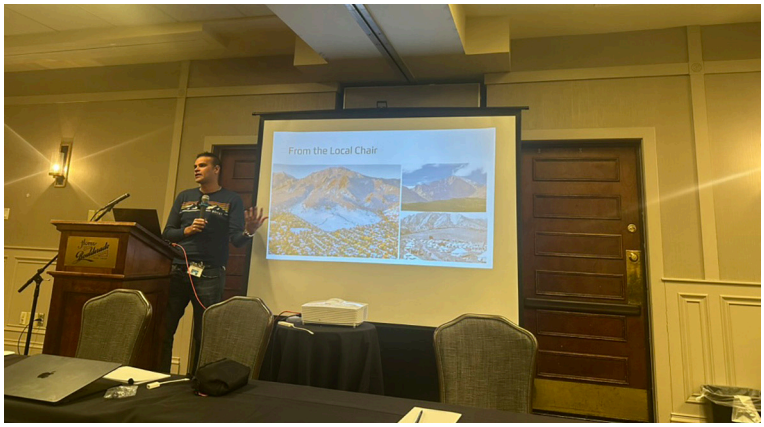
iSAT Snapshot

NSF AI INSTITUTE FOR STUDENT-AI TEAMING Winter 2025

In Brief . . .

A quick look at our activities this quarter!

1. Strand 1 lead, James Pustejovsky, and theme lead, Nikhil Krishnaswamy, presented research findings at the 31st International Conference on Computational Linguistics (COLING) in Abu Dhabi, UAE. The conference had around 2,000 attendees, and 50-60 participants attended James and Nikhil's talk on "Bridging Linguistic Theory and AI: Usage-Based Learning in Humans and Machines".
2. iSAT director Sidney D'Mello gave a keynote talk Student-AI Collaborations for the 21st Century Classroom at Dolby's Future of Learning Summit Agenda in San Fransisco.
3. iSAT director Sidney D'Mello was the local organizer for the Winter Text & Discourse Conference, where more than 50 participants engaged on topics related to discourse processing especially in the age of generative AI. iSAT External Advisory Board (EAB) member Alyssa Wise was a panelist for the Opening Panel on AI and Education Discourse and iSAT Executive director Peter Foltz demonstrated the Jigsaw Interactive Agent (JIA).
4. Dr. Marcelo Worsley, PhD, Associate Professor of Computer Science and Learning Sciences, Northwestern University and the chair of iSAT's External Advisory Board (EAB) visited CU Boulder in February. While he was there, he met with several iSAT team members and was an Institute of Cognitive Science Colloquium presenter. It was great for the team to get his feedback on several iSAT projects!
5. iSAT's Early Career Scholar (ECS) Forum kicked off the spring semester! This group aims to support iSAT's students, postdocs, and research scientists who are early on in their careers. This semester they are participating in several events including a Q & A session with Institute of Cognitive Science's colloquium speakers, a writing workshop, feedback and brainstorming on resume building, conference presentations and journal publications, and share-outs on research tools and techniques.
6. iSAT's high school internship program is in full swing. This program, now in its third year, is designed to enhance students' ability to envision, co-create, critique, and apply student-AI collaborative learning technologies both in the classroom and beyond. The interns choose an area of iSAT that best matches their interests and then conduct targeted research projects under the guidance of iSAT mentors.
7. Members from Strand 3 presented to educators at a workshop in Costa Rica! They shared iSAT's work with the Sensor Immersion curriculum unit as well as a demo of iSAT's AI partner Community Builder (CoBi).
8. Strand 2 Lead Sadhana Puntambekar has been selected as a 2025 AERA (American Educational Research Association) Fellow! One of only 29 Fellows chosen this year, this prestigious honor recognizes exceptional scholars for their excellence in educational research. Congratulations Sadhana!



iSAT Director Sidney D'Mello presents at the Winter Text & Discourse Conference.

From the PI

We hit a major milestone this Winter - the classroom testing of our second AI Partner, the Jigsaw Interactive Agent (JIA). JIA was organically conceived through a meticulous process which involved analyzing several hours of student collaborations to identify areas where they excel and places where they could use AI partner support. We already developed and tested a version of JIA in the iSAT Lab over the past couple of years, but are very proud to now field it in real-world classrooms, with more testing scheduled over the next few months. While we're still in the process of analysing the data, initial observations suggest that students really appreciated and benefited from this type of AI partner support. Supporting multiple groups of students in these classroom pilots also revealed unique challenges that were not seen in the lab studies, highlighting the importance of doing this research in the "wild". Kudos to our all-star team representing all three research strands and institute-wide members for accomplishing this major milestone.

As we prepare to wrap up our first performance period (2020-2025), we are emphasizing our dissemination efforts through publications, presentations, and outreach events to communities of practice. We're also ensuring that we have policies and procedures in place to ensure iSAT's sustainability and a smooth transition to the next phase of the project. And we're preparing for a busy Spring season of reporting and evaluation. Since this is our final year, the annual site visit will provide a unique opportunity to showcase all we have accomplished over the past five years, in terms of our scientific and technological achievements, but also the lives we have positively impacted and hope to impact in the years to come.

Feature: JIA in the Classroom

Testing JIA in the classroom with the iSAT Moderation unit!

Jigsaw activities are a type of collaboration where students first develop initial expertise and then come together to pool their knowledge and jointly solve more complex problems. The iSAT curriculum units include jigsaw activities to foster knowledge and collaboration among students. iSAT's Jigsaw Interactive Agent (JIA) is designed to enhance these activities by facilitating more effective knowledge sharing and brainstorming during small group discussion, potentially leading to increased social cohesion in collaborative Jigsaw activities. We have been iteratively developing and testing JIA in the lab over the past year. In the last two months, we have deployed and tested versions of JIA in middle school classrooms.

Curriculum Context

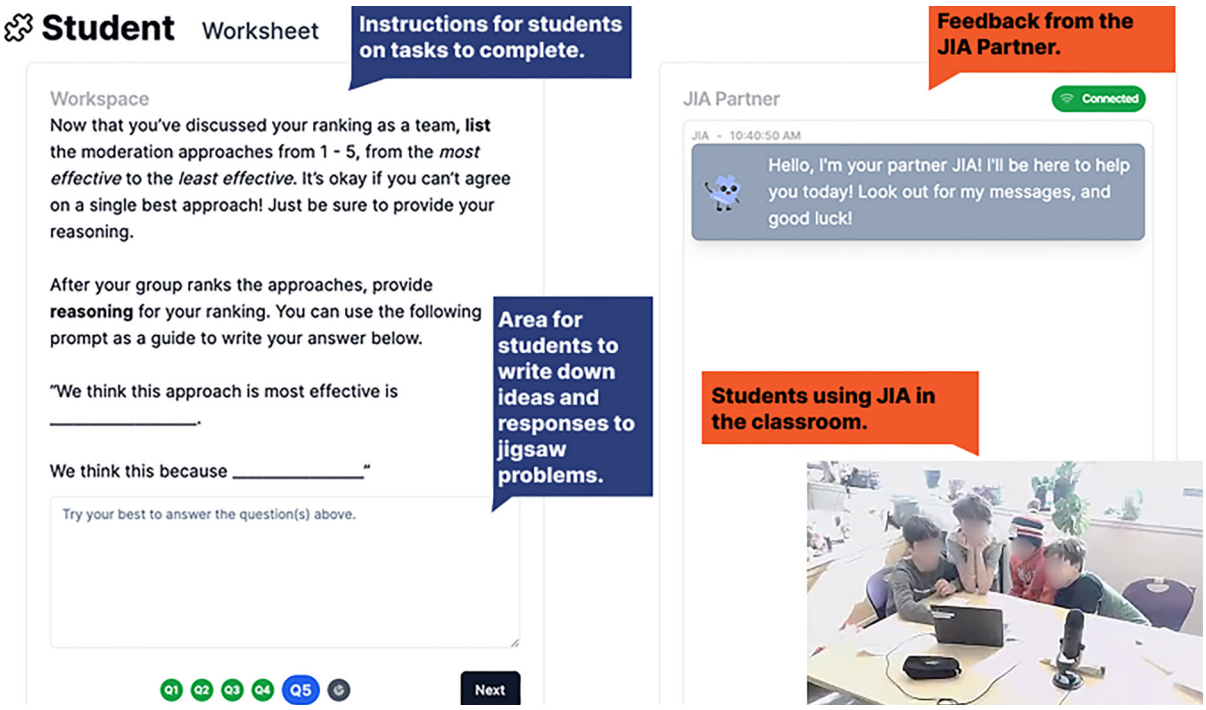
The curricular context for this study is a 3-week unit that is focused on the topic of AI and human moderation in Online Gaming (Moderation Unit). The unit was designed using the storyline approach (Reiser et al., 2021), which has been used to design open-education resources (OERs) in science education. There are several hallmarks to the storyline approach: each unit introduces students to an anchoring phenomenon where students discuss and then generate questions about the phenomenon to form a driving question board (DQB, Weizman et al., 2010). The board guides the class's inquiry for the remainder of the unit (e.g. after they engage in investigations that help answer a subset of questions, they move on to the next set of questions, and so forth). Over the course of the curriculum, students also learn about sentiment analysis and build sentiment bots to better understand how the amount of data and the perspectives of the human trainers can influence data modeling and refinement. Additionally, they develop and test a set of moderation rules for both humans and AI and employ them during gameplay to better understand the different affordances of human-AI moderation systems. The culminating task for the unit invites students to draw on what they learned to design a game space for players.

Embedding JIA in the Curriculum

The JIA partner was embedded in a lesson where students brainstorm possible criteria they might use to evaluate the effectiveness of online moderation systems. Students independently read about different moderation approaches (e.g. AI-only keywords, AI-only bag-of-words, human-only volunteer, etc.). Students then form small groups and summarize the approaches they read about; they work together to rank these approaches from most to least effective, using the criteria they generated earlier in the lesson. The task is completed in an online worksheet. JIA provides feedback and guidance to the group that is displayed on the right-hand side of the screen, while students type responses into a workspace on the left side of the screen (see the Figure below).

How JIA Works

JIA processes real-time small-group discussions to provide feedback, using two distinct implementations: a *rule-based* model and a *generative AI* model. Both operate on one-minute intervals, analyzing a previous segment of conversation to determine the appropriate feedback for



JIA interactive worksheet and AI partner in lesson 4 of the Online Moderation unit.
Bottom right: a group of students using JIA in the classroom.

recent research in both learning sciences and science education. Feedback is given along the following indicators:

Problematize: Juxtapose, compare, or contrast aspects of students' talk to seed uncertainty
Social Support: Provide social support for ongoing collaboration
Connect: Invite others into conversations, make connections between ideas
Stabilize: Take stock of where conversation is, marking areas of convergence or divergence

The *rule-based* JIA agent integrates utterances with their associated predictions from iSAT's Multimodal Intelligent Analyzer (MMIA), a pipeline that applies various Natural Language Processing (NLP) models to classify student utterances. These models provide insights into collaboration dynamics, such as collaboration (Community Builder), teamwork functions (Collaborative Problem Solving (CPS)), topic and task relevance, speaker dynamics, and verbosity. Based on thresholds given to these predictions, JIA determines the current state of collaboration, for example, when there is no speech, off-topic speech, uneven discussion, or low uptake based on CPS predictions. These states are then mapped to the feedback categories and hand-crafted responses specific to the students' current activities.
The generative AI implementation leverages LLaMa, a large language model, to analyze the student conversations from the transcript itself. That is, it receives the past minute of the transcribed conversation along with a structured prompt containing its role, task instructions, assessment criteria, and output formatting guidelines. Every 10 seconds, the model processes the previous minute of discussion and generates a structured response with a summary, intervention type, justification, and personalized feedback for the student groups.

Feedback is provided at one-minute intervals to avoid overwhelming students. The rule-based approach ensures structured feedback based on a theory driven approach with predefined heuristics, while the generative AI model provides more flexibility and awareness of context in its responses. These two implementations offer different trade-offs in flexibility, adaptability, and generalizability as well as explainability, allowing for comparative analyses of their impact on student collaboration. Current research is investigating how to combine the best of both approaches.

Classroom Use of JIA

Building on insights from the lab research, we deployed JIA in two middle school classrooms implementing the AI and Moderation curriculum unit. One classroom was an English Language Arts class and the other was a math class. We had several groups of students using JIA while they completed lesson four in the curriculum. While an analysis of this data is ongoing, we present some initial insights from our observations.

Preliminary Findings

Students were attending to the feedback from both the rule-based and LLM-based agents. In the dialogue snippet below with the rule-based agent, JIA prompts the students with a statement and question that is designed to stabilize the conversation by having them *elaborate on their thinking* and *provide reasoning* (two highly effective discourse practices for academically productive talk moves). Student A reads the feedback from JIA to the group and then responds with additional information about the group's reasoning on their choices.

Rule-based JIA: I'd love to hear more about how you all are deciding which moderation approach is most effective and why. Can you all provide a bit more reasoning as you make these decisions? Be sure to be listening to one another as they share!
Student A: Okay, so how did you decide? How did you decide which group was the most effective?
Student B: We figured that 3 was the most effective because it used both AI and streaming.

In the next dialogue snippet with the LLM-based agent, the agent incorporates information from the students' prior conversation about AI-powered moderation and asks them to compare/contrast it to their current discussion about using human volunteers to moderate, which is a form of *problematizing*. One student reads the JIA feedback while another student provides more information about how the moderation approaches are similar.

LLM-JIA: How does the volunteer moderation approach relate to the AI-powered moderation we discussed earlier?
Student A: Oh, how does the volunteer...
Student B: Well, they're both moderating, they're both ensuring that there's like...

Results suggest that both JIA agents are recognizing and responding to the context of the student conversations and that JIA's feedback is then being incorporated into student discussions. However, through our field note observation and reflective conversations with teachers, we have noted instances where students remarked that the rule-based feedback could be repetitive and that the LLM-based feedback would sometimes misinterpret the context. Ongoing analysis is focusing on: a) examining the contextual and pedagogical appropriateness of the JIA feedback, b) annotating the transcripts to determine how well the feedback supports students in maintaining productive struggle, and c) iterative redesign of the agents and user interface for classroom use. We plan to deploy a revised version of JIA in classrooms in April.

References

Reiser, B. J., Novak, M., McGill, T. A., & Penuel, W. R. (2021). Storyline units: An instructional model to support coherence from the students' perspective. Journal of Science Teacher Education, 32(7), 805-829.

Weizman, A., Shwartz, Y., & Fortus, D. (2010). Developing students' sense of purpose with a driving question board. In R. E. Yager (Ed.), Exemplary science for resolving societal challenges (pp. 110–130). NSTA Press.

Listen, Understand, and Connect

Strand 1

Strand 1 is guided by the foundational question: What advances in Artificial Intelligence are needed to facilitate collaborative learning conversations? The three research themes that help Strand 1 answer this question are: *Speech Processing & Diarization*, *Content Analysis & Dialog Management*, and *Situated Grounding*.

Speech Processing & Diarization

This team is working to improve Automatic Speech Recognition (ASR) models to make it easier for the AI partners to better tell apart who is speaking in a group and when a group conversation switches from one speaker to another.

Recently, the team conducted experiments on custom-trained, Transformer-based ASR models using the MyST dataset (containing children's speech). The experiments suggest that fine-tuning a Transformer (pre-trained on LibriSpeech adult speech) on children's speech data has a greater effect on the performance of the encoder rather than the decoder.

Considerable progress has also been made on improving speech diarization – who is speaking when – in the classroom. The team submitted a paper to InterSpeech 2025 on using a Large Language Model (LLM) Agent-based approach to refining the outputs of ASR (e.g., Whisper) by harnessing information about the context and also phonetic similarity between different words. The goal is to improve ASR accuracy particularly on named entities where contemporary ASR systems often make mistakes. The team further had a paper accepted to the Journal of Educational Data Mining on Optimizing Speaker Diarization for the Classroom focusing on applications in timing student speech and distinguishing teachers from students. Findings from continuing experiments on multimodal speech processing using discrete Transformer-decoders has been submitted to the International Conference on Multimedia & Expo (ICME) 2025; here, the team considered the case where video containing both speech and visual context is transcribed into text. In particular, the team is exploring motion cues (e.g., a ball bounces first against a wall and then onto the floor rather than just static image information) and how vital these are for correct understanding, and what kinds of encoding mechanisms are most effective for capturing motion cues. Lastly, the team is preparing a review paper on multi-person simultaneous speech recognition using end-to-end neural architectures. This has implications for the iSAT context where there is speech from multiple students at the same table and from neighboring tables.

Content Analysis & Dialog Management

This theme is dedicated to helping the AI partners make sense of what they are hearing and seeing and then determining optimal interactions between students and teachers. This work is intended to enable the partners to understand key content words and concepts uttered by students.

The team is currently adapting the coding conventions of Abstract Meaning Representation (AMR) data to optimize its representation for more efficient processing within computational constraints. Given the need to fit a greater volume of AMR text within predefined length limits, they are refining their approach to maximize information density without compromising AMR completeness. To support this effort, the team is fine-tuning the latest LLaMA 3 and Qwen 2.5 models, leveraging their advanced capabilities to enhance performance on this specialized task.

Team members working on Dependency Dialog Acts (DDA) continue to make progress on annotation. They have completed double annotation and adjudication for the Sensor Immersion (SI) sample and are halfway through annotating the remaining data. Additionally, they are getting ready to begin annotation of transcripts from the most recent lab studies.

An ongoing goal under this theme is to improve the Jigsaw Interactive Agent (JIA) by integrating curriculum materials from iSAT into an advanced AI system. The team is determining the best way to represent this educational content—either through large language model embeddings or AMR. To organize the materials for knowledge retrieval, they are using an open-source tool called the Nomic embedding model, which encodes the content sentence by sentence. Additionally, for both the curriculum materials and some of iSAT's lab studies, they have high-quality AMR annotations that serve as a benchmark. These annotations allow the team to compare a student's spoken or typed response with the curriculum content to identify the most relevant knowledge. The retrieved information is then used to guide JIA's responses. The responses are evaluated based on level of control as well as responsiveness. This work was just submitted to the 2025 conference on Educational Data Mining and is a combined effort of several team members.

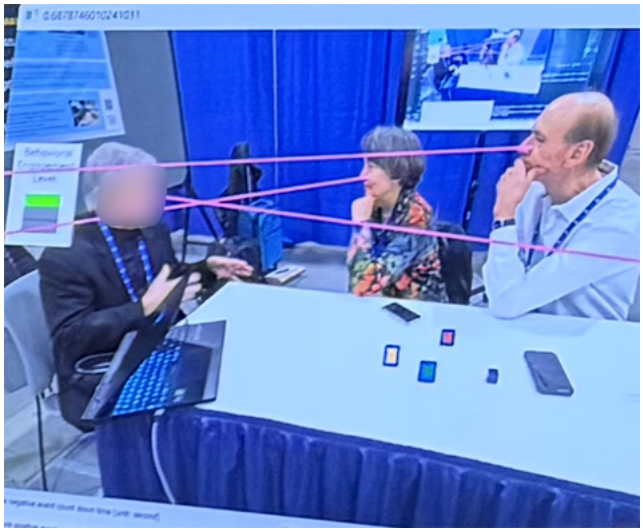
The AMR team has also been busy adding annotations to lab study transcripts and student responses. Over the winter, they created around 900 new sentence-level annotations and reviewed 500 existing ones for quality control. They also introduced new features, such as (anonymous) speaker ID labels for transcripts and a special tag to standardize references to MakeCode blocks. Since the Sensor Immersion (SI) curriculum does not directly track which code blocks students use, the team built an inventory of these blocks from the lab study tutorials. They then summarized this information in both sentence and graph formats to help determine the "correct" answers for a key lab study question. Additionally, the team is working on multi-sentence (MS) coreference annotations, which help to link ideas across sentences. Their efforts include training annotators, collaborating with tool developers to improve efficiency, and refining guidelines for tricky cases. So far,

they have created four MS coreference files and plan to start testing this method on transcript annotations very soon.

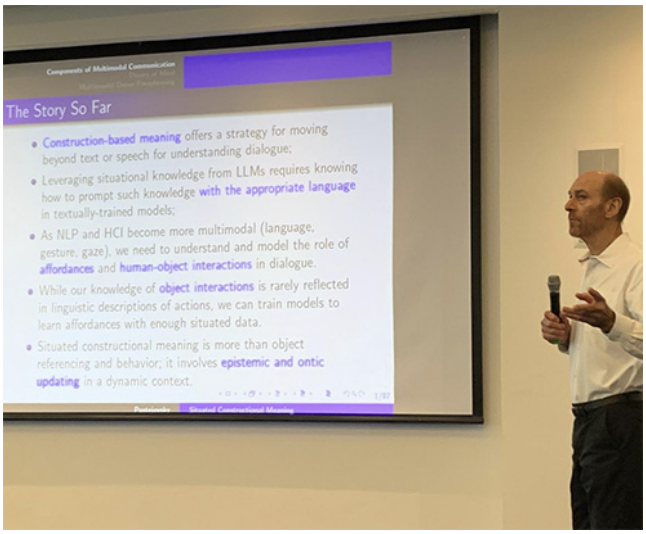
Situated Grounding

This theme is identifying common ground—established when students interact with one another through both behavior and verbal cues—through discourse and gesture. The Situated Grounding team has been actively preparing for the Association for the Advancement of Artificial Intelligence (AAAI) demo presentation scheduled for this winter. Their efforts include drafting, designing, and refining the poster presentation, as well as completing the initial draft of the demo interface. Additionally, the team has conducted local testing to ensure the functionality of the live demo. They have also designed a new SI experiment, which encourages participants to collaborate and maintain a high level of engagement throughout the session.

Meanwhile, several other team members have supported the integration of the AAAI demo with the underlying demo platform. They have also conducted experiments on recovering the active speaker from a single audio stream using multimodal signals—an approach that will help determine who is likely to be the active focus of attention in single-microphone scenarios. Additionally, they have performed ablations of multimodal features in common ground tracking.



iSAT Strand 1 leads, Martha Palmer and James Pustejovsky, demo JIA's multimodal features at the 39th Annual AAAI Conference on Artificial Intelligence.



Strand 1 Lead, James Pustejovsky, presents at the 31st International Conference on Computational Linguistics (COLING) in Abu Dhabi, UAE.

Orchestrating Effective Interactions

Strand 2

Strand 2 is guided by the foundational question: What advances in theories, interaction paradigms, and frameworks are needed to orchestrate effective student and teacher interactions with AI partners? The three research themes identified to help Strand 2 answer this question are: *Dynamic Framework & Measures of Collaboration*; *Collaborative Learning* (non-verbal and verbal communication, peer scaffolding); and *UX Design & Multimodal Modeling*.

Dynamic Framework & Measures of Collaboration

For this theme, the team works on identifying and measuring the basis of collaborative problem-solving skills in social, affective, and cognitive processes, and how to promote interactions in team problem-solving. In January, the team began running an experiment that investigates the impact of emotionally valenced AI feedback on team physiological and communication states in collaborative problem solving. The experiment involves two participants performing different roles (Controller, Guide) to collaboratively solve a series of block programming problems. Using the Wizard of Oz (WoZ) method, the AI partner provides hints to problem steps through chat that either have a neutral, positive, or negative tone. The AI feedback was quantified using sentiment analysis. The goal of the experiment is to investigate the transfer of positive versus negative experiences with an AI partner to other tasks between participants. This research is developing novel physiological metrics of teamwork that are sensitive to emotional states of team members and a communication-based metric of knowledge reorganization through collaboration.

Collaborative Learning

By identifying verbal and non-verbal modes of collaborative engagement and identifying peer scaffolding moves, Strand 2 hopes to better understand how student peers support each other during collaborative learning activities. Using current Jigsaw Interactive Agent (JIA) lab sessions as well as the classroom Online Moderation curriculum, the team has been continuing their work training additional researchers on how to use the Nonverbal Interactions in Collaborative-Learning Environments (NICE) framework to annotate classroom and lab videos. Expanding iSAT's team members who can perform this sort of task will significantly enhance our work on multimodal modeling. The team has been focused on cross-strand collaboration to review and finalize session videos to annotate with NICE; and to discuss which modalities are priorities and whether they can be captured to drive intervention decisions.

UX Design & Multimodal Modeling

The iSAT lab team is collecting data centered on collaboration to understand how data can be used to model collaborative behaviors and correspondingly implement these models as part of the AI partners.

The group has been busy working with team members from other strands to analyze a large multimodal dataset (including Kinect video, eyetracking, speech, and other non-verbal behavioral measures) that was collected in the iSAT lab study consisting of youth (aged 12-17) working on Jigsaw activities with support from JIA. In particular, the team focuses on transcribing the JIA lab study audio, analyzing eye tracking data using Tobii Pro Lab, and analyzing non-verbal behavior data using the NICE and Gesture Abstract Meaning Respresentation (GAMR) annotation schemes.

They have also created high-fidelity prototypes for a new JIA feature, "content support", which provides extra curriculum content for students when they work through the jigsaw activity. In addition, a couple of JIA User Interface (UI) design change requests derived from the JIA lab studies. These UI design changes (e.g., updated notification forms) will be implemented in the next version of JIA. They are also analyzing feedback from the classroom testing of JIA and expect to include corresponding UI changes in the next version.

Additionally, a system is being developed to rapidly evaluate responses generated from JIA partners through a combination of human and automatic evaluation. Part of this effort has been dedicated to adding the ability to run JIA agents on historical data, allowing future agents such as the multimodal agents to be evaluated without the need for further data collection. Also, the quality of the data has been improved by reprocessing video data through updated versions of the Multimodal Intelligent Analyzer (MMIA) - which is a real-time speed analyzer tool used to evaluate collaboration in small groups. These efforts will culminate in an evaluation study examining different features of AI agents and evaluating their effectiveness in creating quality responses.

Engaging Youth

Strand 3

Strand 3 is guided by the foundational question: In what ways can co-design processes empower stakeholders to envision, co-create, critique, and apply Artificial Intelligence learning technologies? The research themes identified to help Strand 3 answer this question are: *Learning Futures Workshop* (LFW), *Online Moderation Unit*, and *Sensor Immersion/Self-Driving Cars Units*.

Learning Futures Workshop

These are annual workshops consisting of high school students resulting in crucial feedback from the participants including what youth want and need the AI partners to be able to do in the classroom, and—just as important—what they do not want them to do. The LFW team is currently planning a spring workshop in Berkeley, CA, with a small group of educators to explore how AI can support learning in schools. The goal is to develop a process for engaging educators in envisioning AI's potential in education. As with past LFWs, the workshop will begin by examining current uses of technology, particularly generative AI tools like ChatGPT. From there, it will go on to explore broader applications, such as using AI for lesson planning and teaching rehearsals—a common teacher education method that allows educators to practice responsive teaching through simulated interactions.

Online Moderation Unit

The Online Moderation unit focuses on how and when AI can be used to help people moderate online communities, and the role that moderation plays in online communities.

The unit implementation continues in middle school classrooms this winter and spring. Team members have begun to incorporate iSAT's new AI partner, the Jigsaw Interactive Agent (JIA), into the curriculum and are testing and refining the approach based on feedback from teachers and students. With support from the Institute-wide data collection team, the unit has been used in several middle and high school classrooms over the span of three weeks.

The research team continues to analyze the data gathered from the classroom implementations, with a focus on: 1) how students are engaging with the accuracy of AI systems, 2) how their understanding of model building, training, and testing AI systems deepens through key lessons, and 3) how they are evaluating the role of norms in shaping online behaviors, especially in gaming communities. A key aspect of their learning here is supported by engaging with iSAT's first AI partner, the Community Builder (CoBi), and then critically reflecting on the accuracy of its output based on their own collaborations with one another. Several of iSAT's graduate students are supporting the analysis of the classroom data and assessment artifacts to help the team get a better understanding of what students learned as a result of the unit.

Sensor Immersion & Self-Driving Cars Units

The Sensor Immersion (SI) curriculum teaches students how to program a sound, a soil-moisture, and an environmental sensor using a micro:bit controller. Students gain a deeper understanding of how sensor systems work and how they can be used to collect, analyze, and display scientific data to support scientific investigations.
SI implementation continues with more teachers implementing the curriculum. To support the continued classroom implementation, the team regularly holds Professional Learnings (PLs) with teachers for them to engage with machine learning and artificial intelligence topics. During the PLs, teachers explore how to use CoBi in their classrooms and gain comfort with the routines for how CoBi supports collaboration discussions.

Strand 3's Jeff Bush hosts a Professional Learning workshop to introduce partner teachers to our curriculum units and the latest version of the AI Partner, "CoBi".

The Self-Driving Cars (SDC) curriculum has students engage with the decisions that a self-driving car needs to make to keep people safe and get them where they need to go. Students program their own cars and let them run on a race track, taking potential obstacles into consideration. The curriculum continues to move towards public release with fine tuning of support materials for teachers and students. Several of iSAT's partner teachers have tested these revisions and are working towards integrating feedback from testing to create publicly available curriculum support for the unit. Most recently, the SDC team is preparing publication of the SDC curriculum on CU Boulder's Inquiry Hub–a website that provides research-based STEM curriculum materials, professional development support, and assessment tools for educators.

Data Collection, Architecture & Implementation

Institute-wide

The Institute-wide team collaborates with all research strands by providing expertise, resources, and staff to promote and support research convergence. Our main themes include *AI Partner Design, Implementation, and Testing*; *Technical Architecture*; and *Collaborative Problem-Solving Framework*.

AI Partner Design, Implementation & Testing

This theme focuses on the main activities supporting the design, implementation, and testing of the AI partners - the Community Builder (CoBi) and the Jigsaw Interactive Agent (JIA). This requires managing a cross-Institute roadmap ranging from conceptualization to implementation and validation for different versions of AI partners.

Both of iSAT's AI partners, CoBi and JIA, are being used in the Sensor Immersion, Self-Driving Cars, and Online Moderation curricula classrooms. The CoBi team is also studying the extent to which students demonstrate improvements in their collaborative problem solving skills and learning outcomes as a result of using the tool. The team is starting to analyze the data collected in the Fall for publication in Summer 2025.

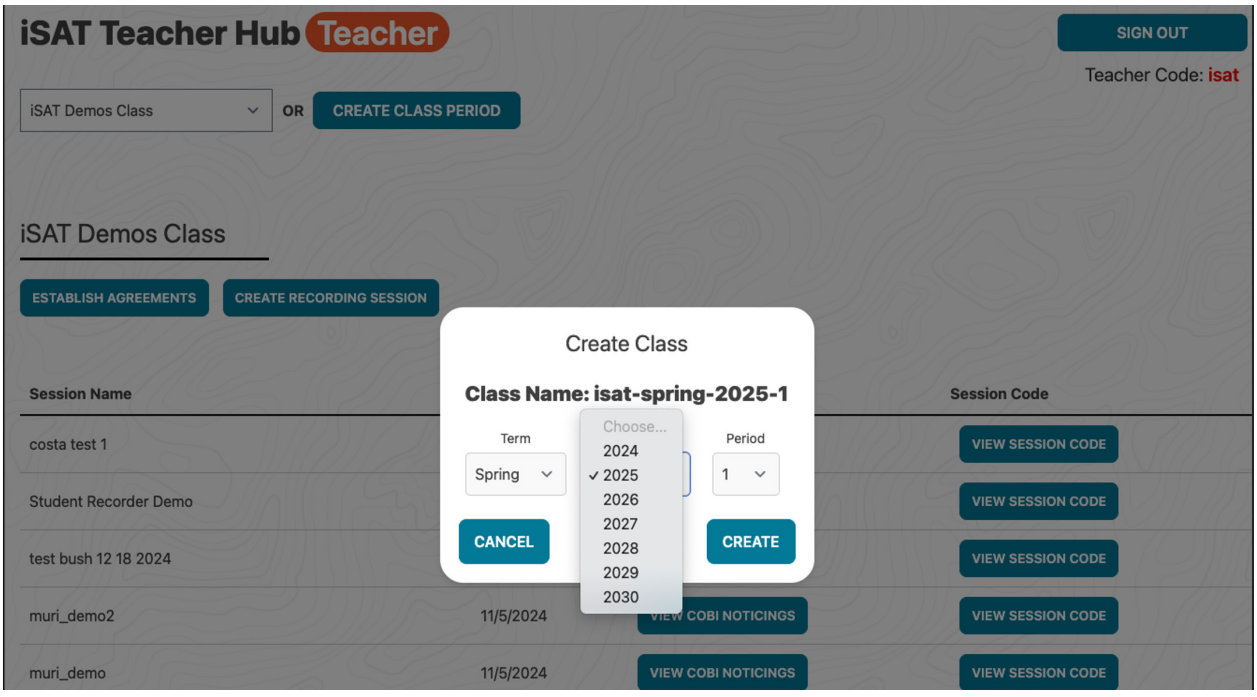
The team also uses these classroom studies - including feedback from students and tutors - to refine CoBi. The team recently launched a new Teacher Hub interface (see below) to enable teachers to more seamlessly integrate the AI partners into their classroom practice, which is an important feature as iSAT scales its technologies, JIA data collection is focused specifically on a jigsaw activity in the Online Moderation curriculum, in which students share their knowledge with others in small groups after having built expertise in different areas. For a full update on JIA in the classroom, please see the Feature Story on page 3.

Technical Architecture

This theme focuses on developing the core infrastructure and components for the AI partners, supporting the integration of AI models and services into the AI partners, and supporting secure data storage to classroom and lab study data.

Several improvements have been made to the AI partners in the last several months. For CoBi, the team implemented the designed previously mentioned of the new Teacher Hub user interface (UI), which enables teachers to view, create, and manage classes and sessions from a single UI.

Refinements to JIA implementations include an optimized agent architecture deployed in classrooms. This architecture provides a faster, near real-time speech processing pipeline used by the JIA agents. In addition to partner improvements, the team has made improvements to the speech processing pipeline, which includes implementing an improved sentence embedding diarizer that more accurately splits utterances at the sentence level.



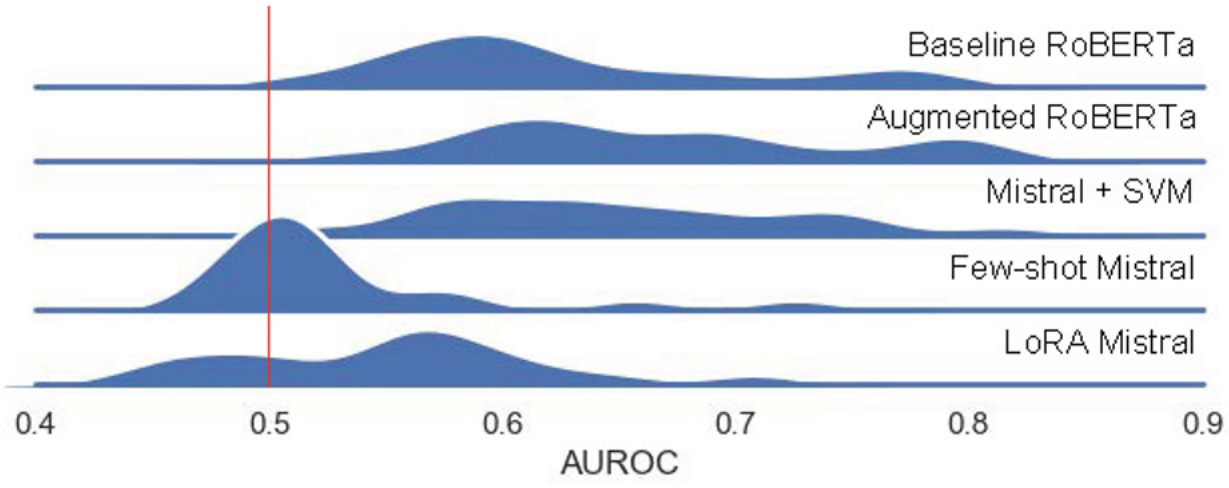
Creating a class in the new iSAT Teacher Hub.

Collaborative Problem-Solving (CPS) Framework

This theme focuses on the need to successfully identify the basis of collaborative problem-solving skills including social, affective, and cognitive processes, identify features of successful collaboration, and identify how to promote equitable and trusted interactions in team problem-solving.

Their current work dives into how different approaches to Collaborative Problem Solving (CPS) define and utilize behavioral indicators. Many coding schemes use the same vocabulary but apply it differently, while others use different terminology to describe the same concept. By examining these inconsistencies, team members are uncovering both the commonalities among, and the unique contributions of, various approaches to measuring CPS. To better understand how these indicators measure CPS across different tasks and contexts, the team has adapted the Generalized Competency Model of CPS to multiple collaborative activities in three different middle school curricula: Sensor Immersion, Self-Driving Cars, and Online Moderation. The annotation team has then used this updated CPS coding scheme to annotate behavior in small groups of middle school students performing these collaborative tasks. These human annotations of collaborative problem solving across task contexts create ground truth data for training computational models.

Relatedly, the team has been developing computational models of collaborative processes that generalize across five datasets encompassing different student populations, subject domains, physical contexts, and interaction modalities. The findings revealed that traditionally fine-tuning RoBERTa models on a single dataset (serving as the baseline) led to overfitting, with the model failing to generalize beyond the training data's specific curriculum and language patterns. In contrast, fine-tuning RoBERTa with embedding augmented data led to significant improvements in generalization, as did pairing Mistral embeddings with a support vector machine classifier. However, fine-tuning and few-shot prompting Mistral did not yield similar improvements. These updated models are being integrated into the AI partners so they can better generalize across curricula.



Distributions of test set AUROCs (model accuracy) for each of the five models. The red line indicates an AUROC of .05 or random guessing.

Member Spotlight

Learn more about our members!

Meet John Weatherley, a Senior Research Professional and iSAT Institute-wide member at CU Boulder. At iSAT, he leads the software development and architecture teams.

Q: What does your research focus on?

A: I lead a team that focuses on integrating the designs and technologies for iSAT's AI partners from members in each strand, and deploying and scaling them in the cloud for use in classroom and lab settings.

Q: What is the coolest thing about your research?

A: There's nothing more exciting than seeing a piece of complex technology come together and work for the first time. I get to have this experience on a regular basis whether it be running a new script to tabulate the analysis data from a collaborative student session, or watching a new AI partner interacting with students in the classroom for the first time. This was the case when members from each strand worked for months to create the original CoBi AI partner. The effort involved many steps and seeing the CoBi tree come to life while students looked on in the classroom was priceless. I also got to hear my own two middle schoolers talk with excitement about JIA after participating in the JIA lab study, which was pretty cool.

Q: What has been a turning point or defining moment in your career?

A: A key turning point in my life was discovering that computer science and software engineering were my true calling. After college, I was working as an executive selling management skills training to corporations. While I was interacting with customers over the phone all day, I found myself drawn to the mainframe database system that housed my customers' sales history, finding joy in constructing the SQL-like queries to scan the data and generate sales leads. In addition, I was intrigued by how the database software itself worked. Who builds software like this? Knowing that computers and technology excited me, I interviewed a range of technical and creative people working in the industry as software engineers, web designers, and photoshop professionals, and determined that software engineering was the right fit. I knew that I needed more knowledge and skills and went on to get a Masters in Computer Science from CU Boulder, which led to a career as a software engineer and a focus on education technology.

Q: What do you like to do outside of work?

A: I love spending time outdoors with my family and friends - bicycling, skiing, hiking, backpacking, climbing, and camping. Along the way I enjoy photography and sharing stories. I also enjoy singing.

Meet Yifan Zu - a PhD student in computational linguistics at Brandeis University whose research focuses on multimodal meaning representations, through language and gestures.

Q: What are you currently working on?

A: Recently I just finished a live demo test for the Association for the Advancement of Artificial Intelligence (AAAI) conference. I am now working on a paper for HCII (the 27th International Conference on Human-Computer Interaction) on Neuro-Symbolic AI systems.

Q: How does your research contribute to iSAT?

A: My work at iSAT focuses on integrating multimodal analysis with dialogue management, contributing to building situation awareness. This helps the AI partners understand and facilitate conversations. I am studying how multimodal representations contribute to forming common ground in classrooms and other educational settings.

Q: What do you like most about your work at iSAT?

A: iSAT has given me the opportunity to meet many wonderful people and I have received significant support from my co-workers. Being part of such a diverse organization has given me a broader perspective which has been invaluable in shaping my career vision!

Q: What is a fun fact about you?

A: I hold two Masters degrees - a MA in linguistics and a MS in computational linguistics. Outside of work, one of my favorite things to do is travel and I also enjoy watching movies.

New Quarter, New Faces!

Growing our team and our impact.

We're excited to welcome new members to our team! Some of our new faces include:

Collette Heskett is a graduate student at CU Boulder. Her research interests are focused on the impacts of game-based science education on secondary science students, focusing on how it shapes their social interactions and deepens their understanding of scientific concepts. She is currently obtaining a MA in Learning Sciences and Human Development, focusing on the research and understanding of professional knowledge building in multiple different environments including school and communities.

Chetan B. is a graduate student at CU Boulder, who is pursuing a master's degree in computer science. He earned his bachelor's degree from R.V. College of Engineering in Bangaluru, India, and subsequently worked as a cloud developer at Hewlett Packard Enterprise for three years. At iSAT, he is part of the Institute-wide team as a Software Engineer.

Jalynn Nicoly is a first-year PhD student in the Computer Science Department at CU Boulder, who is interested in the intersection between augmented reality and brain-computer interfaces. She recently graduated from Colorado State University with four years of research experience and was the first author of an ACM SAP published paper. Now in the SHINE Lab with Leanne Hirshfield, Jalynn works simultaneously with the iSAT Lab to address multimodal analysis.

Welcome Collette, Chetan and Jalynn!

Collette Heskett
Strand 3
Graduate Student

Chetan B.
Institute-wide
Graduate Student

Jalynn Nicoly
Strand 2
PhD Student

Get Involved!

Email, subscribe, and follow us to learn more!

info.ai-institute@colorado.edu

www.colorado.edu/research/ai-institute/

@NSF_iSAT

iSAT Publications and Outreach

Published / Accepted

Anderson, E., Lin, G., Farid, A., Fenech, M., Hanks, B., Klopfer, E., Doherty, E., Hirshfield, L., Ko, M., Foltz, P., Nguyen, H., Nguyen, V., Ludovise, S., Santagata, R., Cao, L., Scardamalia, M., Soliman, D., Resendes, M., Khanlari, A., Costa, S., Chan, C.K.K., Nguyen, A. (2025). Exploring GenAI Technologies within Collaborative Learning. International Society of the Learning Sciences (ISLS 2025).

Chang, M. A., Tissenbaum, M., Philip, T. M., & D'Mello, S. K. (2025). Co-designing AI with Youth Partners: Enabling Ideal Classroom Relationships through a Novel AI Privacy Framework. Computers & Education: Artificial Intelligence.

Dey, I., Doherty, E., Zhang, R., Hoang, N., Bush, J., Hirshfield, L., & Puntambekar, S. (2025). Analyzing Support Moments During Small-Group Work. [Poster Presentation]. American Educational Research Association (AERA 2025).

Doherty, E., Perkoff, M. E., Von Bayern, S., Zhang, R., Dey, I., Puntambekar, S., Hirshfield, L. (2024). Piecing Together Teamwork: A Responsible Approach to an LLM-based Educational Jigsaw Agent. The ACM (Association of Computing Machinery) CHI conference on Human Factors in Computing Systems (CHI 2025).

Foltz, P. W. & Spain, R. (2025). Generative AI for Assessing and Supporting Complex Performances in Teams. Accepted Chapter in *Generative AI in Intelligent Tutoring Systems*.

Ko, M., Chandler, C., Weatherley, J., Raju, R., Muddaluru, R., Foltz, P. W. (2025). Detection and support of states of collaboration through discourse using the Jigsaw Interactive Agent (JIA). Demonstration given at the Society for Text and Discourse Winter Text Conference, Boulder, CO.

Palmer, D., Zhu, Y., Lai, K., VanderHoeven, H., Bradford, M., Khebour, I., Mabrey, C., Fitzgerald, J., Krishnaswamy, N., Palmer, M., and Pustejovsky, J. (2025). Speech Is Not Enough: Interpreting Nonverbal Indicators of Common Knowledge and Engagement. In AAAI Conference on Artificial Intelligence (AAAI): Demos Program. AAAI. (Accepted for publication).

Penuel, W. R., Bakker, A., & Akkerman, S. Design research in between the past and future: Linking the epistemic, ontological, and axiological dimensions of generative projects. Special issue on Design Research in Mathematics Education in Mathematical Thinking and Learning.

Wang, J., He, X., Southwell, R., Dudy, S., Wang, Z., and Whitehill, J. (2025). Optimizing Speaker Diarization for the Classroom: Applications in Timing Student Speech and Distinguishing Teachers from Children. Journal of Educational Data Mining.

Submitted / Under Review

Breideband, T. and D'Mello, S. (under review). Intelligent Tools to Enhance Collaboration in Small Groups. Submitted to Computational Group and Team Dynamics: Forging an Interdisciplinary Science.

Breideband, T., Bush, J., Chandler, C., Chang, M. A., Hirshfield, L., von der Wense, K., Penuel, W. R., Reitman, J. G., Rose, S., Weatherley, J., D'Mello, S. (under review). CoBi: AI-based Collaboration Support and Community Building for K-12 Classrooms. Submitted to International Journal of Artificial Intelligence in Education (IJAIED).

Bush, J. B., Biddy, Q., Chang, M., Benedis, G. G., Breideband, T., Sumner, T., Penuel, W., Weatherly, J., Singh, T., D'Mello, S. (under review). Community-Building in K12 Classrooms with a Collaborative AI Agent and Associated Curriculum Routines. Long paper submitted to the International CoOnference on Computer Supported Collaborative Learning.

Bush, J.B., Dey, I., Hoang, N. (under review). Investigating teacher support for collaboration: lessons from a mixed-methods study in middle-school CS classrooms. Long paper submitted to the International Conference on Computer Supported Collaborative Learning.

Chandler, C., Rohit, R., Reitman, J.G., Penuel, W.R., Ko, M., Bush, J.B., Biddy, Q., and D'Mello, S.K. (under review). Improving the Generalizability of Models of Collaborative Discourse. Submitted to Educational Data Mining Conference (EDM 2025).

Chang, M., Penuel, W. R., & Philip, T. (under review). "It goes both ways": Envisioning mutually compassionate relations between teachers and students. Submitted to The International Conference of the Learning Sciences (ICLS 2025).

Dey, I., Doherty, E., Zhang, R., Hoang, N., Bush, J., Hirshfield, L., & Puntambekar, S. (under review). Examining support moments with a conversational AI partner. Submitted to the International Society of the Learning Sciences (ISLS 2025).

Submitted / Under Review Continued

Dey, I., Ko, M., Puntambekar, S. (under review). The Blind Men and the Elephant: Visualizing Collaboration Using Multiple Modalities. Submitted to the International Society of the Learning Sciences (ISLS 2025).

Dey, I. and Puntambekar, S. (under review). The NICE framework: Examining nonverbal interactions in collaborative groups. Submitted to Discourse Processes: Special Issue on Multimodal Measures of Collaboration.

Doherty, E., Reitman, J., Lai, K., Cooke, N. J., & Foltz, P. W. (under review). Verbal discourse in collaboration. Submitted to Discourse Processes: Special Issue on Multimodal Measures of Collaboration.

Guan, Y., Trinh, V.A., Voleti, V., and Whitehill, J. (under review). Multi-modal Speech Transformer Decoders: When Do Multiple Modalities Improve Accuracy? Submitted to the International Conference on Multimedia and Expo 2025.

Ko, M., Chandler, C., Weatherley, J., Raju, R., Muddaluru, R., Foltz, P. W. (under review). Facilitating Productive Uncertainty in Small Group Jigsaw Activities: Designing and Implementing an AI partner in the Classroom. Submitted to the Artificial Intelligence in Education conference (AIED 2025).

Ko, M., Penuel, W.R., Hoang, N., Howard, A., Heskett, C. (2025) How does an AI-Partner Embedded Curriculum to Support Students in Critically Evaluating AI systems? A Pilot study. Submitted to the Artificial Intelligence in Education conference (AIED 2025).

Perkoff, M., Bayern S. V., Zhao, Z., Cai, J. Z., Doherty, E., Wright-Bettner, K. E., Martin, J., Palmer, M. (under review). Golden Retrievers: Incorporating Curriculum Knowledge into Pedagogical Agents. Submitted to Educational Data Mining Conference (EDM 2025).

Spencer, Yin, X., Gorman, J. (under review). Multimodal Influence Across Speech and Eye Gaze During Collaborative Problem Solving (CPS) Predicts Socio-Cognitive-Affective States and CPS Outcomes. Submitted to Discourse Processes: Special Issue on Multimodal Measures of Collaboration.

Trinh, V.A., He, X., Whitehill, J. (under review). Improving Named Entity Transcription with Contextual LLM-based Revision. Submitted to Interspeech 2025.

VanderHoeven, H., Bhalla, B., Khebour, I. Youngren, A., Venkatesha, V., Bradford, M., Fitzgerald, J., Mabrey, C., Tu, J., Zhu, Y., Lai, K., Pustejovsky, J. and Krishnaswamy, N. (under review). Real-Time Multimodal Common Ground Tracking in Situated Collaborative Dialogues. Submitted to Meeting of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL).