

Presenter: Yuan Sui

Poster title: Building an AI Data Technician: Collaborative Agentic Program Synthesis for Scientific Data Processing

Abstract: Scientific progress is often bottlenecked not by model capacity, but by transforming raw measurements into usable data. We introduce an AI data technician: a collaborative agent that converts sparse expert feedback into interpretable, executable programs for scientific data processing. We formulate this as a language-guided program synthesis problem, where experts provide lightweight feedback on small samples and the system induces structured rules that generalize across datasets. Our approach combines language models with evolutionary search to generate compact, human-auditable programs that formalize expert heuristics. As a case study, we apply this framework to neural recording data cleaning, where artifact rejection is subjective and labor-intensive. This work illustrates a scalable, interpretable paradigm for scientific data processing, where expert knowledge is captured as an executable structure rather than static labels.