RAPID: DRL-AI: Investigating A Community-Inclusive AI Chatbot to Support Teachers in Developing Culturally Focused and Universally Designed STEM Activities

We conducted research to begin the development of CATpc: Critical Activity Teacher Planning Companion. CATpc is a generative AI (genAI) chatbot companion for teachers to support them in adapting and creating culturally focused and universally designed STEM activities and learning environments while honoring, sustaining, and building upon the cultural capital and wealth of communities (Habig et al., 2021; Yosso, 2005) that students bring into the classroom.

Our work explored two intersecting research questions:

1) What are the directions, features, and dispositions that educational researchers, teachers, and families of marginalized and underrepresented students to design a genAI chatbot to support teachers create culturally relevant activities and lessons?
2) What are possible technical approaches to train and tune a genAI chatbot that reflects and represents these features and dispositions?

To explore question 1, we convened a series of virtual listening sessions with four groups: 1) educational and AI experts from across the US; 2) teachers; 3) families; and 4) a combined group of teachers and families, a subset of the original teachers and family members. Most teachers and families were affiliated with Riverbank Elementary School[1] in a district that is a member of the Council of Great City Schools in an urban neighborhood of a large city in the Midwest. Question 2 was explored through an iterative technology design and development process that engaged off-the-shelf open source technologies in dialogue with the requirements and narratives offered by the participants to result in a final (beta-level) system.

Research Methods

The listening sessions included two expert groups (one with 3 experts and another with 2), one group of 5 teachers, one group of 6 family members, and a combined session with 2 teachers and 2 family members. The Expert Sessions was aimed at identifying the opportunities and risks of the project's efforts to support culturally relevant practices in STEM teaching. The sessions with teachers and families also sought out opportunities and risks, but was guided by a values-centered design research methodology specifically tailored for AI (Calo et al., 2021). These sessions were designed to invite ideas for the chatbot's features, dispositions, and use cases.

Expert Sessions

Experts were selected based on their extensive knowledge in AI, education, culturally relevant frameworks, or a combination thereof. They were invited through existing academic networks or newly established connections. Initially planned as a single session, scheduling constraints necessitated two Expert Sessions, conducted via Zoom.

[1] River Bank Elementary School, and all geographic identifiers, are pseudonyms.

Participants represented diverse geographic locations across the United States (West Coast, Midwest, and East Coast), various academic roles, and intersectional identities. **Table 1** provides a detailed demographic overview of the Expert Session participants.

Session	Role	Expertise	Self-Ascribed Identities
1	Computer Science Faculty	Computer Science/AI	South Asian Man
	Learning/Computer Sciences Doctoral Student	Education and Computer Science/AI through Culturally Relevant Frameworks	Black Woman
	Teacher Education/Educational Technology Faculty	Educational Technology and Teacher Education through Culturally Relevant Frameworks	South Asian Man
2	Teacher Education Faculty	Teacher Education and Computer Science Education through Culturally Relevant Frameworks	Black Woman
	Information Design Faculty	AI through Ethical and Culturally Relevant Frameworks	White Woman

Table 1. Expert Session participants.

The sessions employed a semi-structured focus group format, enabling targeted data collection while allowing flexibility to explore topics in depth or make connections across themes. Each session opened with an icebreaker question about participants' first memorable interaction with AI-based technology. The discussions then centered on three key areas: potential data sources for AI training; identifying strengths and supports for teachers' growth; and opportunities and limitations of current AI technology.

Educator and Family Sessions

Listening sessions with educators and family members were held in-person at Riverbank Elementary. Several individuals in these groups were members of an Educator Advisory Board and a Community Caucus connected with a previous project. Table 2 provides demographic information for the educator and family member participants.

Session(s)	Role	School Affiliation	Self-Ascribed Identities
1	Educator	Riverbank Elementary	Black Woman
1	Educator	Riverbank Elementary	Black Woman
1	Educator	Hawthorn Charter High	White Man
1 & 3	Educator	Riverbank Elementary	Black Woman
1 & 3	Educator	Riverbank Elementary	White Woman
2 & 3	Family Member	Riverbank Elementary	Latine Woman
2 & 3	Family Member	Riverbank Elementary	Black Woman
3	Family Member	Riverbank Elementary	Black Woman
3	Family Member	Riverbank Elementary	Latine Woman
3	Family Member	Ashford Middle	Latine Woman

Table 2. Educator and Family Member participants.

Sessions 1 (only Educators) and 2 (only Family Members) were oriented around a semi-structured process that touched upon the same three key areas. As an icebreaker activity, however, following the values-centered design (Hendry et al., 2021) culturally relevant AI methodology (Calo et al., 2021), participants wrote things they want teachers to know about their students and what they need to succeed on a stone, which we referred to as Insight Stones. Writing on a stone imbues importance and value to these ideas symbolizing permanence (Calo et al., 2021). During Session 3, the participating Educators and Family Members were able to revisit and organize the Insight Stones into categories that made sense to them. They then used the Insight Stones as a framework to create Stories of AI in the Classroom. They then shared and we discussed these stories together as a group.

Data Analysis

The data was analyzed to sustain quality and rigor, particularly with respect to credibility through peer debriefing and member checks, dependability through the maintenance of an audit trail and a code-recode strategy, and confirmability through triangulation and practice reflexivity (Anfara et al., 2002). As a participatory design research effort (Cumbo & Selwyn, 2022; Design-Based Research Collective, 2003), the thematic identification process was oriented towards practical criteria that can be applied to the design and development of the genAI chatbot (Easterday et al., 2018; Edelson, 2002) in a manner to contributes to increased agency and voice by marginalized individuals and communities and educational justice (Bang & Vossoughi, 2016).

Expert Sessions. The two listening sessions were held via Zoom, recorded, and uploaded to a Kaltura-based system (Kaltura Corp, 2024) maintained by the university. All members of the research team took ethnographic field notes and wrote research memos based on the experience (Emerson et al., 1995). The listening sessions were machine-transcribed and human-corrected. Each member of the research team then wrote analytic research memos (Emerson et al., 1995) of each listening session. All artifacts were then discussed and coded thematically to identify salient cross-cutting themes over several rounds.

Family and Educator Sessions. The recorded dialogue was analyzed in the same fashion as the Expert Sessions, although these sessions generated additional data. Sessions 1 and 2 resulted in Insight Stones and Session 3 resulted in Stories of AI in the Classroom. The Insight Stones were analyzed through qualitative content analysis (Mayring, 2000); they were categorized thematically over several rounds of categorization and recategorization to identify salient ideas across all participants. Unique contributions by individual participants were also identified. The Stories of AI in the Classroom were analyzed through a narrative analytic approach, with a focus on opportunities for genAI in the context of teacher practice (Fina, 2021; Gazan, 2005) and as contributing to the feature set of the genAI chatbot (Alvarez & Urla, 2002).

Findings

As a result of the expert panel, tensions emerged including:

- **Accuracy vs. Errors**, a concern endemic to genAI that while a chatbot's output is useful, it can include numerous errors that look real (referred colloquially as "hallucinations");
- **Efficiency vs. Critical Thinking**, that the genAI may facilitate increased efficiency and speed in creating educational materials, but lessen the teacher's tendency to question the purposes, enactments, and outcomes of the learning experiences;
- **Ideal Case vs. Reality**, that the genAI may generate outputs that are aimed at an ideal case of a classroom that does not exist, rather than the actual dynamic and diverse conditions of real classrooms in real schools; and similarly,
- **Local Wisdom vs. Universal Assumptions**, that despite the intents, given genAI's tendencies to create outputs that convey broad, generalized, and surface-level statements, locally- and culturally-situated knowledge and practices, particularly in STEM, will be further hidden from view in the classroom.

The teachers and families, who are closer to the students and their communities, sometimes reinforced the experts' ideas and concerns, and at other times viewed them as opportunities. The Insight Stones were typically student-focused and positive (**Figure 1**).

Figure 1. Example Insight Stones.

The teachers and families emphasized:

- **Providing assistance in planning and set-up**, supporting teachers with relevant ideas for activities that would provide sustained engagement with their students;
- **Instilling in the teacher a genuine love and respect for students**, honoring each individual student for who they are, where they live, and what they carry with them;
- **Attending to the socioemotional needs of individual students**, such as engaging students who may need to talk about events in or out of school in a supportive conversation, alerting the teacher of the student's emotional state;

- **Managing auxiliary and mundane classroom tasks**, such as pulling up presentations for teachers, facilitating station activities, or completing attendance, freeing the teacher up to teach; and
- **Providing the time and space for teachers to engage with students** in a more direct, one-on-one manner to meet their individual cognitive, socioemotional, and cultural experiences and needs.

In essence, the Experts, the Educators, and the Family Members pointed to two overarching behaviors they would like to see the AI accomplish: 1) making things easier and more efficient for the teacher for auxiliary tasks; and 2) acting as a thought partner to support the teacher in thinking critically about their work. Both overarching behaviors are oriented toward providing the time and support to meet and address the cognitive, sociocultural, and linguistic needs and backgrounds of the students in their classroom, treating them as whole human beings.

Training Data Identification Process

Based on data collected and analyzed from the Listening Sessions, we were identified categories of research literature that intersected with the concerns and opportunities identified by the participants. Literature was selected from these categories to train the AI. The categories were: Universal Design for Learning, Multilanguage Learning, Culturally Relevant and Sustaining Pedagogy and Curricula, Critiques of the STEM Pipeline, Caring and Ethics, and Cultural Approaches to STEM Learning and Teaching. For example, literature in the Caring and Ethics category was primarily centered on the ethics of care perspective (Kohn, 2023; Noddings, 1988), as well as caring orientations of critical pedagogies (Flores & Alfaro, 2022). Every effort was made to include literature that addressed structural, relational, and individual perspectives in each category.

Summaries of each article were written, with explicit acknowledgement of the authors, to avoid copyright infringement issues. These summaries were provided as training data for the LLM. In addition, all summaries, research memos, and Stories of AI in the Classroom, carefully edited to ensure deidentification, were provided as training data.

Technology Innovations

An early prototype of the AI chatbot is complete. The development of the current version involved leveraging a custom script and an augmented dataset to fine tune the open-sourced Llama 3.1 8B (Touvron et al., 2023) model to generate culturally inclusive and relevant STEM lesson plans, refine conversation guidance, and address knowledge gaps to optimize performance.

We started by developing chat format templates by defining structures for system instructions, user inputs, and model outputs, ensuring that our responses are culturally relevant and context-aware in educational contexts. Drawing on research in culturally responsive pedagogy, we designed these templates to guide the model in addressing

diverse classroom scenarios. On the Ollama platform, we customized the model file to run LLMs locally for educational use cases, such as lesson planning, ensuring the system messages promoted cultural relevance.

We also designed and tested prompt scripts with the Llama 3.1 8B model to help teachers create inclusive lesson plans. The custom script inquired about classroom context, lesson goals, and educational standards, prioritizing cultural inclusivity. Furthermore, we created a structured dataset of question-answer pairs from research paper summaries to fine-tune the model on culturally relevant pedagogy, using tools like LangChain and Ollama LLM.

Lastly, we explored the application of Retrieval-Augmented Generation (RAG) (Gao et al., 2023), an advanced AI technique that combines information retrieval with text generation to produce more accurate and contextually relevant responses. The development process for our RAG model involved several key steps. First, we created external data specifically for designing and testing the model. This process entailed extracting text and metadata from various documents and converting this information into vector embeddings, numerical representations of text that capture semantic meaning. These embeddings were then stored in a PostgreSQL database, creating a vector database for our model to reference.

The RAG model functions by enhancing user input through a combination of relevant retrieved information (chunks) with the original query. This approach allows the model to generate responses that are more contextually accurate. During our rigorous testing phase, we ensured that the model's responses were not only accurate but also aligned with our intended educational outcomes.

Our implementation of the RAG model demonstrated several significant advantages. Notably, it enhanced the AI's cultural responsiveness, showing an improved ability to generate culturally inclusive responses. The implementation also offers flexibility and scalability, allowing for adaptability to different contexts and potential for future growth. By combining information retrieval and text generation, our RAG model produces higher quality and more accurate responses compared to traditional language models.

One of the key benefits of using RAG is its ability to incorporate up-to-date information. Instead of relying solely on the pre-trained knowledge of the underlying Large Language Model (LLM), our RAG model retrieves current information from its vector database. This approach significantly reduces the problem of "hallucinations" in AI, the generation of false or nonsensical information that can occur in traditional language models.

Furthermore, our RAG implementation enhances the credibility and transparency of the generated information. The model can cite sources for specific details, providing a clear trail of information provenance. By maintaining metadata of the augmented data, we ensure transparency in the information's origin and processing.

The RAG model implementation enhanced the AI's ability to generate culturally relevant and accurate lesson plans. This approach offers a promising direction for educational

technology, combining the strengths of information retrieval and text generation to improve the quality and accuracy of AI-generated responses in educational contexts.

Conclusion

This research has illuminated both opportunities and challenges in implementing genAI chatbots in educational settings. Moving forward, we aim to navigate the tensions highlighted by experts and more diligently incorporate the perspectives of teachers and families in future iterations of our genAI chatbot, while maintaining robust and accurate STEM content.

Through our research, we identified three key insights:

1. **Methodological Efficacy:** The value sensitive design methodology (Calo et al., 2021; Hendry et al., 2021), when combined with other participatory reflexive, critically-oriented approaches (Bang & Vossoughi, 2016; Fine et al., 2021), proves to be a robust framework for comprehensively understanding the implications of AI in the classroom. This integrated approach allows for a nuanced examination of both opportunities and concerns.
2. **Dual Role Potential:** The tension participants identified in genAI between performing auxiliary tasks and serving as a thought partner presents a promising avenue for further exploration. This duality offers an opportunity to investigate and address the tradeoffs, unequal distributions, and sociocultural ecological effects of genAI in educational contexts (Postman, 1998).
3. **Technological Advancement:** Retrieval-Augmented Generation (RAG) emerges as a significant technological trajectory for addressing current limitations of genAI. Importantly, RAG offers a mechanism to explicitly incorporate and acknowledge educator and community voices within genAI chatbot applications, potentially enhancing their relevance, focus, and accountability.

These insights will guide our ongoing work in co-constructing genAI applications. Our goal is to support educators in creating more inclusive, culturally relevant, and sustaining STEM learning environments that are deeply rooted in local contexts. By addressing the identified challenges and leveraging the opportunities, we aim to develop AI tools that promote teacher growth and development in designing and facilitating inclusive, culturally relevant and sustaining, and locally focused STEM learning environments.

References

Alvarez, R., & Urla, J. (2002). Tell me a good story: Using narrative analysis to examine information requirements interviews during an ERP implementation. *SIGMIS Database*, *33*(1), 38–52. https://doi.org/10.1145/504350.504357

Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative Analysis on Stage: Making the Research Process More Public. *Educational Researcher*, *31*(7), 28–38.

Bang, M., & Vossoughi, S. (2016). Participatory Design Research and Educational Justice: Studying Learning and Relations Within Social Change Making. *Cognition and Instruction*, *34*(3), 173–193. https://doi.org/10.1080/07370008.2016.1181879

Calo, R., Friedman, B., Kohno, T., Almeter, H., & Logler, N. (2021). *Telling Stories: On Culturally Responsive Artificial Intelligence*.

Cumbo, B., & Selwyn, N. (2022). Using participatory design approaches in educational research. *International Journal of Research & Method in Education*, *45*(1), 60–72. https://doi.org/10.1080/1743727X.2021.1902981

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 5–8.

Easterday, M. W., Rees Lewis, D. G., & Gerber, E. M. (2018). The logic of design research. *Learning: Research and Practice*, *4*(2), 131–160. https://doi.org/10.1080/23735082.2017.1286367

Edelson, D. C. (2002). Design Research: What We Learn When We Engage in Design. *Journal of the Learning Sciences*, *11*(1), 105–121. https://doi.org/10.1207/S15327809JLS1101_4

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing Ethnographic Fieldnotes*. University of Chicago Press.

Fina, A. D. (2021). Doing narrative analysis from a narratives-as-practices perspective. *Narrative Inquiry*, *31*(1), 49–71. https://doi.org/10.1075/ni.20067.def

Fine, M., Torre, M. E., Oswald, A. G., & Avory, S. (2021). Critical participatory action research: Methods and praxis for intersectional knowledge production. *Journal of Counseling Psychology*, *68*, 344–356. https://doi.org/10.1037/cou0000445

Flores, J., & Alfaro, A. R. (2022). Critical pedagogy: Loving and caring within and beyond the classroom. *Curriculum Inquiry*, *52*(3), 385–396. https://doi.org/10.1080/03626784.2022.2072665

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv Preprint arXiv:2312.10997*.

Gazan, R. (2005). Imposing structures: Narrative analysis and the design of information systems. *Library & Information Science Research*, *27*(3), 346–362. https://doi.org/10.1016/j.lisr.2005.04.004

Habig, B., Gupta, P., & Adams, J. D. (2021). Disrupting deficit narratives in informal science education: Applying community cultural wealth theory to youth learning and engagement. *Cultural Studies of Science Education*, *16*, 509–548. https://doi.org/10.1007/s11422-020-10014-8

Hendry, D. G., Friedman, B., & Ballard, S. (2021). Value sensitive design as a formative framework. *Ethics and Information Technology*, *23*(1), 39–44. https://doi.org/10.1007/s10676-021-09579-x

Kaltura Corp. (2024). *Kaltura EDU Solutions*. Kaltura. https://corp.kaltura.com/education-video-platform/

Kohn, A. (2023, May 4). *The Caring Subversion of Nel Noddings: An Appreciation*. https://nepc.colorado.edu/blog/nel-noddings

Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2). http://www.qualitative-research.net/index.php/fqs/article/view/1089

Noddings, N. (1988). An Ethic of Caring and Its Implications for Instructional Arrangements. *American Journal of Education*, *96*(2), 215–230. https://doi.org/10.1086/443894

Postman, N. (1998). *Five Things We Need to Know About Technological Change*. https://www.cs.ucdavis.edu/~rogaway/classes/188/materials/postman.pdf

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., & others. (2023). Llama: Open and efficient foundation language models. *arXiv Preprint arXiv:2302.13971*.

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, *8*(1), 69–91. https://doi.org/10.1080/1361332052000341006