

FROM RECORDS TO A REVIEWED BRIEF.

A case-manager assistant for post-discharge outreach. One generation, inspectable evidence, and approval tied to the version reviewed.

18 evaluated generations / 3 cases / 22 automated tests passing

THE WORKFLOW PROBLEM

Preparing a post-discharge call may require reconciling a discharge summary, an earlier case-management note, and an outreach log. The hypothesis: a checkable brief can make preparation easier. Usefulness remains untested with case managers.

A COMPLETE FIRST FLOW

Choose a member, inspect three records, generate, open evidence, edit, approve, and save. Organize the brief around the documented situation, follow-up tasks, questions, conflicts, and gaps.

PRODUCT JUDGMENT

Preserve uncertainty. Missing documentation does not prove an event never happened. Show both conflicting accounts.

Make review manageable. Unchecked filters, next-item navigation, collapsible sections, and progress help reviewers move through the full checklist. Undo restores an item as unchecked.

Approve a specific version. Edits invalidate approval. Saving retains original and edited text, evidence, and generation metadata.

ASSUMPTIONS TO TEST

The available records contain useful context; reviewers can access sources and have time and authority to resolve uncertainty. These assumptions need observation in actual preparation work.

THE SMALLEST USEFUL BUILD

React/Vite and Node; one OpenAI Responses API call with all three records. Code resolves passage IDs to original quotations. Local JSON stores approved briefs. No retrieval service or agent framework was needed.

DEFINE SUCCESS BEFORE GENERATION

Freeze expected facts, discrepancies, and gaps separately from model inputs. Run three cases twice per iteration. Preserve every original output, including failures. Assess coverage and semantic support separately from reference validity.

THE FAILURE THAT MATTERS

Structural acceptance improved from **4/6 to 4/6 to 6/6**. Yet the final Theo brief adds a **9:15 AM call time** that its cited passage does not document. That is the log timestamp. Valid citations still need human judgment.

FINDINGS AND LIMITS

Final full coverage: facts **29/36**, discrepancies **5/6**, gaps **16/30**; partial coverage is excluded. Two outputs exceeded 350 words. Model, prompt, and schema changed together, so this was not a controlled comparison. Review was AI-assisted.

WHAT I WOULD LEARN NEXT

Observe which omissions change a call, which sources staff trust, and how they resolve disagreements. Add independently authored cases and human adjudication. Shared evidence, evaluation, and approval services become platform candidates when needs repeat.