

## Friendly Testing Plan

### 1. The test data

- **\*\*Simulated trouble:\*\*** We script fake power drops, battery leaks, and sneaky slow declines.
- **\*\*Old mission tapes:\*\*** We replay real telemetry from past satellites that limped or failed.

### 2. What we measure

- **\*\*Did we notice in time?\*\*** (Time to Detection)
- **\*\*How often were we right?\*\*** (Precision & Recall)
- **\*\*Did we cry wolf?\*\*** (False alarm rate)

### 3. Scenarios to run

- Plain-vanilla good days to be sure we stay quiet.
- Sudden “battery unplugged” events.
- Long, boring, slow failures that creep up over weeks.
- Throw in random noise just to be mean.

### 4. Validation style

We slice the timeline train on the early months, test on later months. Then we shuffle and repeat. A red-team crew can feed in bizarre edge cases to keep us honest.

### 5. Humans in the loop

Stage 1: Every alert pops up on an engineer’s screen for thumbs-up / thumbs-down.

Stage 2: Once trust is high, we only ask for help when the AI is not sure.