ForgeFlow ML — TestSprite / Cursor Full-System Test Plan

Purpose: End-to-end verification of the ForgeFlow ML application (Insight Hub + Dashboard). This document is a single authoritative prompt for use in Cursor or TestSprite to validate functionality, security, and readiness for production.

Test Preflight (replace placeholders before running)

Parameter	Value
BASE_URL	e.g., https://staging.forgeflow.example
HOST	e.g., staging.forgeflow.example
USERNAME / PASSWORD	Test user with project:write permission
DATASET_PATH	Path to dataset (use customer_churn.csv supplied)

Primary Test Prompt (copy into Cursor / TestSprite):

Execute a full end-to-end test of ForgeFlow ML against the target environment. Follow the numbered steps below

1) HEALTH & AUTHENTICATION - Verify GET {{BASE_URL}}/health returns 200 and payload {{"status":"ok"}}. - Log in using provided credentials; confirm dashboard loads.

2) INSIGHT HUB (IN-PLACE MODE) - From dashboard, click the Insight Hub control. VERIFY: no navigation occurs; top controls (Problem Characterization, Model selection, Data splitting, Start Training, Live Feed, Test Model) collapse/hide with smooth animation. - New heading 'Insight Hub' must appear and the prompt bar must auto-focus.

3) UPLOAD & PROMPT-DRIVEN RUN - Upload dataset at {{DATASET_PATH}} (customer_churn.csv). Confirm dataset preview shows first rows and inferred types. - In prompt box paste the following exact prompt: "Train a binary classifier to predict customer churn. Primary metric: F1-score. Prioritize recall. Limit model size <50MB. Export ONNX. Provide feature importance and confusion matrix." - Click Run. If AI consent/cost modal appears, accept and confirm. - Confirm API returns a task_id and status 'queued'.

4) LIVE MONITORING & WEBSOCKET - Subscribe to ws://{{HOST}}/ws/tasks/{{task_id}}. Expect streaming messages of types: metric, log, checkpoint, done. - Ensure Live Training Metrics panel updates and chart displays train/validation lines. Expect at least one metric message per epoch.

5) ARTIFACTS & EXPORT - On completion (done), verify artifact record exists via GET /api/artifacts and contains format ONNX. - Trigger Export: complete payment (■15) or simulate payment; then download artifact. Verify SHA256 checksum matches recorded checksum.

6) TESTING (IF SELECTED) - If user accepts testing, trigger the automated QA suite and verify a report with robustness, calibration, and fairness results appears.

7) RESILIENCE & RECOVERY - Simulate a page reload mid-run; verify UI recovers state using GET /api/tasks/{{task_id}}/events/replay. - Simulate worker failure; verify graceful retry or clear error in audit and UI.

8) SECURITY & AUDIT - Confirm PII detection warning appears when expected. - Confirm telemetry logs do not store raw prompts (only prompt hash) and that prompts/responses are stored encrypted in audit log.

Reporting: produce a test report with screenshots at key steps, first 20 WS messages, task_id, artifact_id, payment evidence, pass/fail per checkpoint, and attach logs.

Step-by-step Checkpoints (for automation & verification)

Checkpoint	Expected Result
Preflight	Health endpoint OK; login successful; dataset accessible.
Insight Hub entry	Click triggers in-place mode; top controls hidden; prompt bar focused.
Dataset upload	Preview shows rows; column types correct; PII flagged if present.
Prompt validation	Prompt meets min length and passes forbidden patterns.
Job submission	POST /api/projects/{project_id}/insight-hub/run returns task_id.
WebSocket streaming	Metrics/logs/checkpoints stream; UI chart updates.
Completion & artifact	Artifact record exists; export gated by payment; download checksum veri
Post-test QA	QA report generated with expected sections.
Audit logs	All state changes recorded with user & timestamp.
Recovery	Replay API restores state after reload.

API Endpoints Quick Reference

Endpoint	Method	Purpose
/health	GET	Service health check
/api/projects/{project_id}/insight-hub/run	POST	Submit prompt-driven training job; returns task_
/ws/tasks/{task_id}	WS	Stream metrics/logs/checkpoints/events
/api/tasks/{task_id}/events/replay	GET	Replay recent events for reconnection
/api/artifacts/{artifact_id}/meta	GET	Artifact metadata
/api/artifacts/{artifact_id}/pay	POST	Payment verification for export
/api/artifacts/{artifact_id}/download	GET	Gated download (after payment)

WebSocket Message Schema (examples)

metric: { 'type':'metric', 'payload': { 'epoch':int, 'train_loss':float, 'val_auc':float, 'time_ms':int } }

checkpoint: { 'type':'checkpoint', 'payload': { 'checkpoint_id':str, 'epoch':int, 'metric_value':float, 'artifact_uri':str } }

log: { 'type':'log', 'payload': { 'level':'info|warn|error', 'message':str } }

done: { 'type':'done', 'payload': { 'task_id':str, 'artifact':{ 'name':str, 'url':str }, 'summary':{ 'primary_metric':float } } }

Evidence & Reporting Checklist

• Screenshots: Dashboard before click, Insight Hub prompt, Upload preview, Live chart during training, Training completion modal, Export payment, Download success.

• WS capture: first 20 messages saved as JSON.

• API logs: task_id creation response, artifact meta record, payment confirmation.

• Checksums: downloaded artifact checksum matches DB record.

• Audit entries: record of actions (upload, run, export).

Acceptance Criteria (must all pass)

• End-to-end pipeline completes and artifact downloadable with verified checksum.

• Live metrics stream continuously during the run and UI updates accordingly.

• Insight Hub operates in-place without navigation and restores on exit.

• Audit logs capture all state-changing operations.

• PII handling and consent flows are enforced before external exports.

How to use in Cursor / TestSprite

Paste the 'Primary Test Prompt' block into Cursor or TestSprite as a single script. Replace placeholders with real values. Configure the run to collect screenshots and attach any captured logs/WS messages. Submit the run and review the produced artifacts.

Prepared by: ForgeFlow QA • Generated for Cursor / TestSprite automated testing

Date: 2025-09-20