

Beyond the Chatbot: Deploying Ivy, a Knowledge-Based AI Coach, in a Georgia Tech Online AI Course

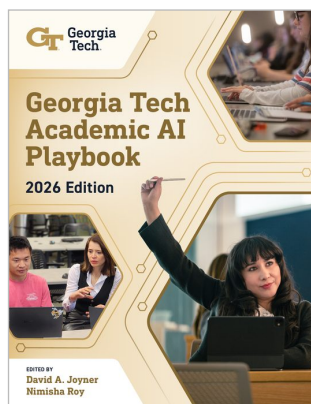
Rahul Dass, Shubham Puri, Arpit Khandelwal, Xiao Jin, Ankita Kundra, John Hall, Ashok Goel

Pages 45–50 of the Georgia Tech Academic AI Playbook, 2026 Edition

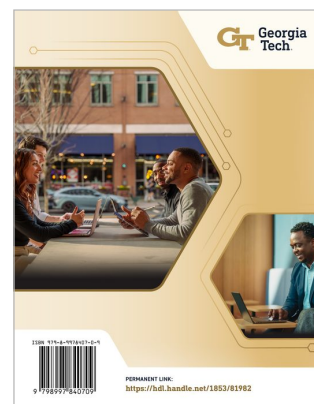
About the Playbook

The Georgia Tech Academic AI Playbook collects stories of how artificial intelligence is being used, taught, and responded to in Georgia Tech classrooms. The 2026 Edition gathers 67 articles from faculty, staff, and researchers across every college at Georgia Tech, along with the Library, the Center for Teaching and Learning, the Georgia Tech Research Institute, the Office of Information Technology, and the Provost's Office.

Each article is a “play”: a specific initiative carried out in a real course, documenting the prior context, the actions the instructor took, the results, and the lessons learned. The volume closes with 20 lessons synthesized across the full collection.



Front cover



Back cover

Copyright and License

Copyright for the content of this volume is retained by each section's author or authors. Copyright for sections not attributed to a specific author is retained by the editors.

ISBN 979-8-9978407-0-9. Edited by David A. Joyner and Nimisha Roy. <https://hdl.handle.net/1853/81982>

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Suggested Citation

Rahul Dass, Shubham Puri, Arpit Khandelwal, Xiao Jin, Ankita Kundra, John Hall, Ashok Goel (2026). Beyond the Chatbot: Deploying Ivy, a Knowledge-Based AI Coach, in a Georgia Tech Online AI Course. In D. A. Joyner & N. Roy (Eds.), *Georgia Tech Academic AI Playbook* (2026 ed., pp. 45–50). Georgia Institute of Technology. <https://hdl.handle.net/1853/81982>

Title: Beyond the Chatbot: Deploying Ivy, a Knowledge-Based AI Coach, in a Georgia Tech Online AI Course

Authors:

- Rahul Dass; Institute for Data Engineering and Science and National AI Institute for Adult Learning and Online Education; rdass7@gatech.edu
- Shubham Puri; National AI Institute for Adult Learning and Online Education; spuri62@gatech.edu
- Arpit Khandelwal; National AI Institute for Adult Learning and Online Education; akhandelwal83@gatech.edu
- Xiao Jin; National AI Institute for Adult Learning and Online Education; xjin96@gatech.edu
- Ankita Kundra; National AI Institute for Adult Learning and Online Education; akundra3@gatech.edu
- John Hall; National AI Institute for Adult Learning and Online Education; jkhall@gatech.edu
- Ashok Goel; School of Interactive Computing and National AI Institute for Adult Learning and Online Education; ashok.goel@cc.gatech.edu

Categorization: Using AI

Sub-Categorization: AI Tutoring & Intelligent Feedback

Formation

As generative AI tools have proliferated in education, most deployments share a common limitation: *they are general-purpose*. A learner asking ChatGPT how case-based reasoning — a skill taught in a graduate AI course — works will receive a fluent answer, but one generated from broad statistical patterns or retrieved from a textbook rather than from a structured understanding of the skill itself. This distinction matters especially in procedural skill learning, where the goal is not simply to teach skills in controlled pedagogical settings, but to help learners develop the kind of deep, adaptive expertise (Schwartz, Bransford, & Sears, 2005) that transfers to problems in the real world.

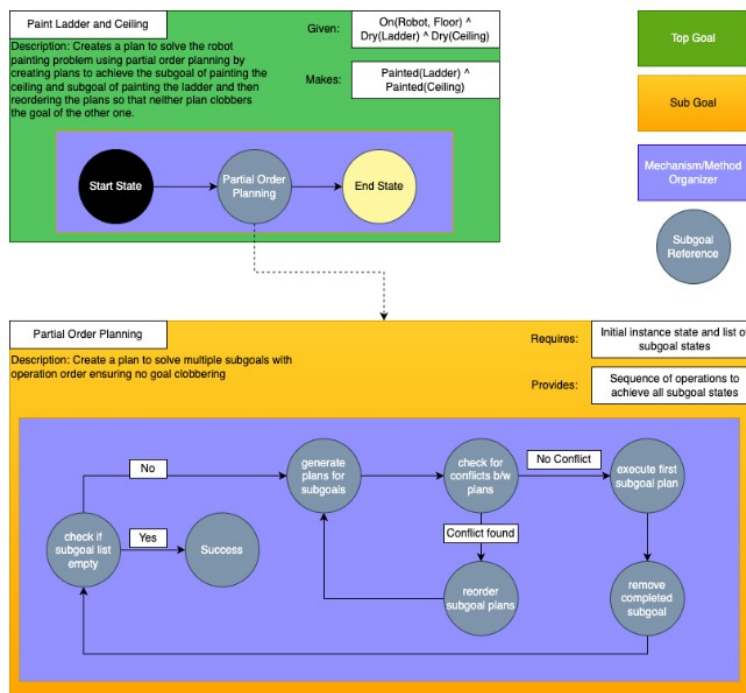


Figure 1: Partial TMK model for a KBAI skill.

Consider an organic chemistry student asking why a particular step in IUPAC nomenclature applies, or a medical student asking how to diagnose a condition. In each case, what the learner needs is not a fluent paraphrase but a causally grounded explanation of the skill's logic. CS 7637 Knowledge-Based AI or KBAI, a core course offered in both Georgia Tech's hybrid (~120 learners) and online (~800 learners) OMSCS formats, is exactly this kind of course. It teaches AI skills including semantic networks, partial order planning, and analogical reasoning through a library of pre-recorded instructional videos, at a scale that makes individualized human support difficult.

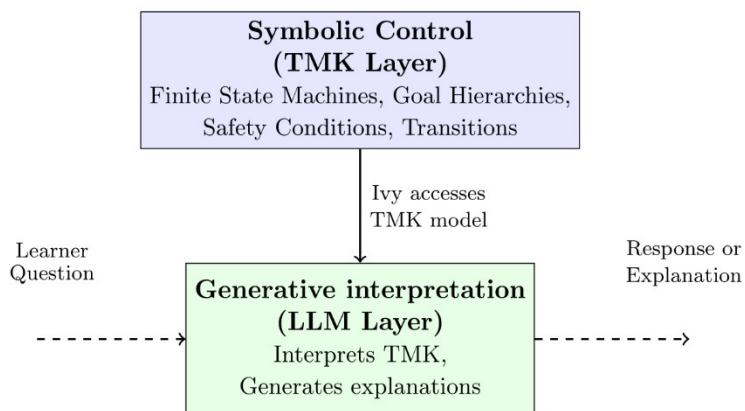


Figure 2: Concept diagram of Ivy's two-layered inferencing.

The Interactive Video, or *Ivy*, is a neurosymbolic AI coach (Dass et al., 2025a; Dass et al., 2025b), developed at the NSF-funded AI Institute for Adult Learning and Online Education (AI-ALOE; <https://aialoe.org/>) at Georgia Tech, co-designed and co-developed with the KBAI course instructor. What distinguishes Ivy from general-purpose AI assistants is that it grounds its responses in an explicit,

symbolic model of each skill: a Task-Method-Knowledge (TMK) model (Goel and Rugaber, 2017) representing the goals, mechanisms, and domain knowledge underlying the lesson (see Fig. 1). Given a learner's question, an LLM infers over this TMK model to synthesize a response — producing causal, goal-oriented explanations that reflect a deep understanding of the procedural skill being taught, rather than surface-level answers (see Fig. 2). Ivy is also embedded directly alongside instructional videos on the Ed Lessons platform (see Fig. 3), meaning learners can ask questions in context, at the moment they are watching and thinking about the material, rather than switching to a separate tool.

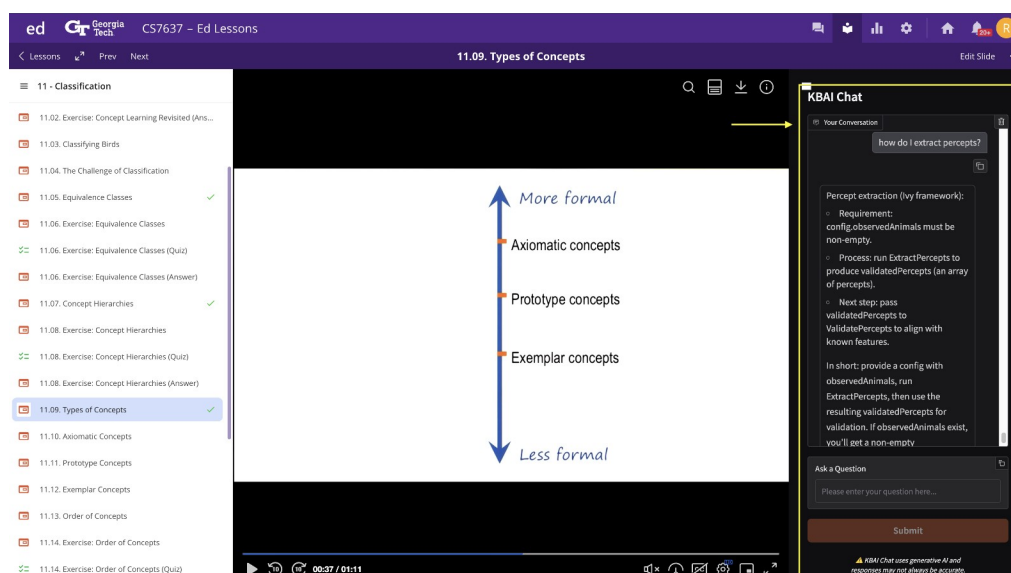


Figure 3: Ivy, accessed via the KBAI Chat interface, embedded alongside an instructional video on Ed Lessons deployed in the Spring 2026 KBAI course.

The Spring 2026 semester marked Ivy's first full-scale deployment, and making it possible required solving two challenges simultaneously. The first was *coverage*: building TMK models manually takes approximately 6-7 hours of expert effort per skill, meaning full-course coverage of 24 lessons would have required hundreds of hours. To address this, we developed a novel LLM-assisted “text-to-model” (TTM) methodology to generate structured TMK drafts from instructional materials such as course transcripts and lesson documents (Dass et al., 2026). This “System1/System2” (Kahneman, 2011) inspired AI-assisted authoring approach cut expert modeling time from roughly 6-7 hours to about 2 hours per skill, making full-course coverage practically feasible for the first time. In this sense, the Spring 2026 deployment used AI not only to coach learners, but to build the knowledge infrastructure that made coaching possible. The second challenge was *engagement*: a formative assignment was added to the course to structure learner interaction with Ivy in a deliberate way, rather than relying on voluntary use alone. The result was two complementary modes of engagement: voluntary, self-directed use alongside course videos, and structured, assignment-driven interaction with a clear pedagogical purpose.

Action

The Spring 2026 deployment was made possible by a knowledge authoring effort described in the Formation section above: using the TTM methodology (Dass et al., 2026), we constructed TMK models for all 24 KBAI lessons, giving Ivy full-course coverage. From the learner's perspective, this translated into a seamlessly integrated experience: a chat panel embedded directly alongside every instructional video on the Ed Lessons platform as an always available coach. Even with full TMK coverage across all 24 lessons, not every learner query maps cleanly onto a skill model — a learner might ask a broader

conceptual question, or phrase something in a way that falls outside Ivy's knowledge base. To ensure no question went unanswered, Ivy was integrated with Jill Watson (Goel & Polepeddi, 2018; Kakar et al., 2024; Taneja et al., 2024), a virtual teaching assistant also developed at Georgia Tech, which largely uses retrieval augmented generation over the KBAI course textbook and course syllabus to offer broader coverage.

The two systems are complementary by design: Ivy provides depth where its TMK models apply; Jill Watson ensures breadth where they do not. Together they formed a dual-agent system designed to maximize the support any learner could receive. Beyond voluntary, self-directed use, a dedicated lab assignment was introduced as a formative assessment to structure and study how learners engaged with Ivy in a more guided way. This design was informed by earlier pilot work showing that novices tend to ask more definitional, surface-level questions, while experts tend to ask more conceptual and causal questions aligned with higher-order thinking (Lum et al., 2025); the assignment was structured to scaffold novice learners toward the kind of higher-order questioning experts ask unprompted. This structured learner engagement in ways that voluntary use alone did not: learners were asked to pose questions spanning multiple levels of cognitive complexity, engage in multi-turn follow-up, and critically compare Ivy's knowledge-grounded responses with those of a general-purpose LLM like ChatGPT, Claude or Gemini. Learners were also asked to reflect on where structured explanations added value and where they felt limited. Guest lectures introduced the learning problem Ivy addresses and the logic of TMK-based knowledge representation, meaning learners engaged with Ivy not just as a tool but as an object of critical inquiry. Together, the dual-agent deployment and the formative assignment created a classroom environment where AI assistance was simultaneously available, structured, and subject to reflection.

Result

Preliminary analysis of the Spring 2026 deployment involved 115 enrolled learners, with a 91.4% participation rate. Among participating learners, engagement appeared substantial. The mean number of questions per learner was 17.7, with nearly 90% of participants asking 11 or more questions over the course of the semester. Together they generated 2,041 questions alongside nearly 75,000 video log events. Analysis of these data revealed a striking behavioral pattern: learners who voluntarily engaged with Ivy while watching videos before the Ivy lab assignment was available showed significantly stronger video engagement overall, watching more videos and remaining active across more days. By contrast, learners who engaged primarily under assignment-driven deadline pressure showed no such correlation with video activity; their use of Ivy appeared decoupled from their broader video-watching behavior. This modality dissociation would be invisible in aggregate usage statistics, and it has direct implications for how instructors interpret AI coach adoption data.

The lab assignment asked learners to pose four categories of questions: basic, complex, critical, and creative. This maps loosely onto increasing cognitive complexity based on Bloom's taxonomy (Krathwohl, 2002), but also means that the distribution of question types in the dataset was shaped in part by the assignment design itself. Notably, assignment-driven questions skewed toward higher-order reasoning: 69% fell in the Apply-to-Create range of Bloom's taxonomy, compared with 41% of questions posed voluntarily and self-directed. This gap indicates that the assignment succeeded in eliciting the intended question types. Whether TMK-grounded structure specifically contributes to this shift, or whether a similar scaffold would produce comparable results with any AI system, is a question the ongoing analysis aims to address.

Currently, we are analyzing structured learner responses and evaluations collected through the assignment, in which learners assessed the quality of Ivy's responses, critically compared them against a

general-purpose LLM, and reflected on where TMK-grounded explanations supported their procedural understanding and where they fell short. The high participation rate and the procedural depth of learner questioning suggest that Ivy is meeting a genuine need. The ongoing analysis will help clarify how well, along two fronts: how online learners actually engage with a structured AI coach relative to the learning problem it is meant to address, and what these usage patterns imply for how instructors might redesign curricula and assignments in a landscape where general-purpose AI tools are already ubiquitous.

Lessons

- **Lesson 1: Our capacity to build TMK models, not learner demand, was the bottleneck.** In Spring 2025, Ivy was partially deployed across four of 24 KBAI lessons. Even within that limited scope, 105 learners asked about 400 questions organically, with no assignment requirement, and over 63% were higher-order questions probing procedural understanding. Learners were ready to use Ivy well before we had built enough TMK coverage to meet that demand. This motivated investing in full-course TMK modeling ahead of Spring 2026 which was made feasible by a novel LLM-assisted “text-to-model” methodology that cut expert modeling time from roughly 6-7 hours to about 2 hours per skill.
Lesson 2: Combine a procedurally grounded coach with a courseware-grounded virtual teaching assistant by design, not as a fallback. Ivy and Jill Watson address different learner needs: Ivy provides causal, goal-oriented explanations grounded in the procedural structure of a specific skill; Jill Watson provides broader, courseware-grounded coverage for questions outside Ivy’s models. Our hypothesis is that deciding this division of labor upfront, rather than discovering it through trial and error, gives learners a more consistent experience regardless of which system answers their question. We suspect this dual-agent setup could be useful wherever learner questions span both procedural depth and broader courseware coverage, though testing that beyond KBAI is a natural next step.
- **Lesson 3: Separate voluntary engagement from assignment-driven engagement before interpreting usage data.** Aggregate usage statistics mask a meaningful behavioral distinction: in Spring 2026, learners who engaged voluntarily while watching videos showed strong coupling with overall video activity, while learners engaging primarily under deadline pressure did not. Separately, the sharp shift toward higher-order questions in assignment-driven interactions vs. organic, self-directed use (69% vs. 41%) reflects the assignment’s deliberate design, not incidental compliance. This confirms that the assignment succeeded in eliciting the intended question types. Whether this kind of scaffolded engagement with Ivy translates into a deepened understanding of the underlying skills is a separate question we are currently analyzing. Instructors evaluating an AI coach should therefore separate voluntary use from assignment-driven use before drawing conclusions from usage data, since combining the two can hide or distort what is happening — a pattern we had already glimpsed in Spring 2025, when short, conversational-looking questions seemed to carry some cognitive depth.

References

- Dass, R. K., Bowlin, T., Li, Z., Jin, X., & Goel, A. K. (2025b). Improving Procedural Skill Explanations via Constrained Generation: A Symbolic-LLM Hybrid Architecture. arXiv preprint arXiv:2511.20942.
- Dass, R. K., Madhusudhana, R. H., Deye, E. C., Verma, S., Bydlon, T. A., Brazil, G., & Goel, A. K. (2025a, July). Ivy: a hybrid knowledge-based and generative AI coach for explaining procedural skills. In *International Conference on Artificial Intelligence in Education* (pp. 233-246). Cham: Springer Nature Switzerland.

Goel, A. K., & Rugaber, S. (2017). GAIA: A CAD-like environment for designing game-playing agents. *IEEE Intelligent Systems*, 32(3), 60-67.

Dass, R. K., Puri, S., Khandelwal, A., Jin, X., & Goel, A. K. (2026, June). Developing Models of Procedural Skills using an AI-assisted Text-to-Model Approach. In *Proceedings of the Thirteenth ACM Conference on Learning @ Scale* (pp. 69-78).

Goel, A. K., & Polepeddi, L. (2018). Jill Watson: A virtual teaching assistant for online education. In *Learning engineering for online education* (pp. 120-143). Routledge.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kakar, S., Maiti, P., Taneja, K., Nandula, A., Nguyen, G., Zhao, A., Nandan, V., & Goel, A. (2024, June). Jill Watson: Scaling and deploying an AI conversational agent in online classrooms. In *International Conference on Intelligent Tutoring Systems* (pp. 78-90). Cham: Springer Nature Switzerland.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212–218.

Lum, C., Deye, E., Brazil, G., Bydlon, T., Verma, S., Madhusudhana, R., Dass, R. K., & Goel, A. K. (2025, June). Designing an AI coaching system for interactive video-based skill learning. In *International Conference on Intelligent Tutoring Systems* (pp. 281-291). Cham: Springer Nature Switzerland.

Schwartz, D., Bransford, J., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1-51). Greenwich, CT: Information Age Publishing.

Taneja, K., Maiti, P., Kakar, S., Guruprasad, P., Rao, S., & Goel, A. K. (2024, July). Jill Watson: A virtual teaching assistant powered by ChatGPT. In *International Conference on Artificial Intelligence in Education* (pp. 324-337). Cham: Springer Nature Switzerland.