From Search Engine to Thought Partner: How Do Teachers Use AI Chatbots?

Ana T. Ribeiro(✉), Benjamin Leiva, Carly Robinson, Chris Agnew, and Susanna Loeb

Stanford University, Stanford, CA 94305, USA
{anactr,bleivas,carly_robinson,agnewc,sloeb}@stanford.edu

Abstract. Chat-based interfaces have transformed how people interact with AI, generating a rapid expansion of chatbots and human–AI interaction data; yet little is known about how educators use these tools. We study teachers' use of AI chatbots designed for education, analyzing more than 15,000 threads created over one month by more than 4,400 teachers on SchoolAI.com, an AI-powered platform for K–12 educators. Although the platform offered 10 assistants—nine specialized and one multi-purpose—most teachers relied primarily on the multi-purpose assistant, and only 22% initiated five or more threads. To analyze prompt content, we use the "Asking", "Doing", and "Expressing" taxonomy to classify teacher prompts by intent, and a taxonomy of education domains to classify prompts into six different categories of topics. More than half of prompts are "Doing", followed by 31% "Asking" and 15% "Expressing". Across intents, curriculum and instructional content is the most common topic, accounting for more than 40% of prompts. Together, these findings suggest that most teachers used AI in limited but practical ways, primarily for task automation and curriculum-related support rather than sustained strategic integration.

Keywords: Human–AI interactions · Teachers · Intent

1 Introduction

Chat-based generative AI tools have rapidly entered K–12 education, changing how teachers access information, plan instruction, and respond to classroom challenges. Conversational interfaces built on large language models lower barriers to use by allowing educators to work with AI through extended dialogue embedded in everyday practice. Yet we still know little about how teachers actually use these tools. Existing evidence offers limited visibility into how teachers engage with AI, capturing perceptions and intentions through self-reports [1,2] or by focusing mainly on task-oriented instructional support [3,4].

Survey and interview studies show that teachers use AI to generate lesson plans, assessments, rubrics, and other instructional materials, positioning it as a productivity aid or pedagogical co-designer [3,5]. Interaction logs provide a

E. G. Blanchard et al. (Eds.): AIED 2026, CCIS 3031, pp. 38–44, 2026.
https://doi.org/10.1007/978-3-032-29788-4_6

complementary source of evidence by capturing what teachers choose to use and what they type, how exchanges unfold across turns, and how purposes shift within a conversation.

This study analyzes naturally occurring teacher interactions with AI chatbots designed for K–12 education and hosted on the platform SchoolAI.com, an AI-powered platform for K–12 educators. Using chat logs from more than 4,400 teachers who created over 15,000 conversation threads in one month, we examine which chatbots teachers choose, how they structure interactions, and what they aim to accomplish. We classify teacher prompts by intent and topic, drawing on frameworks proposed by [6] for categorizing user intent and [4] for education-specific content domains.

We find that although the platform offered 10 chatbots ranging from topic-specific to multi-purpose assistants, most teachers gravitated toward the multi-purpose chatbot. While many prompts seek guidance on curriculum and content, most ask the AI to perform tasks. Teachers also engage in reflective and expressive exchanges at higher rates compared to those reported for work-related uses of ChatGPT [6], suggesting that AI functions not only as a drafting assistant or search engine, but also as a thought partner for teachers.

Research Questions: We examine teachers' intensity of use of the platform's AI chatbots and the content of their prompts to answer:

1. How much do teachers use each chatbot, and how often do they switch between topic-specific and multi-purpose assistants?
2. What are teachers' most common purposes for interacting with AI chatbots, and what topics are most common?

2 Methodology

2.1 Data

Data for this study comes from a research collaboration with SchoolAI, a large AI-powered platform that provides automation and personalization tools for K–12 educators. During the study period, roughly 50,000 educators visited the platform. Because users may disclose personal information about themselves or their students in chatbot interactions, the platform de-identified and scrubbed all personally identifiable information (PII) before sharing records with the research team. Due to the cost of this process, the platform limited the research data to one month of activity, October 2024.

The full sample includes 260,318 user messages (prompts) submitted by 4,597 teachers across 16,158 conversation threads initiated during this month.[1] During October 2024, the platform offered 10 teacher-facing AI chatbot assistants, each

[1] A conversation thread is defined as a discrete conversation tab containing a sequence of prompts and responses. A new thread begins when a user starts a blank conversation; a single thread may span multiple topics.

fine-tuned with content and prompt instructions for professional K–12 use. One served as a general-purpose educator aide, while the other nine were specialized for specific domains or tasks, including research support and essay grading, project-based learning or field trip planning, student-facing chatbot design, and question answering related to curriculum, the Common Core, AI in education, and digital literacy.

The dataset includes an identifier for the AI chatbot associated with each thread and a timestamp for thread creation, but not timestamps for individual messages. As a result, we cannot observe when users returned to and reused an existing thread after its initial creation.

To reduce noise from early exploratory use and focus on teachers with more sustained engagement, we excluded single-day users whose only activity occurred on the day they created an account (3.8% of teachers). The final analytic sample contains 15,640 threads from 4,422 teachers and 251,558 total user messages.

2.2 Message Classification

To classify teacher prompts by intent, we use the framework proposed by [6], sorting prompts into three categories: *Asking, Doing*, and *Expressing*. For topic classification, we used the six top-level domain categories defined by [4]: *Instructional Practices, Assessment and Feedback, Student Needs and Context, Professional Responsibilities, Curriculum and Content Focus,* and *Other*. Descriptions of intent-domain combinations can be found in Table 1.

We used a few-shot prompting strategy with Gemini 2.5 Flash (incorporating LearnLM), providing human-labeled examples and up to 10 prior messages of context. The model returned a classification, confidence score, and justification.

3 Results

3.1 Teacher Use of AI Assistants

We find substantial variation in teacher engagement with AI assistants in our sample. Although most teachers started only a couple of threads with a single assistant and submitted about ten messages over the course of a month, a smaller group of highly engaged teachers started dozens of threads and submitted thousands of messages to multiple assistants.

Figure 1 shows that more than two-thirds of teachers began the first observed thread with the multi-purpose assistant. Among the 2,578 teachers who started more than one thread (58%), 72.6% used a single assistant for all interactions, 21.8% used two, and 5.6% used three or more. Of these teachers, 60% used the multi-purpose assistant exclusively, while 18.4% used only topic-specific assistants. Therefore, most teachers either rely on one assistant for everything—usually the multi-purpose one—or stay within specialized tools rather than switching between the two.

Figure 2 shows the distribution of teacher messages per thread across assistants. Although the multi-purpose assistant was used most often, its median

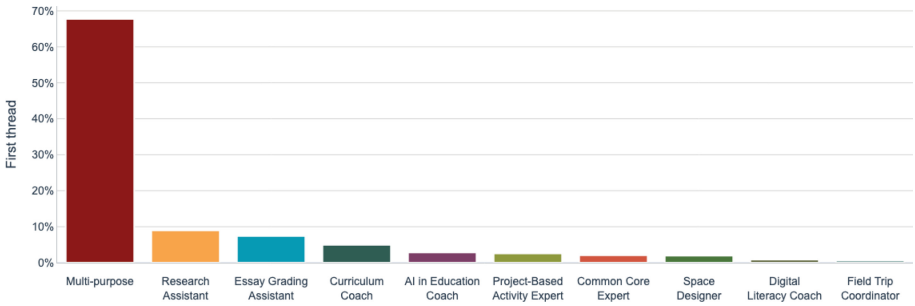

Fig. 1. Distribution of Teachers by First AI Assistant Interaction

thread contained only three teacher messages, placing it among the lowest of all assistants. The Essay Grading Assistant is the only one with threads reaching 10 teacher messages, likely because teachers submit one prompt per student for each graded assignment. Overall, the small number of messages per thread suggests limited iteration in most conversations, though a few outliers contain thousands of messages and reflect sustained exchanges.

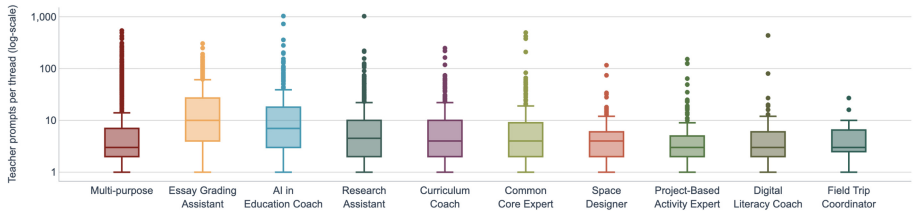

Fig. 2. Teacher messages per chat thread, by AI assistant. The y-axis shows the number of messages sent by teachers in each chat with a specific chatbot, on a logarithmic scale due to highly skewed data.

3.2 Teacher-AI Conversation Intent and Topic

We combine intent and topic into a single view in Fig. 3. Analyzing teacher messages across intent and topic area, we found that most teacher prompts ask the AI assistant to create something. Just over half of all messages were classified as "Doing", meaning teachers requested that the AI generate lesson plans, assessments, feedback, or other materials. The remainder were split between "Asking" (31%)—seeking information or advice—and "Expressing" (15%)—sharing thoughts without requesting action. These estimates are much closer to those reported in [6] for work-related uses of ChatGPT (Asking: 35%; Doing: 56%; Expressing: 8%) than overall uses (49%; 40%; 11%), reflecting the platform's

positioning as a professional tool for teachers. However, one noteworthy difference between ChatGPT work-related use and our teacher platform is the share of Expressing prompts, which is 8% and 15%, respectively. This difference suggests that teachers may have a distinctive demand for a thought partner, a role AI assistants may be well-suited for.

Across intent categories, prompts related to the "curriculum and content" domain are the most common: roughly two out of every five messages relate to what to teach or how to align materials with standards. This finding is consistent with the idea of AI assistants as "curriculum co-planners", helping teachers unpack standards, generate examples, and revise tasks so they are more accessible to students. Assessment and Feedback, Instructional Practices, and Other each account for 11–17%, suggesting teachers also use AI for grading support and concrete classroom moves.

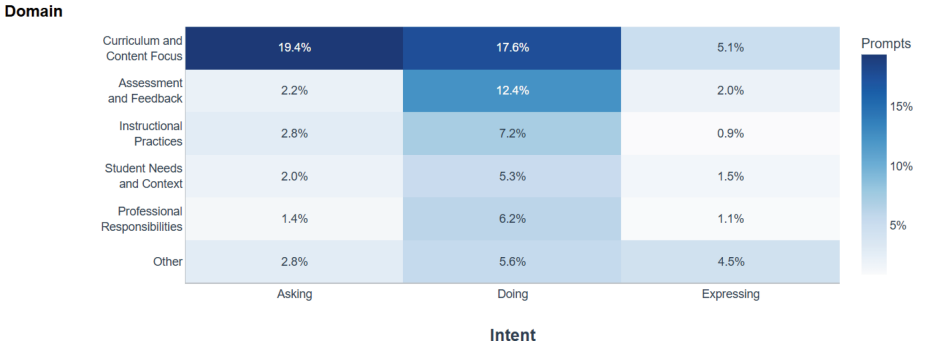

Fig. 3. Message Distribution across Teacher Intent & Pedagogical Domain Notes: This heatmap shows the distribution of messages across pedagogical domains (y-axis) and intent categories (x-axis). Cell values show the percentage of total messages. Darker blue indicates higher concentrations. See the appendix for definitions. (Color figure online)

We also analyze the flow of teacher-AI assistant conversations and document transitions between topics in Fig. 4. The regularity of cross-domain shifts suggests teachers are using SchoolAI chat tools across multiple layers of their work. The transition analysis highlights a curriculum- and content-centric workflow: four of the five most common cross-topic transitions lead into Curriculum and Content Focus at least 25% of the time. These patterns suggest teachers in multi-turn conversations with AI assistants move fluidly across curriculum, instruction, assessment, and student needs within a single conversation. More than half of messages are driven by creation needs ("Doing"), but a substantial share involves seeking information (particularly around curriculum and content). In practice, this means that AI assistants are functioning simultaneously as reference tools and drafting partners within the same conversation.

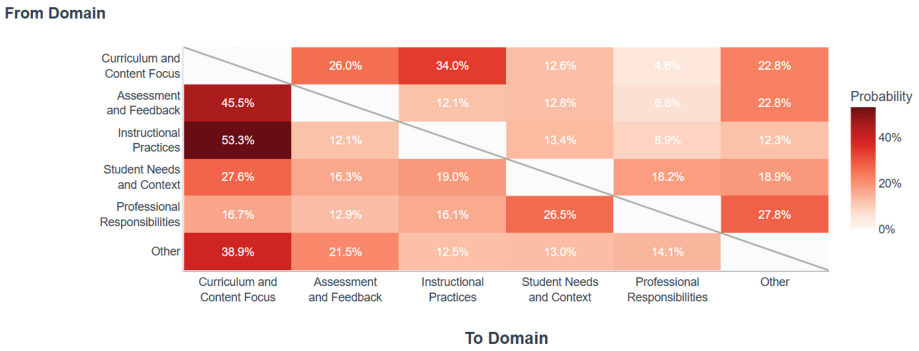

Fig. 4. Topic Transition Patterns Within Conversations. Notes: This heatmap shows the probability of transitioning from one topic (y-axis) to another domain (x-axis) when teachers change topics within a conversation. Cell values show the percentage probability. Darker red indicates more common transitions. Diagonal line marks self-transitions (not counted in analysis). (Color figure online)

Table 1. Taxonomy of teacher prompts by domain and intent

Instructional Practices: Asking = teaching strategies, discussion, checking understanding; Doing = activities, directions, modeled explanations; Expressing = reflections or feelings about instruction
Curriculum and Content Focus: Asking = content, standards, concepts, alignment; Doing = content-aligned materials, problems, texts, lesson outlines; Expressing = reflections on alignment, rigor, texts, tasks, or units
Assessment and Feedback: Asking = checks for understanding, student work, rubrics, assessment; Doing = quizzes, rubrics, feedback, grading; Expressing = workload, consistency, or feedback effectiveness
Student Needs and Context: Asking = differentiation, engagement, behavior, inclusion, classroom situations; Doing = tailoring for ELLs, IEPs, behavior needs, or local contexts; Expressing = concerns about students, climate, participation, or motivation
Professional Responsibilities: Asking = conferences, collaboration, meetings, norms; Doing = emails, meeting notes, summaries, artifacts; Expressing = role, workload, wellbeing, burnout, growth, or goals
Other: Asking = general questions, tool use, curiosity; Doing = formatting, generic drafting, troubleshooting, non-teaching tasks; Expressing = general thoughts, emotions, tiredness, curiosity, or chatting

4 Conclusion

Much like other users of AI assistants for work-related tasks, most teachers rely primarily on a multi-purpose assistant to perform or automate tasks and seek information, particularly around curriculum and content. Although teacher interactions with AI assistants can be relatively light or sporadic, this type of use could help teachers more easily align their teaching with standards, and access and generate high-quality instructional materials faster. Moreover, a substantial share of teacher-AI interactions involve offloading time-consuming tasks or sharing reflections about challenges and successes. In practice, this means that AI assistants are functioning simultaneously as reference tools, drafting partners, and reflective sounding boards within the same platform.

For researchers and practitioners, these findings suggest several directions for future work. One avenue involves linking patterns in teacher–AI interactions to measures of instructional quality or student outcomes, in order to better understand when and how AI use is most productive. Another is to examine how usage varies across school contexts, subject areas, and teacher characteristics, and whether particular supports help broaden access to AI's potential benefits. Qualitative work—such as interviews, classroom observations, and design partnerships with teachers—will also be necessary to interpret the intent and impact of teacher-AI conversation data, especially in sensitive domains like student needs and teacher wellbeing.

Acknowledgments. This study was made possible through a collaboration with SchoolAI.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Angelone, L., Burton, B.: MagicSchool: an exploration of the use of generative AI by southwest Ohio teachers. Ohio J. Teach. Educ. **39**(1), 4 (2024)
2. Esbenshade, L., et al.: Emerging patterns of GenAI Use in K-12 science and mathematics education. arXiv preprint arXiv:2509.10747 (2025)
3. Kerr, R.C., Kim, H.: From prompts to plans: a case study of pre-service EFL teachers' use of generative AI for lesson planning. Engl. Teach. **80**(1), 95–118 (2025)
4. Liu, A., et al.: How K-12 educators use AI: LLM-assisted qualitative analysis at scale. arXiv preprint arXiv:2507.17985 (2025)
5. Handa, K., et al.: Which economic tasks are performed with AI? Evidence from millions of claude conversations. arXiv preprint arXiv:2503.04761 (2025)
6. Chatterji, A., et al.: How people use ChatGPT (No. w34255). National Bureau of Economic Research (2025)