# Process-Traceable Student Agents: Regulating Teachable Behavior of LLM-based Learners with Model-Tracing

**Research Overview & Research questions:**

This research is situated at the intersection of generative AI and structured cognitive modeling. It proposes to formalize, implement, and evaluate a novel framework that synthesizes CoT reasoning, Model-Tracing constraints, and supervised fine-tuning to create Process-Traceable Student Agents. This methodology involves a two-phase process. The first phase leverages the integrated CoT and Model-Tracing framework as a "data factory" to generate a high-quality corpus of structured, cognitively plausible error paths. The second phase then uses this corpus to fine-tune a base LLM. This approach is expected to become a robust paradigm for addressing the cognitive degradation found in free-form generation. It is also intended to resolve the instability of prompt-based methods, thereby fostering a specialized agent with inherent pedagogical alignment. The ultimate goal is to achieve process-traceable, cognitively authentic, and instructionally valuable simulations. These simulations aim to set a new standard for educational AI in enhancing teacher development and personalized learning. To systematically achieve this goal and validate the proposed framework, this study is structured around three pivotal research questions.

RQ1: How can Chain-of-Thought and Model-Tracing be systematically combined to generate process-traceable and cognitively plausible student reasoning paths?

RQ2: Do the fine-tuned agents generate problem-solving trajectories with higher structural similarity to expert paths and realism in error manifestation than prompt-based agents?

RQ3: Do the process-traceable error chains generated by the agents provide enhanced cognitive interpretability that supports accurate diagnosis of knowledge gaps and strategy misuse?