

Developing and Evaluating an AI agent to Read and Interpret Learning Maps

Dante Cisterna
dcisterna@ku.edu

Jonathan Schuster
jgfsla@ku.edu

Amber Samson
ambersamson@ku.edu

University of Kansas, ATLAS

Abstract

We describe the development and evaluation of an AI conversational agent designed to interpret learning maps, consisting of graphic models of student cognition for a particular domain. Results indicate the agent can accurately identify map components and structures, especially when different informational formats are combined. Findings highlight the tool's feasibility to inform the development of tools that use and interpret learning maps for teachers.

1 Introduction

Learning maps are graphic models of student cognition that display multiple pathways of knowledge and skills that increase in complexity toward a specific learning goal (Bechard et al., 2021; Kingston & Broaddus, 2017). Nodes in learning maps are fine-grained knowledge, skills, and understandings (KSUs) that represent cognitive foundations and academic content related to a learning target (Kingston et al., 2016). The connections between nodes show directional relationships from cognitively less to more complex KSUs, and the multiple pathways formed by these interconnected nodes reflect research on how students acquire academic content (Dynamic Learning Maps Consortium, 2022). Unlike learning progressions, which often consist of a single, linear pathway with few nodes, learning maps provide multiple, fine-grained routes for learners to reach academic targets, helping teachers better track and support student learning (Kingston & Broaddus, 2017). Learning maps have been used to develop assessments (e.g., DLM assessments; Dynamic Learning Maps Consortium, 2022) and guide teacher instruction (ATLAS, 2025). In addition, there is growing interest in using learning maps to support teachers' learning in planning and

instructional work (Clark et al., 2023; Swinburne Romine et al., 2025). For example, the Shared Writing Instructional Model (SWIM) project developed a learning map-based writing intervention model that is adaptable and effective for students across the full range of writing abilities (Swinburne Romine et al., 2023a). In the SWIM approach (Swinburne Romine et al., 2023b), teachers use clusters of conceptually related nodes in learning maps to identify initial instructional goals for individual students according to their current writing level. They then select a customized, flexible plan focused on the student's progress toward these goals to guide instruction. Finally, teachers reflect on and evaluate their writing instruction and the student's performance, suggesting the student's location on the learning map and appropriate goals for future instruction.

The development of new technologies focused on generative artificial intelligence (AI) can create opportunities to develop learning and support tools to help teachers make informed decisions that guide students toward their next steps in learning. For example, AI-based agents can help teachers efficiently identify the specific knowledge and skills a student displays at a given moment and support student progress to a more advanced level of performance. Therefore, a first step would be to design AI-based agents that accurately interpret learning map information to serve as a basis for developing teacher support tools. Several research and development efforts have focused on using AI and machine learning to develop graphic representations that map student ideas and skills based on knowledge corpora such as open educational resources or online materials (e.g., dos Santos et al., Reales et al., 2024; Schicchi et al., 2025; Zhai, 2025). Zhai (2025) highlights the opportunities of Large Language Models (LLMs)

to create concept maps and scale them up, although limitations related to validity, user interpretation, and pedagogical implications remain, suggesting that using human-in-the-loop approaches, adding specialized knowledge sources for agent development, and combining different evaluation methods can enhance the pedagogical value of map-creator agents. Schicchi et al. (2025) reported that concept maps developed by AI models were evaluated similarly to teacher-created maps. Still, the role of experts with specialized knowledge in both the discipline domain and pedagogy remains underexplored. In this study, we investigate how an AI agent can interpret cognitive representations of learning and replicate the knowledge of subject matter experts (e.g., map developers).

2 Research Question

Our study describes the development and evaluation of an AI-based conversational agent capable of interpreting learning maps. We investigate whether the AI agent can learn to read learning maps with accuracy and efficiency as map developers would expect. Therefore, the study’s research question is, “Can the AI agent accurately interpret different types of learning maps and formats of map data?”

3 Methods

We conducted successive rounds of development and testing of a conversational agent using the full version of Microsoft Copilot Studio (using the Claude Sonnet 4.6 model). During the agent’s training, we included learning map documentation, such as guidelines for developing learning maps, information from the Dynamic Learning Maps program website, and narratives on learning map structure. We also provided the agent with additional clarifying information about specific learning map characteristics.

Our iterative development and testing focused on the impact of two evaluator-provided features on the AI agent’s responses: 1) specific types of interactions and 2) different learning map structures, views, and formats. For the first feature, we explored whether different interaction types, specifically in the ways the evaluator interacts and responds to the agent, may facilitate the agent’s interpretation of learning maps presented in visual representations (i.e., PDFs).

The interaction condition types are:

1) No Feedback to Agent: Provided the agent with no additional content outside of the prompts targeting specific map characteristics.

2) Limited Feedback to Agent: Provided the agent with limited feedback aimed at eliciting reactive, actionable information from the agent following the prompts to improve its performance. For example, for the question of “What are the individual pathways through the learning map,” example feedback could be “That’s not quite what I was looking for. I want to know how to get from an entry-point node in the learning map to the final node” or “I’m not sure I understood that. Could you put it another way?”

3) Probe Agent: Gained insight into the agent’s understanding of learning maps through active, in-depth, and scaffolded questioning. For example, for the question “What is the overall topic covered in the learning map?”, example probes could include “How did you arrive at that response?” or “What information in the learning map did you use to identify that topic?”

For each condition, evaluators used a specific set of prompts to assess the agent (see Table 1).

No Feedback Condition	Limited Feedback Condition	Probing Condition
How many nodes are in the learning map?	What are the smaller topics related to the overall topic covered in the learning map?	What is the overall topic covered in the learning map?
How many connections are in the learning map?	What standards would be covered in the learning map?	Which node(s) represent the learning map’s learning target?
What node(s) are at each linkage level?	Which node(s) represent the learning map’s learning target?	What standards are represented by the learning target?
What is the first node(s) in the learning map?	What nodes precede the node(s) at a critical point in a student’s content learning?	What nodes significantly contribute to the acquisition of the node(s)?
What nodes come between two critical points in a student’s content learning?		What are the individual pathways through the learning map?

Table 1: Example prompts asked of the agent for each prompting condition.

3.1 Agent Evaluation

We developed two testing protocols and trained two human evaluators with expertise in learning maps' design and use to test and interact with the agent under different conditions, covering the specific interaction types and different formats of learning map information (12 outputs for the first feature and eight outputs for the second one, 20 cases in total). For the different interaction types (first feature), evaluators asked the agent specific prompts under the No Feedback, Limited Feedback, and Probing conditions using the four small learning maps (two maps per condition).

To examine the second learning map feature, the evaluators also asked the agent specific prompts in three conditions: Map Formats, Map Views, and Map Views plus Map Formats. The Map Formats condition examined whether specific format combinations improve agent performance, using a PDF image and either a spreadsheet or screen-reader text of the same learning map (i.e., one of the two non-linear mini-maps). In the Map View condition, evaluators explored whether the agent can accurately interpret the two learning map neighborhoods. The final condition, Map Views plus Map Formats, examined whether combining different formats (PDF and spreadsheet) improves agent performance on the neighborhoods. The prompts the evaluators asked the agent were the same as the ones used in the prompting type conditions except for the inclusion of ones focused on the number of node clusters in the learning map neighborhoods, what nodes represent each node cluster, and what nodes significantly contribute to the acquisition of each node cluster in the Map Views and Map Views plus Map Formats conditions.

Across both features, the evaluators began their conversations with the agent by saying "Hello" before waiting for the agent's response. This conversational starter aimed to evaluate whether the agent greets the user, describes its purpose, and asks the user to upload a learning map file for interpretation. Upon completion of each testing round, the evaluators uploaded the transcripts of their conversations with the agent before briefly summarizing their evaluation of, or experience interacting with, the agent for each condition.

We reviewed the evaluators' transcripts and evaluation summaries for the 20 sessions. One team member read the transcriptions and became familiar with the data corpus. Then, guided by the

research question, we sorted conversation extracts and assigned them to themes that described patterns and reflected interaction moves suggesting the agent's ability to interpret learning maps, as described by Ryan and Bernard (2003). We selected excerpts for each topic and created operational definitions. Two additional researchers reviewed and verified the selected categories.

In addition, we further analyzed the conversation extracts and themes demonstrated in the evaluators' transcripts and summaries through a descriptive topic analysis. To accomplish this, we used specific criteria adapted from Denecke (2023) to evaluate the agent's performance. The criteria included the following evaluation categories: Accessibility, Ease of Use, Engagement, Flexibility in Dialogue Handling, Content Accuracy, Context Awareness, and Error Tolerance. The Accessibility criterion examines the ease with which an individual can use the conversational agent and the agent's ability to process data in different modalities and to adjust the conversational level of its responses (e.g., the amount, clarity, and use of specific language and vocabulary) to increase their comprehensibility, while the Ease of Use criterion assesses the extent to which an individual believes that the use of the conversational agent would be simple and easy, which encompasses the degree of control and flexibility users have when communicating with it, the perceived quality of the agent's responses, the time required to provide the responses, and the number of errors that occur. The Engagement criterion judges the value an individual places on using the communication agent, increasing the likelihood of continued use. The rate at which they can accomplish a desired goal when using the communication agent, the time it takes them to achieve the goal, the amount of dialogue required, and their satisfaction with the entire process contribute to this criterion. Next, the Flexibility in Dialogue Handling criterion evaluates the communication agent's ability to maintain a conversation and handle all the user's questions or prompts, including the number of back-and-forth exchanges required to achieve the desired goal when unexpected input, questions, or prompts arise. For the Content Accuracy criterion, the proportion of responses from the communication agent that are consistent with those of map developers for the same set of prompts is evaluated, while the Context Awareness

criterion judges the extent to which the communication agent uses the contextual knowledge in a conversation to appropriately respond to users' prompts and questions, reflecting how well the agent can efficiently notice topic changes, even when the new topic is not clearly articulated by the user. Finally, the Error Tolerance criterion determines the extent to which the communication agent appropriately handles ambiguous or incomplete information or a lack of background knowledge to provide a correct response to a user's prompt or question. It would also cover conditional responses based on the information currently available.

Results

The agent's overall performance demonstrated that, following training, it could accurately and efficiently interpret learning maps as map developers would. Although the agent struggled in some areas, it proved to be very effective, accurate, flexible, and suitable for describing a learning map's content and structure. Regarding the Accessibility criterion, the agent had no difficulty reading the content and structure of the learning maps in PDF and spreadsheet formats. In addition, the agent successfully rephrased its responses or used simpler language to help the evaluators better understand them. Regarding the Ease-of-Use criterion, the agent consistently referenced the provided background documents in response to their prompts. The evaluators also reported a high degree of control and flexibility when communicating with the agent, easily guiding it through the series of prompts, asking clarifying questions when given unclear or incomplete information, and probing it about the rationale behind its responses. Most of the agent's responses were high-quality, based on their relevance and comprehensiveness, were delivered quickly, and produced few errors, which were quickly resolved. Regarding the Engagement criterion, the agent met the desired goals but took longer and required more dialogue to do so for more difficult learning map characteristics. Next, for the Flexibility in Dialogue Handling criterion, the agent readily maintained conversations about the learning maps and their characteristics with the evaluators and responded quickly and appropriately to unexpected prompts and questions. Regarding the Content Accuracy criterion, the agent was accurate in most cases, despite initially struggling to perceive certain

learning map characteristics. Interestingly, its performance improved as the evaluators provided additional information and guidance on their interpretation, enabling it to achieve high accuracy across prompts and questions by the end of the process. For the Context Awareness criterion, the agent quickly recognized and responded to topic shifts and variations in the evaluators' expressions. Finally, for the Error Tolerance criterion, the agent accurately inferred the intended meaning of prompts even when information was ambiguous or incomplete or when background knowledge was lacking, responding only conditionally as needed.

Table 2 presents identified topics, operative definitions, and example quotes that illustrate agent's responses when prompted by the evaluator. Thus, the AI agent could satisfactorily recognize learning map components, structures, and views, although challenges emerged in specific cases involving particular map views and some map formats. Regarding map structure, the analysis suggested the AI agent correctly and quickly identified the content and structure of learning maps with simple (linear and non-linear) structures such as nodes and some linkage levels, and any challenges it encountered (e.g., counting the number of connections and identifying the map's learning target) were easily resolved. However, for more complex maps, the agent successfully identified some of the easier map characteristics (e.g., identifying the overall topic and counting the number of nodes) but encountered some difficulty interpreting more difficult ones (e.g., identifying pathways between nodes and through the maps), which could have been caused by remaining grounded in the provided files rather than making broad inferences. In addition, these difficult characteristics remained a constant challenge for the agent because of the size of the more complex learning maps. Additional guidance and supplementary information resolved most of the agent's challenges, improving its performance.

Across map views, the agent could more readily interpret the critical content of the mini-maps than that of the learning map neighborhoods. The agent could identify most of the nodes in the mini-maps aligned with the linkage levels, as well as those surrounding and contributing to them. It encountered some trouble identifying multiple nodes aligned at the same linkage level, but its performance improved when provided with additional instructions, guidance, and

background knowledge. On the other hand, the agent struggled to do anything beyond detecting node clusters in the learning map neighborhoods without additional information beyond the PDF images.

Topic: Accuracy The agent often correctly identified the overall topic and associated learning standard when asked about general map content. Example interaction quote Agent: “The overall topic of this learning map is early literary comprehension, focusing on recognizing story elements and structure—progressing from identifying familiar objects and characters to understanding key narrative components like beginnings, middles, ends, and changes in stories.” [Output_Condition_2_ELA.EE.RL.4.5]
Topic: Response revisions When prompted, the agent was able to provide a stronger, more precise conceptual distinction. Example interaction quote Agent: “The target node reflects the grade-level expectation ... while the successor node generalizes the skill to less familiar or broader texts, extending beyond the standard’s immediate scope.” [Output_Condition_3_ELA.EE.RL.3.1]
Topic: Over inference The agent explicitly acknowledged it was using a generalized developmental assumption for describing a node pathway instead of reading the visual connections in the map. Example interaction quote Agent: “The order was incorrect earlier because I relied on linkage level intuition rather than the actual arrow structure.” [Output_Condition_3_ELA.EE.RL.3.1]
Topic: Combined formats The agent often answered accurately and reasonably when both a PDF and a spreadsheet were available. Example interaction quote Evaluator: “What node(s) are at the Distal Precursor linkage level?” Agent: “Based on the graphic image, there is one Distal Precursor (DP) node... the image appears to be the more precise source here.” [compared to the spreadsheet only]. [Output_Condition1_ELA.EE.RL.4.5.]

Table 2: Selected topic categories for agent/evaluator interactions.

Finally, while the agent encountered few problems in interpreting learning maps when using PDFs alone, its performance improved significantly when PDF files were paired with spreadsheets. This benefit was especially evident when identifying individual pathways between nodes and through learning maps, and when detecting the specific nodes within and between node clusters. Screen reader text was less useful for this task because the agent failed to use it when responding to the evaluators’ prompts.

Conclusions

The study’s results demonstrate the feasibility of developing an AI agent that can accurately identify learning map components and interpret map content and structure in ways consistent with subject-matter experts (e.g., map developers). Although challenges emerged when the agent encountered particularly complex map structures and views, as well as certain file formats, these limitations were substantially reduced by using additional guidance, structured representations of map data, and multiple complementary map formats. By establishing the technical feasibility of AI-based map interpretation, this study lays the foundation for future uses by map developers, assessment developers, and educators. Aligned with Zhai’s future directions for developing automated concept maps (2025), we consider that the constant evaluation of map developers and the use of multiple sources of information are essential for refining AI-agents that can read and interpret learning maps. For instance, such an agent could support map development and maintenance activities by quickly identifying existing maps and nodes related to specified topics of interest while leaving final decisions and revisions to human review. It could also be integrated into learning map software to generate preliminary map views for experts to review, validate, and refine. An AI agent that can read learning maps could also promote coherence across map-based products. For example, the agent could help verify that assessment, instructional, and professional learning resources align with the maps and map components with which they are associated, providing an efficient first review that subject-matter experts can confirm. Most importantly, the capacity to interpret learning maps means an AI agent can make complex learning map information

more accessible and actionable for educators through integration into teacher-facing platforms. Such an agent could support teachers' instructional decision-making by helping them identify students' likely locations within learning maps, determine appropriate next learning goals, and locate resources aligned with those goals. Educators could use these recommendations as starting points, adapting them based on their professional judgment and knowledge about individual students and instructional contexts. To support future use cases like these, the next step is to improve the agent's ability to process complex learning maps. This improvement would include identifying minimum technical requirements for map formats, metadata, accessibility features, and knowledge sources and expanding the agent's knowledge base with additional guidance on map interpretation and varied learning map representations. In addition, further investigation into privacy and security requirements will be critical for supporting reliable and scalable implementation. Collectively, these efforts will help determine whether an AI agent can serve as a practical bridge between learning maps and the map-based products and services designed to support educators and their students.

Acknowledgments

This work was funded by ATLAS (Accessible Teaching Learning and Assessment Systems), University of Kansas as part of the ATLAS AI Innovation Funding.

References

- Accessible Teaching, Learning, and Assessment Systems (ATLAS). 2025. *PIE as a proof-of-concept for future summative use: Evidence from a pilot study*. University of Kansas. https://pie.atlas4learning.org/sites/default/files/documents/resources/PIE_Proof-of-Concept.pdf
- Amy Clark, Jennifer Kobrin, Meagan Karvonen & Ashley Hirt. 2023. Teacher use of diagnostic score reports for instructional decision-making in the subsequent academic year. *Practical Assessment, Research, and Evaluation*, 28(1), Article 6. <https://openpublishing.library.umass.edu/pare/article/id/1255/>
- Daniel Reales, Rubén Manrique, and Christian Grévisse. 2024. Core concept identification in educational resources via knowledge graphs and large language models. *SN Computer Science*, 5(8), 1029. <https://doi.org/10.1007/s42979-024-03341-y>
- Daniele Schicchi, Carla Limongelli, Vito Monteleone and Davide Taibi. 2025. A closer look at ChatGPT's role in concept map generation for education. *Interactive Learning Environments*, 34(1), 216-296. <https://doi.org/10.1080/10494820.2025.2497110>
- Dynamic Learning Maps Consortium. 2022. *2021–2022 Technical Manual—Instructionally Embedded Model*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems. <https://2022-ic-techmanual.dynamiclearningmaps.org/>
- Gery W. Ryan and H. Russell Bernard. 2003. Techniques to Identify Themes. *Field Methods*, 15, 1, 109 - 85. <https://doi.org/10.1177/1525822x02239569>
- Kerstin Denecke. 2023. Framework for guiding the development of high-quality conversational agents in healthcare. *Healthcare*, 11(8), 1061. <https://doi.org/10.3390/healthcare11081061>
- Neal M. Kingston & Angela Broadus. 2017. The use of learning map systems to support the formative assessment in mathematics. *Education Sciences*, 7(1), 41. <https://doi.org/10.3390/educsci7010041>
- Neal M. Kingston, Meagan Karvonen, Sue Bechard and Karen Erickson. 2016. The philosophical underpinnings and key features of the Dynamic Learning Maps alternate assessment. *Teachers College Record: The Voice of Scholarship in Education*, 118(14), 1–30. <https://doi.org/10.1177/016146811611801410>
- Russell Swinburne Romine, Meagan Karvonen and Emily Thatcher. 2023a. *Meaningful writing instruction for students with intellectual disabilities* (SWIM Project Brief #1). University of Kansas, Accessible Teaching, Learning, and Assessment Systems. <https://swim.atlas4learning.org/sites/default/files/documents/RC/Project%20Briefs/SWIM%20Project%20Brief%201%20-%20Meaningful%20Writing%20Instruction.pdf>
- Russell Swinburne Romine, Emily Thatcher, and Meagan Karvonen. 2023b. *Shared writing instructional model: Planning, delivering and evaluating writing instruction for students with intellectual disabilities* (SWIM Project Brief #2). University of Kansas, Accessible Teaching, Learning, and Assessment Systems. <https://swim.atlas4learning.org/sites/default/files/documents/RC/Project%20Briefs/SWIM%20Project%20Brief%202%20-%20Shared%20Writing%20Instructional%20Model.pdf>
- Russell Swinburne Romine, Jonathan Schuster, Meagan Karvonen, W. Jake Thompson, Karen

- Erickson, Vanessa Simmering and Sue Bechard (2025). Learning maps as cognitive models for instruction and assessment. *Education Sciences*, 15(3), 365.
<https://doi.org/10.3390/educsci15030365>
- Sue Bechard, Meagan Karvonen, M., and Karen Erickson. 2021. Opportunities and challenges of applying cognitive process dimensions to map-based learning and alternate assessment. *Frontiers in Education*, 6, 653693.
<https://doi.org/10.3389/feduc.2021.653693>
- Vinicius dos Santos, Érica Ferreira de Souza, Katia Romero Felizardo, William Massami Watanabe, Nandamudi L. Vijaykumar, Sandra Maria Aluizio and Arnaldo Candido. 2019. A Conceptual Map Creation from Natural Language Processing: a Systematic Mapping Study. *Revista Brasileira de Informática na Educação*, [S. l.], v. 27, n. 3, p. 150–176, 2019.
<https://doi.org/0.5753/rbie.2019.27.03.150>.
- Xiaoming Zhai. 2025. Generative large language models for knowledge representation: A systematic review of content map generation. *arXiv:2509.14554* [cs.CY]
<https://doi.org/10.48550/arXiv.2509.14554>