Data Collection Evaluation Task 1: Refactor the... Task 2: Prover that... instruction gen .diff repair & apply ➤ Lean compile gold .diff LLM judge ./.lake ./.Mathlib worktree