# <u>Research Statement</u>

My research focuses on developing **neurosymbolic AI systems** that can **learn and teach like humans do**. My work takes a computational approach to studying human learning, aimed at building hybrid machine learning and generative AI systems that can **efficiently and flexibly learn from the rich forms of natural instruction** that arise in human-to-human tutoring. During my PhD at **Carnegie Mellon University** and postdoctoral research at **Georgia Tech**, I developed simulated learners that can induce and refine **well-formed, reliable knowledge structures** with as few training interactions as human learners (~12 or fewer) [4]. My simulated learners can learn from diverse forms of natural instruction, such as worked examples, correctness feedback, and natural language instruction. In my theoretical work, I have used these AI systems as **computational models of learning** [1,9] to precisely investigate how distinct learning mechanisms cooperatively manifest humans' rapid learning and misconception formation. I have also investigated methods for testing and refining these computational theories in comparisons of simulated learners with student data [15]. In practical applications, I have used these **teachable AI systems** as agents that can learn tutoring behavior through bottom-up induction from human instruction, and reliably deliver that adaptive instruction at scale in the form of **intelligent tutoring systems** [5,6,14].

## <u>Thrust 1: Simulated Learners as Computational Theories of Learning</u>

Learning engineering has a theory problem. As founders of the field have argued [20], the learning sciences have produced broad theories of knowledge [19] and pedagogy [18], but lack causal, mechanistic theories of learning for making relative predictions about how variations in instructional methodology affect students' knowledge formation and learning outcomes. While an aeronautic engineer can draw on theories of lift to precisely inform airplane design, an instructional engineer proceeds mostly by guesswork. Traditional statistical educational data mining uses student performance data to guide instructional refinement. However, at the individual level, what we observe in learners is only a glimpse into a complex cognitive landscape within developing minds. How profoundly might the learning sciences progress if we could look into and interpret the formation of knowledge within the minds of simulated students, and subject them as cognitive crash dummies to alternative forms of instruction? As the track lead for the Computational Models of Learning track at CMU's LearnLab summer school—a first-of-its-kind workshop series—I have **trained nearly 100 learning engineers** to use my simulated learners in precisely these ways.

A central outcome of my PhD research was the development of a theory called **Decomposed Inductive Procedure Learning (DIPL)** [4], which builds on decades of simulated-learner research. DIPL specifies how different abductive and inductive learning mechanisms work together to produce human-like learning of procedural tasks common in STEM domains. This computational theory specifies how humans generalize from small amounts of natural instruction, such as demonstrations and correctness feedback, to achieve knowledge-level learning. Beginning as nearly a blank slate with limited prior knowledge, my DIPL-based simulated learners can learn as quickly from intelligent tutoring system instruction as human students, which I have shown can be **1000x more efficient than deep reinforcement learning** [4]. I have also demonstrated in several publications that these simulated learners can **replicate human learning curves without fitting to student data** [16,13]. Additionally, applications of my agents have **predicted experimental outcomes** in advance of running experiments with students [11]. In ongoing follow-up work, we have extended DIPL to learn from natural language instruction, including mechanisms for mutual disambiguation between worked examples and hints [10], setting the stage for agents that learn from scratch, naturally from human tutoring.

## <u>Thrust 2: AI Evaluations and Epistemic Perspectives in EdTech</u>

While most research artifacts never find practical application, AI systems have a way of being applied in advance of evidence of their efficacy. In some cases, forms of evidence that are generally accepted by the majority of researchers can be epistemically flawed, having no bearing at all on real-world outcomes. In my own work, I have found evidence of this sort of **widespread epistemic oversight in research on knowledge tracing systems**, data-driven systems that learn to estimate students' mastery of individual concepts and skills as they work in educational technology, and recommend appropriate next practice problems. As I have shown in simulation [7], the goodness-of-fit-based measures by which knowledge tracing systems are typically compared have no precise correspondence to their ability to appropriately prescribe next practice problems. This means new knowledge tracing systems that are cutting-edge in terms of their fit to data have the potential to harm student learning outcomes.

Vastly greater epistemic problems arise from generative AI's ability to convincingly replicate human behaviors while learning and performing in ways that are fundamentally distinct from human cognition. I have written [1] on how decades of philosophy of mind and cognitive science have paved a path for approaching these developments wisely. As a practical tool, learning engineers are simultaneously optimistic about generative AI yet troubled by its inaccuracy [3], unpredictability, and the absence of applied learning science in its design and common usage.

Toward evaluating AI tools on their ability to deliver **personalized cognitive support in automated tutoring** and facilitate the **precise study of human learning**, I created **TutorGym** [2], a standard API for evaluating generative AI and computational models of learning at adaptive tutoring and student simulation. TutorGym interfaces AI agents with 223 different existing Cognitive Tutors, Apprentice Tutors, and OATutors. In our initial evaluations, we found that LLMs were poor at tutoring—none did better than chance at labeling incorrect actions, and next-step actions were correct only ~52-70% of the time—but they could produce remarkably human-like learning curves when trained as students with in-context learning. Used in combination with datasets collected from its tutoring systems, TutorGym will provide a testbed for my future work in student simulation and computational theory refinement.

## Thrust 3: Teachable Agents and Interactive Task Learning

My simulated learners are interactive task learners; they can learn not only from data, but also rapidly from direct human instruction. For my PhD dissertation, I built **AI2T**, an interactively **teachable agent for authoring intelligent tutoring systems (ITSs)**. Authors tutor AI2T by providing a few step-by-step solutions and then grading AI2T's own problem-solving attempts. In just 20-30 minutes of interactive training (instead of hundreds of hours of programming), AI2T can induce robust rules for step-by-step solution tracking (i.e., model-tracing) that are 100% complete and accurate at automating intelligent tutoring system behavior. This enables learning engineers to **author custom AI-delivered adaptive tutoring** experiences with well-defined cognitive supports that are **free of hallucination-based inaccuracies**. Through bottom-up induction, AI2T induces and refines interpretable knowledge structures in the form of hierarchical task networks (HTNs), and has self-aware learning capabilities facilitated by an algorithm that I developed called **STAND** [8]. With STAND, AI2T can produce **precise estimates of its own learning progress,** which addresses a significant problem in interactive task learning: determining when the agent has finished learning. In my current postdoctoral research, I am combining AI2T's bottom-up induction with methods of top-down knowledge construction from natural language (i.e., from explanations of HTN task decompositions into subtasks). This work focuses on both academic tasks and broader game-based and web automation-based tasks. Our approach has the **potential to significantly outperform VLM-based agentic AI systems** in task automation scenarios, laying a path for AI systems that are interpretable, easily customizable, self-aware of their capabilities, and extremely robust to error.

**Future Research:** I aim to be at the forefront of AI systems that learn and teach like humans. My simulated learners solve the difficult problem of replicating human learning efficiency. Yet, there is a great deal to be done in terms of **data-driven refinement of existing computational theories of learning.** I hope to use student data to form precise points of comparison between human and simulated learners and **develop a comprehensive cognitive architecture that addresses a wide range of facets of learning and cognition**. I aim to expand upon my computational modeling work to build an extensive and flexibly usable simulated learner toolset that can execute a theory of learning on instructional technologies to make **prescriptive and comparative instructional design predictions** [17]. Generative AI presents an opportunity to shortcut the specification of complex prior knowledge like natural language interpretation within this system, but it is not a viable standalone tool for precise theory development. Additionally, I am waiting on two grants with the possibility of subcontracting to my new institution to explore the applications of teachable agents. One proposal to **NSF CISE Core** explores authoring-by-tutoring as a means of engaging learning engineers with the precise cognitive demands of their instruction as they author it. As they tutor an AI agent to author tutor behavior, they engage in a form of **reflective cognitive task analysis** by observing the learning of, and **looking into the mind of, a simulated student**. In a proposal to the **Kaizen Toyota Research Initiative**, users teach an AI agent to automate specialized workplace tasks, and the executable rule-based knowledge is used to **teach operational knowledge to new employees**. Both proposals include explorations of how knowledge learned by an AI agent can be used to **flexibly deliver a variety of instructional approaches** in mastery-based learning systems. As a long-term research goal, I am also interested in using simulated students as **digital twins** that **reconstruct student knowledge states from data**, and prescribe individualized instruction and diagnose misconceptions.

## References


1. **Weitekamp, D.**, & Koedinger, K. [Manuscript under review] Computational models of learning: deepening care and carefulness in AI in education with Theory-Driven Simulation. International journal of artificial intelligence in education
2. **Weitekamp, D.**, N. Siddiqui, M., & J. MacLellan, C. (2025). TutorGym: A Testbed for Evaluating AI Agents as Tutors and Students. International Conference on Artificial Intelligence in Education, 361–376. Springer Nature Switzerland Cham.
3. Gupta, A., Reddig, J., Calo, T., **Weitekamp, D.**, & MacLellan, C. J. (2025). Beyond final answers: Evaluating large language models for math tutoring. International Conference on Artificial Intelligence in Education, 323–337. Springer Nature Switzerland Cham.
4. **Weitekamp, D.**, MacLellan, C., Harpstead, E., & Koedinger, K. (2025). Decomposed inductive procedure learning: Learning academic tasks with human-like data efficiency. Proceedings of the Annual Meeting of the Cognitive Science Society, 47.
5. **Weitekamp, D.**, Harpstead, E., & Koedinger, K. (2024). AI2T: Building trustable AI tutors by interactively teaching a self-aware learning agent. arXiv Preprint arXiv:2411. 17924.
6. **Weitekamp, D., (2024).** Building Educational Technology Quickly and Robustly with an Interactively Teachable AI. Doctoral dissertation, Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburg, PA.
7. Rachatasumrit, N., **Weitekamp, D. [equal contrib.],** & Koedinger, K. R. (2024). Good Fit Bad Policy: Why Fit Statistics Are a Biased Measure of Knowledge Tracer Quality. International Conference on Artificial Intelligence in Education, 183–191. Springer Nature Switzerland Cham. **[Best Paper Nominee 🥈]**

8. **Weitekamp, D.**, & Koedinger, K. (2024). STAND: Data-Efficient and Self-Aware Precondition Induction for Interactive Task Learning. arXiv Preprint arXiv:2409. 07653.
9. **Weitekamp, D.**, & Koedinger, K. (2023). Computational models of learning: deepening care and carefulness in AI in education. International Conference on Artificial Intelligence in Education, 13–25. Springer Nature Switzerland Cham.
10. **Weitekamp, D.**, Rachatasumrit, N., Wei, R., Harpstead, E., & Koedinger, K. (2023). Simulating learning from language and examples. International Conference on Artificial Intelligence in Education, 580–586. Springer Nature Switzerland Cham.
11. Rachatasumrit, N., Carvalho, P. F., Li, S., & Koedinger, K. R. (2023, June). Content matters: A computational investigation into the effectiveness of retrieval practice and worked examples. In *International conference on artificial intelligence in education* (pp. 54-65). Cham: Springer Nature Switzerland.
12. **Weitekamp, D.**, & Stevens, P. (2022). A Mobile Invented Spelling Tutoring System. International Conference on Artificial Intelligence in Education, 492–496. Springer International Publishing Cham.
13. **Weitekamp, D.**, Harpstead, E., & Koedinger, K. (2021). Toward stable asymptotic learning with simulated learners. International Conference on Artificial Intelligence in Education, 390–394. Springer International Publishing Cham.
14. **Weitekamp, D.**, Harpstead, E., & Koedinger, K. R. (2020). An interaction design for machine teaching to develop AI tutors. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–11.
15. **Weitekamp, D.**, Ye, Z., Rachatasumrit, N., Harpstead, E., & Koedinger, K. (2020). Investigating differential error types between human and simulated learners. International Conference on Artificial Intelligence in Education, 586–597. Springer International Publishing Cham.
16. **Weitekamp, D., III**, Harpstead, E., MacLellan, C. J., Rachatasumrit, N., & Koedinger, K. R. (2018). Toward Near Zero-Parameter Prediction Using a Computational Model of Student Learning. International Conference on Educational Data Mining (EDM) 2018.
17. Harpstead, E., MacLellan, C. J., **Weitekamp, D.**, & Koedinger, K. R. (n.d.). The use of simulated learners in adaptive education. AIAED-19: AI+ Adaptive Education, 1–3.
18. Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.
19. Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge‐Learning‐Instruction framework: Bridging the science‐practice chasm to enhance robust student learning. Cognitive science, 36(5), 757-798.
20. Self, J. (1995). Computational mathetics: towards a science of learning systems design.