- **Input:** knife skills (or a query matching a known pattern)
- **Expected:** Response draws from the stored pattern content (Knife Skills: Basic Cuts)
- **Note:** Check response metadata — should indicate pattern override was used

- **Result:** []

Test 4.2 — Domain with no patterns falls through to persona

- **Domain:** peter (has no patterns)
 - **Input:** Bought concert tickets
 - **Expected:** Timestamped echo (poet persona, no pattern override)
 - **Result:** []
-

5. Logging

Test 5.1 — Journal entry appears in domain log

- **Domain:** peter
- **Input:** Testing the domain log
- **Expected:** After response, go to the domain editor/log viewer. The entry should appear with timestamp.
- **FAIL if:** Entry is missing from domain_log.md
- **Result:** []

Test 5.2 — Web search response logged with timestamp

- **Domain:** cooking
 - **Input:** How to cook perfect rice?
 - **Expected:** In the domain log, the entry should be tagged with [WEB_SEARCH - YYYY-MM-DD HH:MM:SS]
 - **Result:** []
-

6. Trace Data

Test 6.1 — Trace tab shows Phase 1 queries

- Run any query in any domain
- **Navigate to:** Traces tab
- **Expected:** Recent query appears with timing info (LLM call duration in ms)
- **FAIL if:** Traces tab is empty or doesn't show Phase 1 queries
- **Result:** []

Test 6.2 — Trace shows LLM timing breakdown

- Click on a trace entry
 - **Expected:** Shows "Step 1: LLM Call (XXXms)" and "Step 2: LLM Response"
 - **Result:** []
-

7. Domain Admin

Test 7.1 — View domain config

- Go to domain editor for peter

- **Expected:** Shows role_context field, persona type, conversation memory settings
- **Result:** []

Test 7.2 — Edit role_context via UI

- Go to domain editor for any test domain
- Change the role_context text
- Save
- Run a query
- **Expected:** AI behavior reflects the updated role_context
- **Caution:** Change it back after testing
- **Result:** []

Test 7.3 — Switch domains

- Query peter domain, then switch to cooking domain, then back
 - **Expected:** Each domain uses its own persona and role_context. No bleed between domains.
 - **Result:** []
-

8. Edge Cases

Test 8.1 — Empty query

- **Domain:** peter
- **Input:** (empty string or just spaces)
- **Expected:** Graceful handling — either a short response or an error message, not a crash
- **Result:** []

Test 8.2 — Very long query

- **Domain:** peter
- **Input:** A paragraph of 500+ characters
- **Expected:** Timestamped echo of the full text (not truncated)
- **Result:** []

Test 8.3 — Special characters in query

- **Domain:** peter
 - **Input:** Meeting at O'Brien's cafe — \$50 budget & "formal" dress code
 - **Expected:** Timestamped echo with all special characters preserved
 - **Result:** []
-

9. Container Restart Resilience

Test 9.1 — Code changes require restart

- Make any code change to a .py file
- Query the system WITHOUT restarting

- **Expected:** Old behavior (Python caches modules in memory)
- Restart container: `docker restart eeiframe-app`
- Query again
- **Expected:** New behavior
- **Result:** []

Test 9.2 — Config changes are live (no restart needed)

- Edit a domain.json file (e.g., change temperature)
 - Query the domain WITHOUT restarting
 - **Expected:** New config is picked up (domain.json is read per-request)
 - **Result:** []
-

Summary

Area	Tests	Pass	Fail
Role Context	4		
Conversation Memory	2		
Persona Types	3		
Pattern Override	2		
Logging	2		
Trace Data	2		
Domain Admin	3		
Edge Cases	3		
Container Restart	2		
Total	23		