

Solving Open Problems in Operations Research using AI

Eric Fithian Rad Niazadeh Pranav Nuti

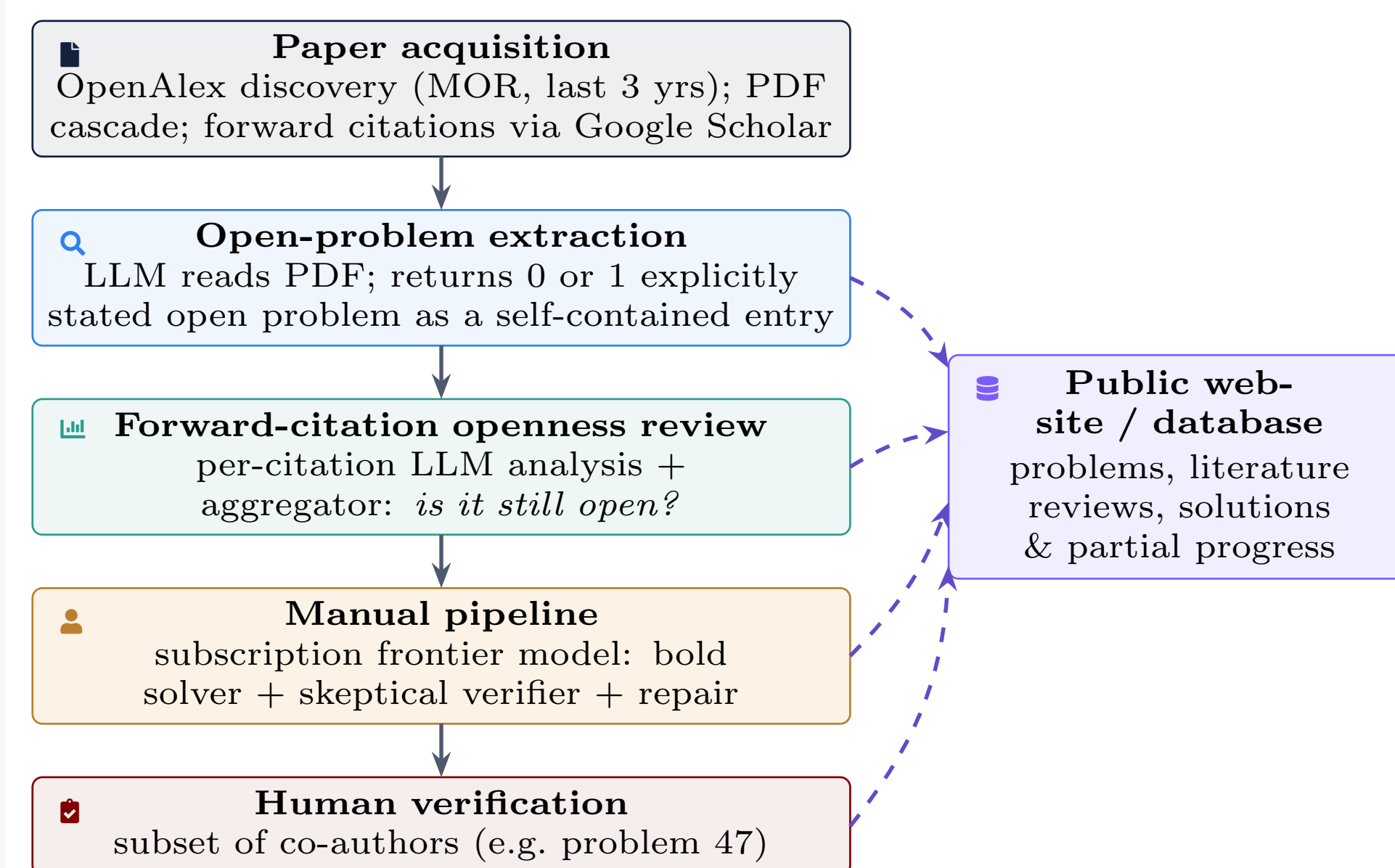
University of Chicago, Booth School of Business — Can AI do Theory? STOC TheoryFest 2026



Motivation

- ▶ LLMs have displayed improved abilities at solving tightly-scoped mathematical problems, particularly problems involving counterexample discovery.
- ▶ Most successes have been in pure mathematics, from researchers experimenting with AI tools on problems they are personally interested in. We ask:
How powerful are frontier models at solving open problems in operations research, and can we build a pipeline that solves such problems at scale?
- ▶ Attempts by researchers trying out AI tools may have an ephemeral nature. If the attempt produces an insufficiently interesting result, it is lost. Storing these attempts might still be useful for authors. We want to:
Create a database of problems, literature reviews, and AI attempts at problems.

Pipeline



Current Results

- ▶ We extracted and published **132 open problems** from the last 3 years of the journal *Mathematics of Operations Research*.
- ▶ We generated **105 literature reviews**.
- ▶ We generated solution attempts for all problems using a solver-verifier loop, and published these attempts. We have **26 claimed solutions**.
- ▶ We personally verified the solution to **problem 47**.

Problem 47: Case Study

- ▶ Problem 47 is about fair and efficient multi-resource allocation with Leontief utilities.
- ▶ There are n agents and three divisible resources, each with unit supply. Agent i has a normalized demand vector $\mathbf{d}_i \in (0, 1]^3$ with $\max_r d_{ir} = 1$, and $v_d(X) = \min_r X_r / d_r$.
- ▶ Bei, Li, and Luo introduced a fair-ratio version of the price of strategyproofness for social welfare:

$$\text{PoSP}_{SW}(3) = \inf_{f \text{ satisfies SI, EF, PO, SP}} \sup_I \frac{\text{OPT}_{\text{fair}}(I)}{SW(f(I))}.$$

and left open its exact value, proving only

$$2 \leq \text{PoSP}_{SW}(3) \leq 3.$$

We prove that the upper bound is tight:

$$\text{PoSP}_{SW}(3) = 3.$$

- ▶ In retrospect, this entry is a natural target for AI-assisted work because the missing result was plausibly reducible to the discovery of a carefully designed hard instance.

Some Lessons Learned

- ▶ **Open problem extraction:** A significant fraction of open problems in papers are vague, and it is hard to automatically say whether they have been solved.
- ▶ **Openness review:** Off-the-shelf literature review agents do not perform better than simply checking every forward-citation of a paper.
- ▶ **Solving pipeline:** Frontier models accessible through subscription perform well with simple prompting.
- ▶ **Human verification:** Human verification is time consuming. AI attempts can be stored with the view that they might eventually be checked if they are used in a result a human cares about.

Future Work

- ▶ Expanding to other sources of papers, with a focus on more concrete open problems.
- ▶ Strengthening the solving pipeline, particularly by attaching a persistent *memory file* to every problem, and recording promising leads and dead ends accumulated across runs. Adding a dedicated memory-management agent to curate this record and decide what information to provide each solver.
- ▶ Formally verifying solutions.
- ▶ Making the pipeline itself available on the website, so researchers can try solving their own problems with their own prompts and models.

We think we ought to build shared databases of problems, counterexamples, and partial results that reduce duplicated use of AI and can speed up automated mathematical research in our field.