

Cognitive Agent Compilation for Explicit Problem Solver Modeling

Hyeongdon (Doni) Moon
Carnegie Mellon University
donim@andrew.cmu.edu

John Stamper
Carnegie Mellon University
jstamper@andrew.cmu.edu

Carolyn Rose
Carnegie Mellon University
cp3a@andrew.cmu.edu

ABSTRACT

Large language models (LLMs) are widely used for tutoring, feedback generation, and content creation, but their broad pretraining makes them hard to constrain and poor substitutes for controllable learners. Educational systems often require inspectable and editable knowledge states: educators want to know what a system assumes the learner knows, and learners benefit when the system can justify actions in terms of explicit skills, misconceptions, and strategies. Inspired by cognitive architectures, we propose Cognitive Agent Compilation (CAC), a framework that uses a strong teacher LLM to compile problem-solving knowledge into an explicit target agent. CAC separates (i) knowledge representation, (ii) problem-solving policy, and (iii) verification and update rules, with the goal of making bounded problem solving more inspectable and editable in educational settings. We present an early proof of concept implemented with Small Language Models that surfaces key design trade-offs, particularly between explicit control and scalable generalization, and positions CAC as an initial step toward bounded-knowledge AI for educational applications.

Keywords

Cognitive Architecture, Knowledge Representation, Learner models, Explainable AI, Intelligent Tutoring systems

Hyeongdon Moon, John Stamper, and Carolyn Rosé. Cognitive Agent Compilation for Explicit Problem Solver Modeling. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyeonjeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 835–835. International Educational Data Mining Society (2026).

© 2026 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.
<https://doi.org/10.5281/zenodo.21040042>