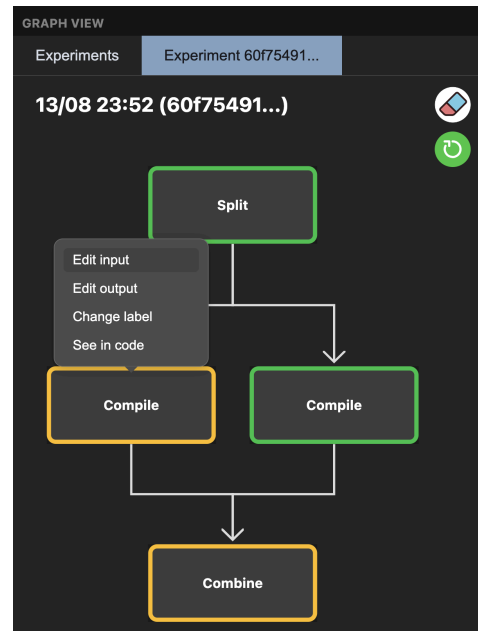


Towards more effective agent workflows

Agent workflows dramatically boost AI performance by chaining together several LLM calls and tool invocations. Their effectiveness hinges on two key factors: **workflow design** and the quality of its **building blocks** (e.g., LLMs, RAG models, etc.).

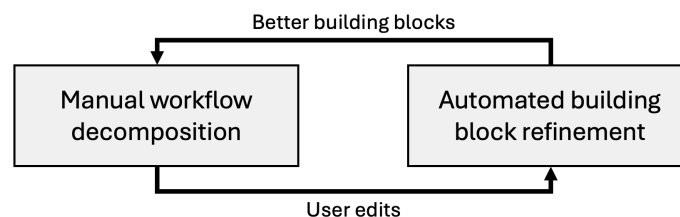
Designing effective workflows often involves a tedious trial and error process, where the user iterates over the many possible ways to decompose a complex task and integrate helpful tools. For example, an agent that converts design documents into runnable code might only involve a single "compile" step, or it may involve steps like "rewrite design doc", "gather relevant context", "write pseudo-code", etc. Orthogonally, optimizing each of these building blocks involves "lower level" techniques like fine-tuning the right models, enhancing RAG systems, or implementing common agent techniques like verification loops.



In this project, we aim to enhance both axes of agent effectiveness and help developers build *better* workflows, *faster*. First, we introduce tools that accelerate the developer's trial-and-error process by allowing them to quickly understand where errors occur, and how they can be fixed. Specifically, we allow the developer to ask *what-if questions* over failed workflow runs (e.g., for a given design doc, what would have happened if the pseudo-code of function X was more detailed?). They do this by editing inputs and outputs in a dataflow graph that we construct as their code is executing (see screenshot on the right).

Then, we learn from the developer's interactions with the tool, as well as their code changes, to understand what they *think* are improvements to the workflow and where they see weaknesses. Akin to weakly-supervised learning problems, the developer's tool interactions helps us tune the workflow's building blocks by, e.g., adding an interactive refinement mechanism to the pseudo-coder so it produces more detailed code.

In summary, our tool can be broken down into a manual and automated component, which each inform each other:



We're looking for UROPs and MEnges who are eager to solve engineering and research challenges with the goal of improving the effectiveness of agent workflows.