

CSM Copilot

Move onboarding forward.
Keep the evidence intact.

Independent portfolio project. Not affiliated with Unity. All customer records are synthetic.

5 synthetic accounts / 8 MCP tools / 18 recorded model attempts
English and Chinese follow-ups | Public evidence snapshot: 14 October 2026

The customer problem

A publisher wants to launch on Friday. An iOS test passed, but a non-test Android device shows a blank Offerwall. The CSM needs a focused follow-up and a useful technical handoff without turning incomplete information into a promise.

The working sequence

Prioritize > inspect evidence > draft > hand off > review
The MCP server supplies scoped findings and drafts. A model-enabled host can choose tools. The website replays actual engine outputs and recorded model runs; it is not a live chat.

One account. Two different next actions.

Integration	Recorded evidence	CSM next action
Moon Garden Android	3 unknown checks; 1 failed device check. Separate settings evidence from a fresh test.	Request missing settings/callback evidence and a new non-test-device result. Keep launch unresolved.
Moon Garden iOS	6 scoped checks met. The records belong to iOS, not Android.	Confirm changes and review the launch plan. Checklist completion is not launch approval.

Trace the Android failure to `lantern_android_device_verified`. Two new setting confirmations still do not replace that failed test.

What a reviewer can inspect

Bilingual drafts, explicit missing handoff fields, platform-specific evidence, a Product feedback digest, and a practice exercise that preserves initial reasoning separately from reflection after guidance.

Contribution and learning
Yuchi selected the project direction and CSM workflow. Implementation, synthetic cases, tests and documentation used substantial AI assistance. Personal mastery exercises and independent human validation are not claimed as completed.

[Open the interactive portfolio](#) | [View code and evidence](#)

AI behavior you can inspect

Correct tool use does not guarantee a faithful explanation. The project preserves the failed outputs as well as the changes made in response.

Observed step	What happened	Design response
1. Tool calls succeed	Qwen3-4B executed 4/4 requested calls, then described iOS as having 'full compliance' and omitted citations.	Separate tool coverage from answer fidelity. Preserve the raw output and compare claims with sources.
2. Format is constrained	Parseable JSON still repeated platforms and included an issue-only source as readiness evidence.	Check scope, unknowns, failures and readiness IDs against actual tool results. Reject mismatches.
3. Fields finally match	After 2 recorded correction requests, the structured fields matched. The explanation still lacked inline source IDs and an explicit nothing-sent sentence.	Render a separate summary from checked tool fields. Keep free prose visibly unverified.

Result: assisted recovery on one task, not a model accuracy score.

Prompts, response contracts and generation settings changed across attempts. Correction feedback supplied expected values. Field agreement does not verify source truth or free prose. The source comparison is AI-assisted, not independent human review.

What the checks establish

Real MCP transport, repeatable workflow outputs and tested evidence boundaries. The lexical triage diagnostic matched 16/18 reused synthetic messages after adapting rules to inspected failures (earlier: 14/18). No production adoption, measured time savings or general reliability is claimed.

How to demonstrate the learning

Explain one decision before opening the guidance. Trace it to a source. Identify what would change your conclusion. Then make and verify a small improvement you can explain yourself. The practice download keeps original reasoning and later reflection distinct; it does not grade them.

Open the evidence

[Android case](#) | [Chinese iOS follow-up](#) | [Practice](#)

[Recorded corrections and review](#) | [Verification record](#)

Source basis: evidence/workflow.json; evidence/model-runs/qwen4-batch.json; evidence/model-runs/qwen4-constrained.json; evidence/evaluation.json. Links open the current public project. The evidence fingerprint in the footer identifies the source bundle used here.