

9

Generative AI as a teaching assistant

Ryan S. Baker, Xiner Liu, Mamta Shah, Maciej Pankiewicz, Yoon Jeon Kim, Yunseo Lee, Chelsea Porter

University of Pennsylvania, United States

This chapter examines how AI Teaching Assistants driven by generative AI can expand and support the traditional roles played by human Teaching Assistants (TAs) and instructors in higher education. AI TAs are designed to automate routine administrative tasks, deliver real-time student support, and provide timely formative feedback at scale in ways that complement and extend human capabilities. Rather than replacements for human TAs, we argue for a model of augmentation AI TAs provide support and feedback that is otherwise infeasible for human instructors and TAs alone: support and feedback that is immediate, personalised, detailed, and available around the clock. The presentation and discussion of JeepyTA, a generative AI system, helps ground these ideas. We also discuss key design and implementation considerations.

Introduction

Much of the recent focus on generative artificial intelligence (GenAI) in education has considered it as a tool used individually either by a student or an instructor, through a commercial off-the-shelf chatbot designed to be an assistant. This role has considerable value (see Gašević and Yan, 2026^[1]; Li and Hu, 2026^[2]), but GenAI can be used in several other fashions, as shown in this report. This chapter considers the role of GenAI in supporting instructors and Teaching Assistants (TAs). TAs serve as a foundational support structure within colleges and universities, bridging the gap between students and academic staff and helping to sustain the quality of instruction in a wide range of disciplines. By leading discussion sections, grading assignments, answering questions, and offering individualised guidance, TAs play a vital role in shaping how students experience their courses. This dual role as both an intermediary and a mentor underscores the significance of the work that TAs do, not just in managing course logistics but in advancing student learning, engagement, and success. As higher education evolves – facing increasing enrolment pressures, budget constraints, and shifts toward online or hybrid delivery – the role of TAs is likely to expand. At the same time, TAs are themselves students, balancing these teaching responsibilities with their own scholarly activities, which can lead to time conflicts, uneven support for learners, and burnout. These challenges raise important questions about how universities can sustainably leverage the benefits TAs provide while addressing the real human limitations of time, expertise, and scalability.

Recent advances in GenAI offer a promising avenue to complement TAs' efforts, while retaining the human element that underpins great teaching. GenAI-driven "teaching assistants," powered by large language models (LLMs) and other advanced technologies, have the potential to handle repetitive administrative tasks, deliver targeted learning support, and provide immediate feedback to students in a way that human TAs alone simply cannot. By examining the evolving role of TAs and exploring how AI can enrich and extend their capabilities, this chapter seeks to highlight both current practices and new horizons for more equitable, accessible, and impactful teaching support. We conclude with a discussion of implications, including for policy.

Teaching assistants

The teaching assistant role

A Teaching Assistant (TA) in higher education (in countries where this role exists) is typically a graduate (Master's and PhD) or advanced undergraduate (Bachelor's) student who supports the primary instructor in delivering course content and assisting students. A TA's responsibilities may vary by institution and discipline, but they generally include facilitating small-group discussions, answering student questions, and supplementing the main instructor's efforts to create an effective learning environment (Park, 2004_[3]). In recent years, as more courses have tended to partly or fully move online, TAs also play a large role in managing discussion forums, answering student questions and supporting discussions there (Wadams and Schick-Makaroff, 2022_[4]). By handling a portion of the teaching and administrative workload, TAs play a critical role in making large or complex courses more manageable, thus enhancing the overall educational experience for both academic staff and students.

The origins of Teaching Assistants can be traced back to the late nineteenth century, when growing student enrolments and expanding research expectations prompted universities to look for ways to extend instructional capacity (Chism, 1998_[5]). In these early stages, TAs often served as informal helpers to more senior academic staff, assisting with tasks like grading or lab supervision as part of their own apprenticeship in academia. Over time, and particularly after World War II when higher education systems expanded rapidly, the role of TAs became increasingly formalised. Universities began creating structured programs that provided clearer job responsibilities, training, and professional development opportunities, reflecting the recognition that TAs could significantly enhance both teaching and learning. This evolution laid the foundation for the modern TA role, in which graduate and advanced undergraduate students are systematically integrated into the educational process (Park, 2004_[3]).

Teaching Assistants carry out a range of tasks that collectively support both the instructor and students in higher education settings. Their responsibilities can be broadly categorised into instructional support and student engagement, though these two areas naturally overlap (Park, 2004_[3]). By taking on these roles, TAs help to foster an environment that promotes understanding, participation, and continuous feedback - key components of effective learning (Hattie and Timperley, 2007_[6]; Chi and Wylie, 2014_[7]).

In terms of instructional support, TAs frequently lead discussion sections, tutorials, or laboratory sessions, serving as facilitators who bridge theory and practice (Park, 2004_[3]). In these smaller and often more interactive settings, TAs clarify course material, demonstrate practical techniques, and encourage student participation. By adapting teaching methods to the needs of specific groups of students, TAs help maintain a dynamic and inclusive classroom atmosphere. Another critical element of a TA's role involves assessing student work. TAs often grade assignments, quizzes, and exams under the supervision of the lead instructor (Marshman et al., 2018_[8]). This process typically includes reviewing submissions, providing constructive feedback, and highlighting areas for improvement - ideally, guiding students to developing and demonstrating deeper understanding (Marshman et al., 2018_[8]). Grading student work not only eases faculty workload but can also offer TAs valuable, instructor-scaffolded experience in evaluating academic performance, helping TAs to develop deeper understanding of student thinking (Karim, Maries and Singh, 2018_[9]).

As for student engagement in their learning process, in many modern courses - particularly those with hybrid or fully online components - TAs serve as key points of contact on discussion forums (Wadams and Schick-Makaroff, 2022_[4]). By responding to questions, facilitating conversation, and sharing clarifications from the instructor, they help maintain an active and supportive online learning community. This work often extends to moderating peer-to-peer exchanges, ensuring that discussions remain on topic and respectful. In addition, TAs often hold regular office hours and meet with students to allow students to seek in-depth explanations, review feedback, or discuss academic challenges (Park, 2004_[3]). These mechanisms often provide learning support beyond what instructors can offer just-in-time and on-demand, particularly for research-active senior academic staff or large courses.

Teaching Assistants provide a *range of benefits* to higher education institutions, more senior academic staff, and students. Universities often find that using TAs is a cost-effective means of managing large course enrolments while still providing individualised support to students, a topic of constant interest to administrators when university budgets are under pressure. For more senior academic staff, TAs offer substantial advantages by relieving some of the workload associated with teaching, grading, and administrative duties. By delegating tasks such as discussion facilitation, assignment feedback, and routine course management, faculty members can devote more time to developing innovative curricula, advancing their research agendas, and mentoring students (including the TAs) at higher levels. In addition, TAs often introduce diverse perspectives or novel approaches to instruction, encouraging

a collaborative environment in which both senior academic staff and TAs refine teaching strategies (Begley et al., 2019_[10]). Finally, students also benefit significantly from the involvement of TAs. In many cases, TAs are more available to answer questions outside of regular class times, and their support on online discussion forums can be accessed asynchronously, providing a flexible option for students who need extra help. TAs' relative proximity to the student experience - whether by age, academic journey, or shared disciplines - can also result in a peer-mentorship-like atmosphere during office hours and informal interactions. As such, TAs' greater relatability and students' perception that they are more understanding (Kendall and Schussler, 2012_[11]) can ease anxieties and foster a sense of community, ultimately enhancing the overall learning experience.

However, several challenges have been noted for current practices involving Teaching Assistants. Regarding instructional support, many TAs may for example lack pedagogical training or skill (Nicklow, Marikunte and Chevalier, 2007_[12]). This lack of formal preparation can undermine the quality of their instruction, as they may be uncertain about how to present information clearly. Moreover, TAs sometimes adopt a surface-level approach to grading, focusing on relatively simplistic aspects of correctness rather than attempting to provide feedback that guides students toward deeper conceptual understanding (Marshman et al., 2018_[8]). Compounding these issues is the fact that TAs typically possess less subject-specific expertise than full academic staff (for example, see Karim et al., 2018_[9]), which can limit their ability to answer complex questions or provide advanced guidance.

Furthermore, TAs often face challenges around workload and time constraints. Many TAs must balance teaching responsibilities with personal academic obligations, such as coursework, research projects, and preparation for required examinations (Park, 2004_[3]). Some TAs may find it difficult to invest the necessary time in class preparation, grading, or providing substantive feedback to students. This overload can also lead to high stress and exhaustion, reducing their effectiveness as a TA while also impairing their other work and personal success (Berta and Pembroke, 2019_[13]). This is exacerbated by the uneven training and faculty support provided to TAs. Some instructors involve TAs extensively in designing lesson plans, assessment rubrics, or instructional materials, while others may provide only minimal training and mentorship (Begley et al., 2019_[10]). Furthermore, many TAs do not have access to teaching mentors other than the instructor. This lack of support can leave TAs uncertain about expectations or best practices, making it harder for them to support students without spending large amounts of their time.

Hence, while TAs fulfil crucial roles in supporting learning and engagement, and help bridge the gap between senior academic staff and students, there are several aspects in which current practices are not optimal for either TAs or students.

Review of AI teaching assistants (AI TAs) and their use

Could generative AI (GenAI) technologies improve things for both TAs and students? The emergence of GenAI models has created a great deal of enthusiasm for the potential of a wide range of educational benefits. One area of rising interest has been the creation of AI Teaching Assistants—tools that extend the capabilities of human TAs. AI Teaching Assistants (AI TAs) use computational methods – in many recent cases GenAI (Sajja et al., 2024_[14]; Yetişensoy and Karaduman, 2024_[15]; Liu et al., forthcoming_[16]) but, before that, machine learning and previous-generation natural language processing (NLP) (Khosravi et al., 2021_[17]; Shermis and Burstein, 2003_[18]) – to perform tasks that were typically in the purview of human TAs, although in many cases beyond what was feasible for human TAs. Their scope has included streamlining routine administrative tasks, providing targeted learning support or rapid formative assessment, and empowering human TAs and instructors with information and insights about their students, enhancing the overall effectiveness of instructional delivery. While they sometimes take over what used to be human tasks, these systems are not intended to replace human educators, but instead to provide support 24/7 and free up valuable time for TAs and instructors to focus on more complex, critical, and high-impact aspects of teaching.

Over the past few decades, the use of AI in educational technology has evolved from simple automation tools – like basic quiz generators and grading scripts (see discussion in Brooks, 2023_[19]) – to advanced AI systems that can process language and consider context, allowing them to respond to students' questions and needs in real time. The previous generation of intelligent tutoring systems and question answering systems could offer sophisticated support (Nye, Graesser and Hu, 2014_[20]; Goel and Polepeddi, 2018_[21]), but was highly expensive to author (Aleven et al., 2006_[22]), often necessitating intense focus on only a single aspect of adaptivity (Baker, 2016_[23]). The contemporary use of GenAI, sometimes combined with previous-generation machine learning, creates the potential for a qualitative leap forward in functionality and sophistication, at much lower development cost.

These technologies, when used to complement instructors, can take on repetitive tasks – such as answering common questions and some parts of the assessment of student work – thereby providing immediate, round-the-clock support to learners and eliminating the bottleneck that often occurs when TAs or instructors

are not available – for instance, for an online learner working from a different time zone than the instructors. At the same time, human TAs and instructors can allocate their expertise to higher-order pedagogical activities, such as facilitating in-depth discussions, offering mentorship, and providing customised feedback for unusual cases and learning challenges. This synergy ultimately helps institutions maintain quality education at scale, addresses the labour-intensive aspects of teaching, and ultimately supports instructors in finding time for high-value personal interaction with learners. In the following section, we will discuss some of the ways that AI TAs can support learners, human TAs, and instructors.

Key functions and use cases of AI-based teaching assistants (AI TAs)

Administrative and logistical support is an area of responsibility for current human TAs that is easy and fairly non-controversial to replace. By automating more administrative processes such as course enrolment or the monitoring of completion of assignments, instructors can devote more time to pedagogical planning and personalised student engagement (Ahmad et al., 2022_[24]). AI TAs can also manage course communications by sending out timely reminders for assignments, examinations, and events, ensuring students remain informed and minimising the risk of missed deadlines (Pardo et al., 2018_[25]). Streamlining these tasks can allow human instructors and TAs to focus on higher-level teaching responsibilities, such as curriculum development and individualised feedback. Beyond routine administrative tasks, AI systems can support course logistics by monitoring student participation and promptly alerting instructors to potential engagement and performance issues (Wise and Jung, 2019_[26]; Khosravi et al., 2021_[17]), as well as distil insights from discussions on course forums for instructors.

Another potential area of application for AI TAs is *instructional support*. AI Teaching Assistants can provide supplementary explanations or resources tailored to students' individual needs. For example, if a student expresses confusion about a specific topic, these systems can supply targeted materials, such as a textual explanation, brief video tutorials, interactive modules, or suggested readings (Sajja et al., 2024_[14]; Essel et al., 2022_[27]; Yetişensoy and Karaduman, 2024_[15]). Such a system can provide more attention to customising learning for a given student than would be feasible for even the most dedicated human Teaching Assistant. By engaging in follow-up discussion, an AI TA based on a chatbot can support a student in ways that would be infeasible with a static resource. In addition, just as current GenAI can recommend resources to a learner, it can also assist educators with content curation and lesson planning, suggesting how to communicate topics more effectively (Karpouzis et al., 2024_[28]).

Furthermore, AI TAs can support instructors in *evaluating* students' progress, particularly when it comes to formative assessment. Summative assessment through AI has been used in some applications but still needs to meet a higher bar for reliability and fairness (Bulut and Beiting-Parrish, 2024_[29]). Formative assessments to inform instructors or support learners can be used safely due to the lower stakes. There is a long history of using automated assessments; decades of work assessed learners with automated quizzes and multiple-choice items (Brooks, 2023_[19]) and a previous generation of NLP afforded short-answer grading (Gao et al., 2024_[30]) and automated essay scoring (Shermis and Burstein, 2003_[18]). Previous work also enabled the generation of new items, through procedural templates for instance (Kurdi et al., 2019_[31]). However, the advent of generative AI has made it possible both to generate new items in sophisticated, tailored ways (Yadav and Tseng, 2023_[32]; IIITD et al., 2024_[33]; Sajja et al., 2024_[14]) and to offer much more detailed, rich feedback on complex artifacts created by students (Chiang et al., 2024_[34]; IIITD et al., 2024_[33]; Stahl, Biermann and Wachsmuth, 2024_[35]; Teng et al., 2024_[36]). Automated announcement tools can then deliver personalised updates to students, supporting both performance and self-regulation (Lim et al., 2021_[37]). By making it feasible to offer detailed feedback in a much timelier fashion, students are likely to revise their understanding and adapt in ways that align with the course goals (Graham, 2023_[38]).

AI Teaching Assistants embedded in course platforms can effectively serve as a first point of contact for students, promptly addressing frequently asked questions about the syllabus, assignment deadlines, and other logistical concerns (Liu et al., forthcoming_[16]). By referencing a structured knowledge base, these systems can also respond to content-related inquiries, offering supplementary explanations or clarifications (Liu and M'Hiri, 2024_[39]; Maiti and Goel, 2024_[40]; Miroyan et al., 2025_[41]; Teng et al., 2024_[36]; Liu et al., forthcoming_[16]). Questions can be asked in an external platform (Alsafari et al., 2024_[42]; Teng et al., 2024_[36]), within the course discussion forum (Liu et al., forthcoming_[16]; Miroyan et al., 2025_[41]; Haaben, 2024 (Unpublished Manuscript)_[43]), a learning management system (LMS) (Maiti and Goel, 2024_[40]), or in the context of a learning activity itself (Lee et al., 2023_[44]). These tools can save considerable time for instructors and human TAs (Sinha, 2024_[45]; Miroyan et al., 2025_[41]).

More importantly, the immediate, round-the-clock availability of these tools supports learners who may need help outside of conventional office hours; for example, one study of an AI TA embedded into a discussion forum found that students received responses significantly more often on weekends with the AI TA than during the previous (only human TA support) semester (Liu et al., forthcoming_[16]). While such a system cannot respond to all student

queries, it can provide support in many cases. In other cases where a query requires more nuanced interpretation or context—such as complex conceptual misunderstandings or unique personal circumstances—AI TAs can escalate the matter to human TAs, thus ensuring students receive appropriate and thorough support. This triaging function can help manage the flow of incoming questions, reducing the volume of simpler queries that human TAs and instructors must handle. As a result, educators are freed to spend more time providing personalised feedback, guiding higher-level discussions, and engaging students in meaningful academic interactions. While this type of question answering functionality was available even before the advent of LLMs, it required considerable engineering (e.g. Goel and Polepeddi, 2018_[21]) compared to the relative ease of deployment now possible.

Key considerations

One of the key steps to moving these types of advancements from one-off research projects to scalable solutions benefitting a large number of learners is ensuring they integrate seamlessly with existing educational infrastructures. Many AI-based tools of this type so far require learners and instructors to use separate platforms rather than being integrated directly into their primary learning management systems (LMS) or discussion forums. This lack of integration or interoperability can create a fragmented user experience, requiring additional sign-ins, duplicating data entry, and making it harder to track student progress across multiple systems. In contrast, compatibility with widely used LMSs (e.g. Canvas, Moodle, Blackboard) and discussion forum platforms (Piazza, Discourse, phpBB, vBulletin, Flarum) would allow AI TAs to seamlessly access course materials, participation records, and student performance data. Such interoperability not only streamlines the user experience but also supports richer analytics and more effective, personalised interventions, ultimately strengthening the teaching and learning process.

Another key step for making these systems usable at scale will be efforts to engineer the human-computer interactions of these systems to facilitate their use by busy human TAs and instructors. Currently, the process of integrating course resources varies in complexity between tools, and the degree of uptake can vary considerably between instructors (Maiti and Goel, 2024_[40]). There are several ways to accomplish this, including shared folders, access to learning management systems as a simulated student, or tools for uploading resources - but whichever approach is chosen, it must be low-effort for human beings. It should also be easy to continually update these resources, as changes to course materials and syllabi will often occur within a semester and across semesters for courses that are offered on a regular basis.

In addition, onboarding and even training is needed for the human TAs and instructors who will collaborate with an AI TA. They will need to understand enough about how the system works, what it can do, and what its limitations are, to ensure that they implement it effectively in their courses. By clearly communicating which queries or tasks the AI TA should handle versus those that call for human expertise, institutions can maintain quality control while maximising efficiency. Over time, incorporating continuous improvement and feedback loops can further refine these boundaries. For instance, platforms should support human TAs and instructors in regularly reviewing the AI TA's responses to student inquiries. Student and instructor feedback collected through short surveys or mining forum discussions can also highlight areas where the AI TA might be underperforming or producing confusing or inaccurate information. Supporting instructors in checking and refining the system will help to ensure that content remains accurate, relevant, and aligned with educational objectives, reduce instructor frustration, and increase the likelihood of long-term sustained use.

Case study analysis: The JeepyTA platform in universities

JeepyTA is an example of a course-specific, AI-driven Teaching Assistant designed to integrate with existing classroom and online practices. Developed by the Penn Center for Learning Analytics at the University of Pennsylvania (UPenn) and launched in Fall 2023 (Liu et al., forthcoming_[16]), JeepyTA utilises a multi-turn conversational architecture of large language models (LLMs) and is not bound to a specific model – it can be configured to run on many LLMs (e.g. GPT, Llama, or DeepSeek). In courses where JeepyTA has been used, a recent OpenAI GPT model has been used (starting with GPT-3.5 Turbo, moving to GPT-4, GPT-4-Turbo and GPT-4o).

JeepyTA has been used in various scenarios: to deliver responses to logistics questions, contextually respond in discussions based on provided course materials, provide targeted feedback to written assignments and coding problems, and to serve as a brainstorming partner. As of Spring 2025, JeepyTA has been deployed across 16 sessions of 14 courses at three higher education institutions in the USA (with deployments also running later in 2025 in Singapore and Colombia). This widespread adoption reflects the growing interest in AI systems that can free human instructors and TAs from repetitive logistical duties, while still delivering responsive, round-the-clock support for learners.

To ensure course alignment, JeepyTA is primed with instructors' chosen reference materials, including syllabi, textbooks, readings, and past instructor feedback. These resources are embedded in the system's knowledge base through a retrieval-augmented generation (RAG) workflow: newly uploaded documents are converted into embedding vectors, enabling JeepyTA to retrieve contextually relevant information via semantic search. In doing so, JeepyTA can address administrative queries – such as answering date-related questions from the syllabus – while simultaneously leveraging contextual readings to stimulate in-depth discussions on course-specific topics. Additionally, through collaboration with instructors, in several cases, JeepyTA's prompts have been iteratively refined to better address specific learning objectives. Separate models have been employed to automate decisions on whether responses appear immediately or await instructor approval, providing finer control over JeepyTA's engagement in forum discussions. Finally, JeepyTA's behavior can be customised by category of tasks, giving instructors the flexibility to choose which topics or discussion forum categories it responds to and with what level of human supervision (human-in-the-loop).

Across courses, JeepyTA has been integrated into the open-source Flarum platform, appearing as a forum user distinctly marked as an AI Teaching Assistant. Through a Progressive Web App, the forum is accessible on mobile devices, allowing students and instructors to stay engaged on-the-go. In addition to traditional email notifications, users of a mobile app can receive push alerts – such as when JeepyTA responds or specifically mentions them – ensuring timely updates and facilitating faster interaction within the discussion forum.

Answer logistics questions

One of JeepyTA's primary functions is answering logistics questions about the course. At the beginning of the semester, JeepyTA can handle enrollment-related inquiries, including prerequisites, add/drop deadlines, and options for changing course registration. When students need accommodations, JeepyTA directs them to official university guidelines and relevant support services. It also provides information on class schedules, classroom locations, and changes due to holidays or special events. When a course has multiple sections, JeepyTA helps students confirm where they need to attend.

To support coursework, JeepyTA clarifies submission guidelines for assignments, specifying required file formats, submission portals, and deadlines. JeepyTA also assists with technical aspects of online learning platforms when required by the course. It helps students log into external platforms used by instructors (for instance, for video discussions) and provides information like login codes, platform access links, and usage instructions. If students experience submission errors or other technical problems, JeepyTA offers guidance in many cases without needing to involve the instructor.


JeepyTA helps students understand grading policies by explaining how grades are calculated based on rubrics, weighted components, and participation requirements. It also assists in interpreting feedback from instructors and TAs and guiding students on resubmissions, appeals, or grade disputes (see Figure 9.1). When students need access to course materials, JeepyTA provides links to lecture slides, reading repositories, and virtual meeting links, ensuring they have the necessary resources.

With recent updates, JeepyTA can remember instructor responses and announcements on recurring topics. If students ask about schedule changes, assignment deadlines, or policy updates, JeepyTA provides the latest information. This reduces confusion and keeps students informed without requiring instructors to repeat themselves.

It is worth noting that JeepyTA's ability to answer logistics questions depends on the information instructors choose to provide. It does not generate responses based on general knowledge but instead pulls from course-specific details that instructors input. If a detail was not provided, JeepyTA directs students to the human TA (if available for the course) and instructor or official course documents rather than guessing or giving incomplete information.


JeepyTA's performance in answering logistics questions is not always perfect, as some student inquiries may go beyond what is covered in the course materials. Therefore, instructors can choose to edit JeepyTA's responses at any time instead of only choosing between fully accepting or discarding them. This option is especially useful when combined with the feature that allows instructors to review JeepyTA's response before it becomes visible to other students (as explained below). This flexibility allows instructors to keep useful parts, make quick edits, and provide students with accurate information while reducing effort.


Figure 9.1. JeepyTA guiding students in making up for part of the assignment


 [Student A] Oct 19, 2023

I am wondering if we're supposed to post 4 times for each creative assignment, or what's the grading criteria for this? I only posted once for each CA and got 1/4. Is there any way that I can make up for it?

Thank you!

 JeepyTA replied to this.

 **JeepyTA** Oct 19, 2023

 [Student A] @ [Student A] Hello!

I just wanted to clarify something regarding the grading criteria for the Creative Assignments. According to the instructions, you are expected to submit your own assignment on time, as well as provide meaningful comments on at least four other students' submissions.

If you have only posted once for each Creative Assignment and received a score of 1/4, it might be a good idea to reach out to Ryan or the TAs. They can guide you on how to make up for the missed posts and provide information on the specific requirements and options available to you.

It's always important to fully understand the grading criteria and make sure that you are meeting all the expectations. So, if you have any questions or concerns, don't hesitate to seek clarification.

Best,
Jeepytee (Virtual Teaching Assistant)

Source: Authors' own work.

Provide feedback to student essays

In several courses, JeepyTA provides feedback on student essays based on the grading rubric specified for the assignment. This consists of both higher-level conceptual elements and aspects of writing. In terms of higher-level conceptual elements, JeepyTA evaluates essays on the goals of the assignment, such as whether (for example) the student has appropriately discussed the needs of stakeholder groups, whether the student has made arguments in terms of theories discussed in class, or whether the limitations of a proposed solution have been concretely detailed (see Figure 9.2). When students submit drafts, JeepyTA acknowledges what they do well and highlights their strengths.

JeepyTA also offers feedback on more mechanical aspects of writing such as argument clarity, evidence use, structure, and writing quality (see Figure 9.2). In addition, JeepyTA comments on lower-level details such as language use, unclear phrasing, grammar mistakes, and wordiness. In these cases, it suggests revisions that can preserve the student's original intent.

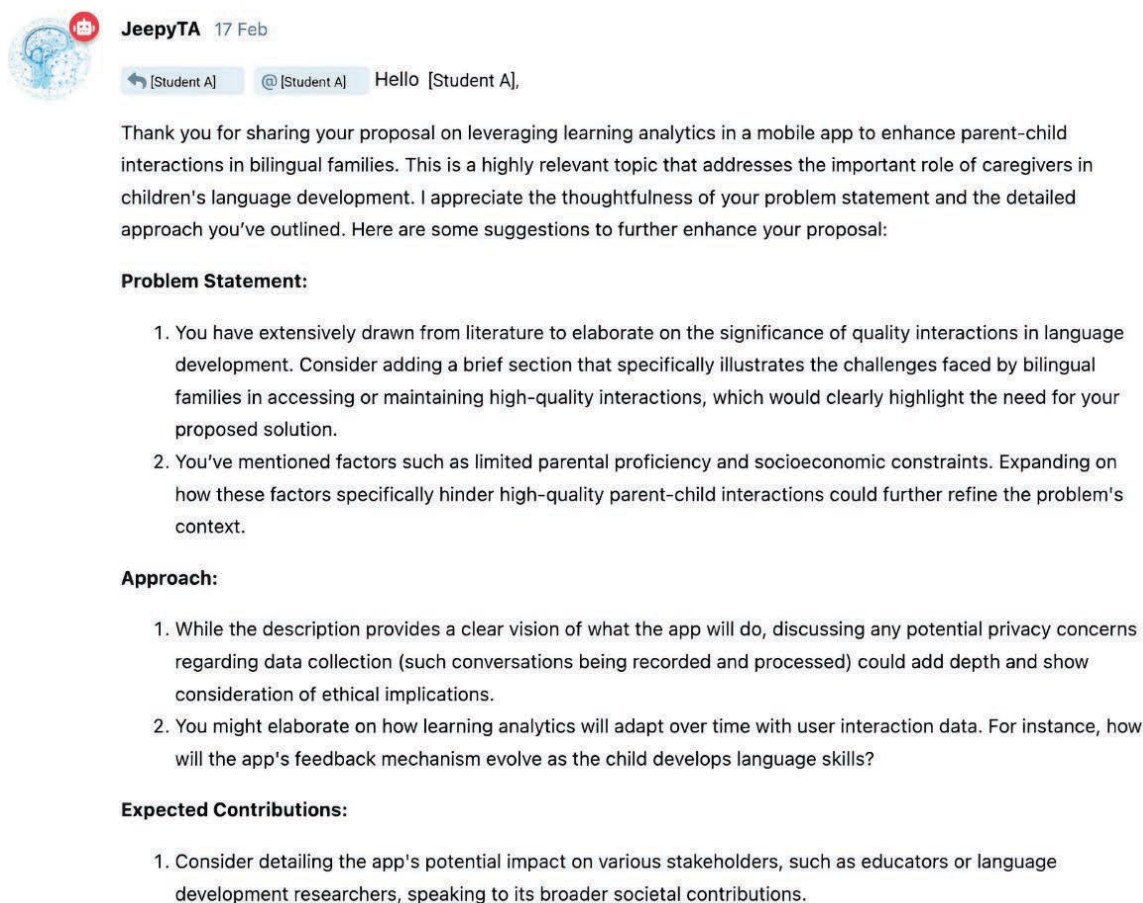
GPT models are trained to provide generic responses that apply across many scenarios, which can make their default feedback vague or overly general. To prevent this, JeepyTA was instructed to "provide actionable insights rather than shallow suggestions". This small detail within prompt engineering makes a difference in helping students receive concrete guidance that improves their revisions.

If students need clarification, they can ask follow-up questions, and JeepyTA refines its guidance based on those questions. Instructors can also adjust JeepyTA's feedback settings to focus on specific aspects of writing or emphasise areas where students generally struggle the most.

Before asking JeepyTA to provide feedback on essays, its responses are first tested on a set of sample essays and the output is reviewed with instructors. This step helps confirm that the feedback aligns with the pedagogical goals of the course. When necessary, the prompts are refined based on the instructor's suggestions in the review process. This process helps JeepyTA provide comments that are clear, relevant to the assignment, and focused on the aspects instructors consider most important. It also creates an opportunity to catch cases where the LLM's default knowledge base produces inaccurate info, such as in cases where much of the content on the web reflects an incorrect understanding of a specific technical point.

In some cases, the prompt specifies a particular tone to shape the feedback style. For example, JeepyTA can be instructed to provide concise and direct feedback or take a more encouraging and supportive tone. This allows the feedback to align with the way instructors and TAs typically communicate with students about their writing. Additionally, past feedback from previous course offerings, along with de-identified student essays from matching pairs, are included in some cases as a reference for JeepyTA. JeepyTA does not use the content of past essays as a source for feedback but instead looks at these examples to follow the structure, level of detail, and key focus areas that instructors and TAs have emphasised. This helps make the feedback more useful to students by reflecting the expectations and priorities set in previous iterations of the course.

Figure 9.2. JeepyTA providing feedback on the first step in a student essay assignment – the essay prospectus.



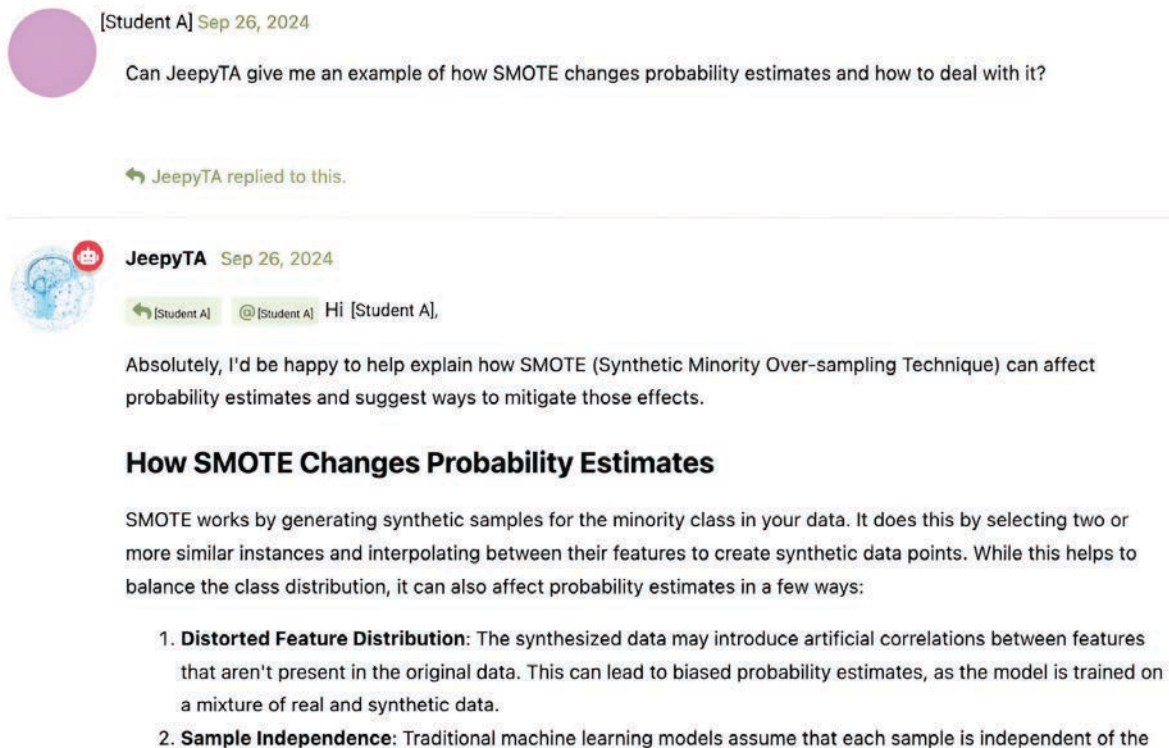
Source: Authors' own work.

Respond to student reflections/questions on the reading/lecture

JeepyTA is also capable of responding to student reflections and questions on the course readings and lectures, offering additional clarification, prompting further thinking, and connecting ideas across course materials. When students share reflections, JeepyTA acknowledges their contributions by reinforcing key ideas from the readings or connecting their insights to broader course themes. If a student raises a question about a concept, theory, or method, JeepyTA provides a response by summarising relevant arguments, explaining terms, or pointing to sections of the readings that address the issue (see Figure 9.3). When a reflection introduces an interesting perspective or critique, JeepyTA may pose follow-up questions to encourage further discussion. To maintain consistency between JeepyTA's responses and the course content, JeepyTA was specifically instructed to first reference the course materials, with specific materials selected based on their similarity score to the student's query, while replying. JeepyTA was also prompted to use course-specific language as defined by instructors before the start of the semester in its responses. If a student's question is not closely related to the course, JeepyTA may be instructed to rely on its knowledge base to respond.

Instructors or TAs can modify the visibility settings of JeepyTA's responses at any time during the semester. If preferred, JeepyTA's responses can be flagged for instructor review before being shared with the students. This option can be turned on or off at any time during implementation. It is also available for selected categories, such as only for answering logistics questions. This helps prevent the provision of incorrect or misleading information, a particular issue in subject areas where misconceptions are highly present on the web and therefore also in the LLM knowledge base. Instructors or TAs can review flagged responses. If a response is inaccurate, they can discard it and reply directly. If the response is mostly correct but needs refinement, they can edit it before posting. When a response is accurate and well-structured, the instructors or TAs can approve it as is. This additional layer allows JeepyTA to provide timely support while making sure students receive information that is accurate, relevant, and aligned with the course objectives.


Figure 9.3. JeepyTA explaining a key detail about an algorithm



[Student A] Sep 26, 2024

Can JeepyTA give me an example of how SMOTE changes probability estimates and how to deal with it?

↩ JeepyTA replied to this.

 JeepyTA Sep 26, 2024

↩ [Student A] @ [Student A] Hi [Student A],

Absolutely, I'd be happy to help explain how SMOTE (Synthetic Minority Over-sampling Technique) can affect probability estimates and suggest ways to mitigate those effects.

How SMOTE Changes Probability Estimates

SMOTE works by generating synthetic samples for the minority class in your data. It does this by selecting two or more similar instances and interpolating between their features to create synthetic data points. While this helps to balance the class distribution, it can also affect probability estimates in a few ways:

- 1. Distorted Feature Distribution:** The synthesized data may introduce artificial correlations between features that aren't present in the original data. This can lead to biased probability estimates, as the model is trained on a mixture of real and synthetic data.
- 2. Sample Independence:** Traditional machine learning models assume that each sample is independent of the

Note: SMOTE is an algorithm used for data pre-processing.

Source: Authors' own work.

Provide debugging support to programming code

For courses involving programming (but not focused on learning to programme), JeepyTA also provides debugging support to students working on assignments by analysing their code and identifying potential errors. When students submit code snippets or describe issues they encounter, JeepyTA reviews the logic, syntax, and structure to pinpoint common mistakes. It then suggests corrections or improvements to resolve the errors (see Figure 9.4).

JeepyTA is generally encouraged to use its existing programming knowledge to solve coding issues while following course-specific conventions or practices. In an Educational Data Mining course, for example, student-level cross-validation is required because this method evaluates how well a model generalises to unseen students. A general LLM chatbot may default to recommending a simple train-test split, a technique that would be acceptable in other contexts but is not the method needed in this course. To prevent this, JeepyTA is instructed to prioritise debugging support based on course materials, assignment requirements, and instructor guidelines rather than relying on broadly used techniques that may not be appropriate in the course context. When addressing programming errors, it refers to the course's preferred approaches and explains why they are used and how they differ from other methods.

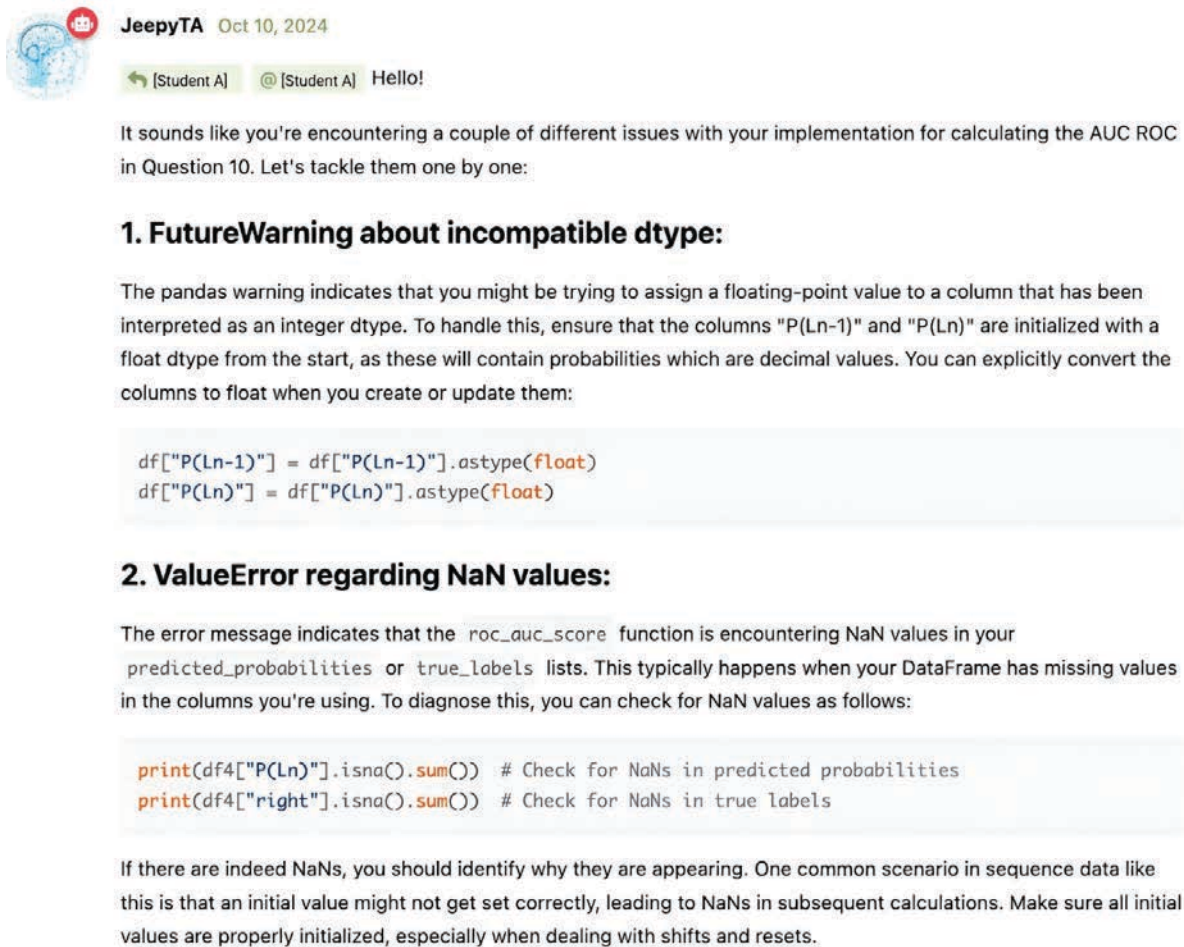
In cases where code produces unexpected output, JeepyTA offers strategies for troubleshooting, such as adding print statements, checking variable values, or breaking down complex functions into smaller, testable parts. If students describe the problem rather than submitting code, JeepyTA suggests debugging techniques based on the nature of the issue and guides them through potential causes and solutions.

If students describe the problem vaguely or provide incomplete context, JeepyTA asks follow-up questions to clarify the issue before offering suggestions. For example, if a student says, "The code isn't working," without specifying the error message or expected output, JeepyTA prompts them to provide more details, such as the error message received, the intended function of the code, or the steps they have already tried.

When providing programming code support, JeepyTA's prompts are designed to avoid simply providing the correct code but instead focus on helping students understand how to diagnose and fix errors themselves. The prompt design encourages students to learn from their own debugging process, which creates opportunities for them to build confidence and capacity to read errors, trace code, and solve problems independently rather than rely on being given the exact fix.

JeepyTA has provided debugging support for two courses across two semesters, but its effectiveness in identifying errors has been limited in some cases. One possible reason is that it did not have access to the datasets students were working on in either implementation, which makes it difficult for JeepyTA to verify data structures, variable values, or dataset-specific errors. On the other hand, in other cases, it has caught unusual mistakes (such as a student using the symbol `\` instead of `|`) and typos which can be difficult for instructors and human TAs to see in a lengthy programme (see Figure 9.4). Even when JeepyTA is unable to pinpoint the exact cause of issues, it has helped students clarify their problems and suggests general debugging strategies. This still reduces the time instructors or TAs need to spend guiding students through the initial steps of troubleshooting.

Figure 9.4. JeepyTA explaining the errors the student encountered while solving a programming problem and giving advice for diagnosis



JeepyTA Oct 10, 2024

[Student A] @ [Student A] Hello!

It sounds like you're encountering a couple of different issues with your implementation for calculating the AUC ROC in Question 10. Let's tackle them one by one:

1. FutureWarning about incompatible dtype:

The pandas warning indicates that you might be trying to assign a floating-point value to a column that has been interpreted as an integer dtype. To handle this, ensure that the columns "P(Ln-1)" and "P(Ln)" are initialized with a float dtype from the start, as these will contain probabilities which are decimal values. You can explicitly convert the columns to float when you create or update them:

```
df["P(Ln-1)"] = df["P(Ln-1)"].astype(float)
df["P(Ln)"] = df["P(Ln)"].astype(float)
```

2. ValueError regarding NaN values:

The error message indicates that the `roc_auc_score` function is encountering NaN values in your `predicted_probabilities` or `true_labels` lists. This typically happens when your DataFrame has missing values in the columns you're using. To diagnose this, you can check for NaN values as follows:

```
print(df4["P(Ln)"].isna().sum()) # Check for NaNs in predicted probabilities
print(df4["right"].isna().sum()) # Check for NaNs in true labels
```

If there are indeed NaNs, you should identify why they are appearing. One common scenario in sequence data like this is that an initial value might not get set correctly, leading to NaNs in subsequent calculations. Make sure all initial values are properly initialized, especially when dealing with shifts and resets.

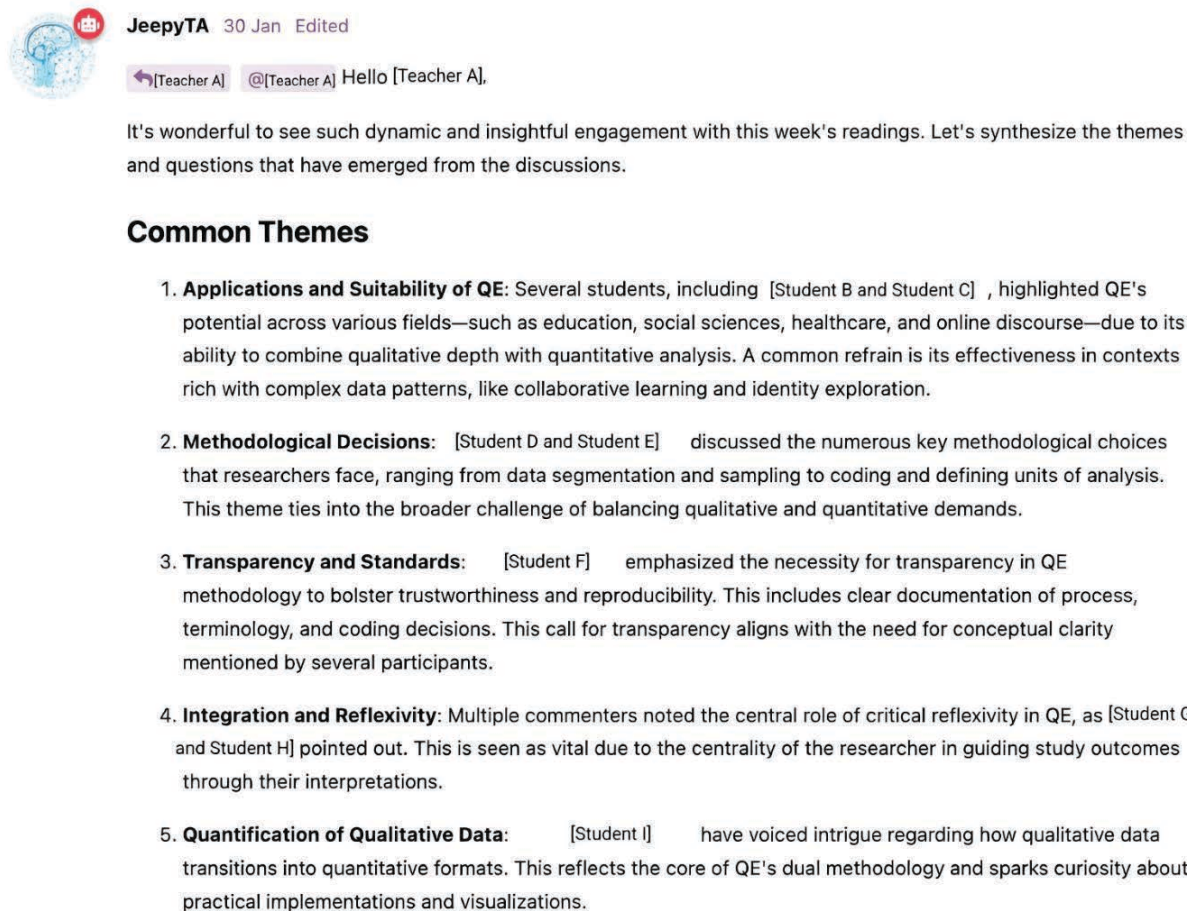
Source:Authors' own work.

Summarise the discussion posts for each week

JeepyTA has also been applied to generate summaries of discussion forum conversations. When the use case was first introduced in Spring 2024, summaries were provided only to instructors and TAs to give them an overview of the student discussions. As of Spring 2025, in some courses, these summaries are accessible to all students on the forum. After each weekly discussion, JeepyTA summarises key themes, groups discussions into meaningful themes and recurring arguments, and identifies important questions from the discussion forum. In doing so, JeepyTA credits students who introduced specific points in order to give the instructor and TAs a sense of specific student participation as well as overall trends (See Figure 9.5).

The purpose of these summaries is not to replace reading or participating in discussions but to provide an additional layer of support in organising and reflecting on what was discussed. Students are still expected to engage in the full conversation, but the summary can help identify patterns, highlight areas of agreement and disagreement, and uncover questions that might require further discussion. In other words, instead of replacing direct engagement, the summary function can serve as a tool to make the overall direction of conversations more accessible.

Figure 9.5. JeepyTA summarising the weekly discussion within the forum



JeepyTA 30 Jan Edited

←[Teacher A] @[Teacher A] Hello [Teacher A],

It's wonderful to see such dynamic and insightful engagement with this week's readings. Let's synthesize the themes and questions that have emerged from the discussions.

Common Themes

- 1. Applications and Suitability of QE:** Several students, including [Student B and Student C], highlighted QE's potential across various fields—such as education, social sciences, healthcare, and online discourse—due to its ability to combine qualitative depth with quantitative analysis. A common refrain is its effectiveness in contexts rich with complex data patterns, like collaborative learning and identity exploration.
- 2. Methodological Decisions:** [Student D and Student E] discussed the numerous key methodological choices that researchers face, ranging from data segmentation and sampling to coding and defining units of analysis. This theme ties into the broader challenge of balancing qualitative and quantitative demands.
- 3. Transparency and Standards:** [Student F] emphasized the necessity for transparency in QE methodology to bolster trustworthiness and reproducibility. This includes clear documentation of process, terminology, and coding decisions. This call for transparency aligns with the need for conceptual clarity mentioned by several participants.
- 4. Integration and Reflexivity:** Multiple commenters noted the central role of critical reflexivity in QE, as [Student G and Student H] pointed out. This is seen as vital due to the centrality of the researcher in guiding study outcomes through their interpretations.
- 5. Quantification of Qualitative Data:** [Student I] have voiced intrigue regarding how qualitative data transitions into quantitative formats. This reflects the core of QE's dual methodology and sparks curiosity about practical implementations and visualizations.

Frequently Raised Questions

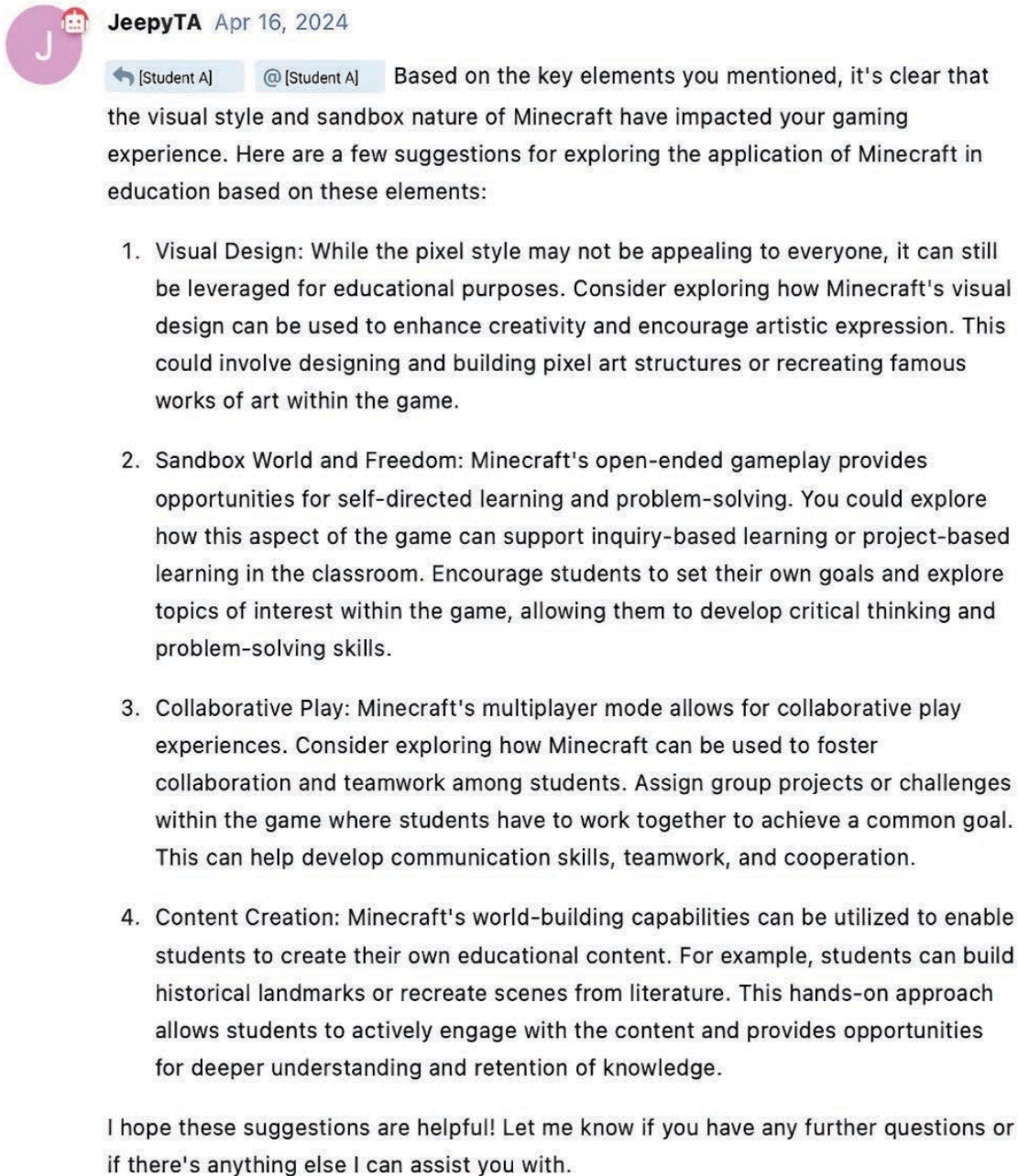
- How does data get sorted and make it past certain phases of screening once it passes identification?

Source: Authors' own work.

Suggest ideas for essays and support brainstorming

In a Games and Learning course offered in Spring 2024, JeepyTA was used as a brainstorming partner in two play journal assignments, where students consulted with JeepyTA to propose educational uses for classic games and Minecraft (Shah et al., 2024_[46]). A play journal is a structured reflective assignment in which students document and analyse their gameplay experiences to critically engage with classic and contemporary video games by examining their design, narrative, and educational potential. These journals help students develop knowledge of a game through both direct and vicarious experiences, providing a baseline understanding that enables them to propose meaningful educational applications. This structured reflection also prepares students with foundational knowledge of the game, which they can draw upon when consulting with JeepyTA – thus streamlining the labour-intensive process of game analysis and educational integration. While recent initiatives have aimed to make the repurposing of commercial entertainment games for education more accessible (Foster and Shah, 2020_[47]), the process still requires significant human effort. This poses challenges for students new to game-based learning and for instructors with limited resources to support student experimentation through direct experience (e.g. playing the game, learning through trial and error) or vicarious experience (e.g. watching YouTube videos, learning from colleagues). JeepyTA offers a solution by enabling users with varying levels of technological, pedagogical, and content knowledge to generate ideas, helping them explore how a game can be adapted for specific educational contexts and learning goals. Students can be guided in developing familiarity with a game and using it in dialogue with JeepyTA (See Figure 9.6).

Figure 9.6. JeepyTA supporting brainstorming and idea generation in a “games and learning” course



JeepyTA Apr 16, 2024

← [Student A] @ [Student A] Based on the key elements you mentioned, it's clear that the visual style and sandbox nature of Minecraft have impacted your gaming experience. Here are a few suggestions for exploring the application of Minecraft in education based on these elements:

1. **Visual Design:** While the pixel style may not be appealing to everyone, it can still be leveraged for educational purposes. Consider exploring how Minecraft's visual design can be used to enhance creativity and encourage artistic expression. This could involve designing and building pixel art structures or recreating famous works of art within the game.
2. **Sandbox World and Freedom:** Minecraft's open-ended gameplay provides opportunities for self-directed learning and problem-solving. You could explore how this aspect of the game can support inquiry-based learning or project-based learning in the classroom. Encourage students to set their own goals and explore topics of interest within the game, allowing them to develop critical thinking and problem-solving skills.
3. **Collaborative Play:** Minecraft's multiplayer mode allows for collaborative play experiences. Consider exploring how Minecraft can be used to foster collaboration and teamwork among students. Assign group projects or challenges within the game where students have to work together to achieve a common goal. This can help develop communication skills, teamwork, and cooperation.
4. **Content Creation:** Minecraft's world-building capabilities can be utilized to enable students to create their own educational content. For example, students can build historical landmarks or recreate scenes from literature. This hands-on approach allows students to actively engage with the content and provides opportunities for deeper understanding and retention of knowledge.

I hope these suggestions are helpful! Let me know if you have any further questions or if there's anything else I can assist you with.

Source: Authors' own work.

Personas in a course

In a course on cultural foundations for teaching and learning offered in Fall 2024, JeepyTA interacted with students through personas, discussing scenarios and stories related to their cultural experiences with teaching and learning.

The persona prompts for JeepyTA consisted of three main components: persona description, situation, and instruction. The persona description section defines the persona's name and role (e.g. Felipe, a teacher educator), the context in which the persona operates (e.g. teaching elementary education majors), and the personal and cultural background that reveals their funds of knowledge, such as home language, family activities, cultural rituals, and hobbies (Gonzalez, Moll and Amanti, 2006_[48]). The situation section contains the persona's role in the interaction (e.g. providing feedback, answering questions, mentoring), the task or topic being discussed (e.g. reviewing drafts on a specific subject), and the participants in the conversation (e.g. the persona interacting with a preservice teacher). Finally, the instruction section provides information regarding the tone of voice (e.g. formal, informal, supportive, critical), the level of detail required in responses (e.g. detailed feedback with personal experiences), and specific elements to include, such as relevant examples.

For each persona defined on the forum, a separate sub-forum category was created using the persona's name. In these categories, JeepyTA responded while acting as the corresponding persona. Students were informed about the personas and instructed that JeepyTA would post there as the persona indicated by the category name.

Since the main goal of this course is to situate novice preservice teachers in culturally relevant and sustaining teaching pedagogies (Ladson-Billings, 1995_[49]; Paris, 2012_[50]), the persona descriptions intentionally emphasise the cultural and linguistic backgrounds to avoid generating general responses that do not centre specific groups of learners. Thus, the personas were developed based on the lived experiences of four individuals who were members of historically underrepresented groups whom the team interviewed (e.g. Mexican American, Hmong American). These individuals also reviewed the draft persona descriptions to ensure that their identities were accurately portrayed and to avoid racial essentialisation (Omi and Winant, 2018_[51]).

For instance, one persona, 'Claire,' who identified as Hmong-American, shared personal stories related to food, spirituality, and family history, providing preservice teachers with concrete examples that fostered a deeper understanding of Hmong learners. Additionally, the personas guided preservice teachers in understanding concrete ways to incorporate funds of knowledge into classroom settings. For example, the persona 'Felipe' suggested specific strategies for adapting class materials to align with Mexican-American families, such as incorporating family tree activities and introducing home craft projects (see Figure 9.7). These approaches offered practical and culturally responsive methods for connecting multicultural perspectives to classroom instruction.

Generate a discussion prompt to start the weekly discussion

One use case being piloted in Spring 2025 is the generation of discussion prompts to start weekly discussions based on the assigned readings. The goal is to provide a foundation for meaningful discussions while maintaining relevance to the course objectives.

When generating the discussion prompt, JeepyTA extracts core arguments, methodologies, and debates from the week's readings. It identifies recurring themes, unresolved questions, or contrasting viewpoints and frames them in a way that encourages meaningful discussion. If students are expected to relate the reading material to their own research, JeepyTA includes questions that prompt reflection on personal experiences or future applications. If the goal is to explore methodological issues, it focuses on the strengths, limitations, and assumptions underlying the methods presented in the readings (see Figure 9.8).

Some prompts JeepyTA suggests may not be immediately useful. For example, it has generated questions that are overly broad and more at the level of the entire course than a specific week's content. Other discussion questions generated may be too complex and require extensive background knowledge or additional explanation before students can engage with them. A highly detailed methodological critique, for example, may be difficult to address within the scope of a discussion forum. Therefore, instructors and TAs have reviewed all of the discussion prompts before making them visible to students. JeepyTA's suggestions have provided a starting point that allows instructors and TAs to refine the wording, adjust the focus, or simplify overly technical questions to improve understandability. As such, JeepyTA does not replace instructor or TA expertise in orchestrating discussions, but rather, it streamlines the process by offering an initial draft that helps structure each week's forum.

Figure 9.7. JeepyTA acting as a Mexican American persona, “Felipe”, bringing a specific persona to recommendation



JeepyTA Nov 22, 2024

← [Student A] @[Student A] To incorporate 'funds of knowledge' effectively into your design project for teaching 5th-grade students, you need to first ensure that you are thoroughly familiar with the cultural, social, and economic backgrounds of your students. This understanding will allow you to tailor your project to reflect and utilize the knowledge and skills that students bring from their home environments. Here are a few detailed suggestions based on my experiences as a student:

1. **Cultural Relevance:** When I was in 5th grade, classroom materials that included or were relevant to Mexican culture could have helped me engage more fully with the content. For your project, consider how you can integrate examples, case studies, or stories from the students' own cultures. For instance, if teaching a history lesson, include historical figures and events from the cultures represented in your classroom.
2. **Family and Community as a Resource:** In my upbringing, family played a crucial role. Involving families in the learning process can greatly enrich classroom learning. You might design projects or assignments that require students to interview family members or involve them in building a family tree, which teaches social studies and personal history.
3. **Language and Communication:** Recognize the linguistic assets students bring into the classroom. Using students' first languages when possible can aid understanding and show respect for their linguistic background. For example, providing bilingual resources or allowing students to present work in their first language can make learning more accessible and engaging.
4. **Skills from Home:** Many students develop specific skills through their home life—be it cooking, mechanical repairs, or other crafts. Projects that allow students to demonstrate and apply these skills can validate their non-academic knowledge and show its relevance to school subjects. For example, if a student's family is involved in construction, a math project could involve calculating materials needed for a model house.

Source: Authors' own work.

Figure 9.8. JeepyTA suggesting discussion questions to start off a weekly discussion on coding qualitative data in the “Quantitative Ethnography and Epistemic Network Analysis” course



JeepyTA 13 Feb Edited

Hello dear class! I have some engaging themes, topics, and guiding questions based on the readings for Week 6:

1. Developing a Coding System:

- What are the challenges and benefits of developing your own coding system versus using a predefined one (e.g., what David mentioned in the QE book: inductive/bottom-up and deductive/top-down) in QE research?
- How do the readings suggest balancing automation (like using AI tools) with human expertise when coding qualitative data?

2. The Role of Codes in QE:

- What is the relationship between codes and the research questions in QE studies? How do the readings illustrate this relationship? How do you see this relationship in your own work?

3. Ethical Considerations in Coding:

- What ethical considerations should researchers keep in mind while developing a coding scheme?

4. Code Consistency and Integrity:

- According to Shaffer & Ruis (2020), how do researchers ensure "fairness" in the coding process across different datasets or studies? Give an example of what it might look like in your work.

These themes and questions aim to create a robust dialogue around the complexities and innovations in coding for quantitative ethnography. I hope they inspire deep and meaningful discussions in your class!

Best,
Jeepy.

Source: Authors' own work.

Evaluation of JeepyTA

JeepyTA was first implemented in Fall 2023 and covered use cases described above. To understand how students viewed the virtual teaching assistant, we distributed a voluntary end-of-semester survey, approved as exempt by the university's Institutional Review Board, and clearly stated that participation was optional and would not affect grades. Students provided informed consent before answering 13 multiple-choice questions. The survey asked about specific aspects of JeepyTA's performance, including how quickly and accurately it responded to questions, how clearly and professionally it communicated, and how well it supported student learning, development, and motivation, compared to a human TA. Response options ranged from 1 ("Human TA is significantly better") to 5 ("AI TA is significantly better"). We used two-sample t-tests to check whether the average scores for each question differed from the neutral midpoint of 3. This allowed us to measure whether students viewed JeepyTA as better or worse than a human TA. Results showed that students rated JeepyTA as comparable to a human TA in several areas, including the speed and clarity of its replies, the accuracy and professionalism of its responses, its ability to support learning without giving away answers, and the overall usefulness and quality of its feedback. However, students rated it lower in three areas: offering useful ideas, supporting student development, and motivating students.

We also evaluated JeepyTA's impact on when responses were available to students. In two identical courses offered one year apart, taught by the same professor at the same institution and involving comparable student groups, a difference emerged in response times to students' queries. During the earlier term, when JeepyTA was not in use, the instructor and the TAs posted 153 responses, with a median response time of 7.09 hours. In the following term of Fall 2023, after JeepyTA was introduced, course staff posted 136 responses, and the median response time dropped to 2.23 hours, a statistically significantly lower time.

Among 89 posts where JeepyTA attempted to answer student inquiries, 22 responses received approval. The AI generated replies in approximately 40 seconds, and course staff approved them within an average of 38 minutes. Because JeepyTA was able to handle these queries, even manually written responses were faster. In the JeepyTA semester, the median human response time was 4.14 hours, statistically significantly faster than the 7.09-hour median time observed in the prior semester. This suggests that JeepyTA improved the efficiency of even fully-human responses, likely because the instructor and TAs had more time to address tasks that deserved more of their attention.

Another of JeepyTA's intended purpose was to support instructors and TAs outside standard working hours. In the term prior to its introduction, course staff posted 62% of replies outside regular U.S. business hours (after 5 p.m. and before 9 a.m.). In the term when JeepyTA was available, this proportion was 60%, not statistically significantly different. However, there appeared to be a difference in the proportion of responses posted during weekends. In the prior semester, instructors posted 10% of all replies on weekends. After JeepyTA was introduced, this number increased to 29% of the total number of replies (including JeepyTA posts approved or edited by the instructors). This increase was statistically significant, suggesting that course staff were better able to focus their work time even on weekends, following the introduction of JeepyTA.

Work is currently ongoing to study the impact of JeepyTA's feedback on student essays (3b). In our initial work, we have found that semester-on-semester, students receive statistically significantly higher assignment grades (according to an independent grader) after receiving JeepyTA's feedback – going from an average of 64% of students receiving an A or an A+ on their final submission to 95% of students receiving an A or A+. In follow-up work, we are investigating whether students specifically fix the issues identified by JeepyTA in that same essay, and whether they make the same mistakes in subsequent essays (including in a different class where JeepyTA is also offered).

A study on JeepyTA's brainstorming support functionality (3f) (Shah et al., 2024_[46]) revealed that JeepyTA helped students generate a higher volume of ideas (averaging 2.78 per student compared to 1.7 for student-led ideation) and increased the production of fully formed, detailed concepts. However, this came with important trade-offs: JeepyTA-driven ideas often showed thematic overlap, with common suggestions like "teamwork and collaboration" appearing across multiple students' work, while student-driven ideas exhibited greater diversity (cf. Doshi and Hauser, 2024_[52]). In analysing students' descriptions of their process of using JeepyTA, the researchers identified five distinct patterns: 36% of students found the suggestions insightful and aligned with their own ideas, 18% acknowledged JeepyTA's role in idea generation but did not say if it was actually helpful in doing so, 18% referenced using JeepyTA but did not say how/if its use was connected to their proposed ideas, 18% made no reference to JeepyTA, and 13% critically evaluated and built upon JeepyTA's recommendations. Notably, 68% of students proposed multiple educational applications spanning knowledge types essential for 21st-century learning (Kereluik et al., 2013_[53]): meta-knowledge (collaboration, problem-solving), foundational knowledge (mathematics, history, computer science), and humanistic knowledge (digital citizenship, ethical awareness). Especially novel recommendations emerged when students integrated insights from readings, game experiences, and their domain expertise, suggesting that JeepyTA works best as a brainstorming tool when students already possess foundational knowledge they can build upon (Nasiar, forthcoming_[54]).

Other case studies of AI-augmented TAs

Example from India

At the *Indian Institute of Technology Kanpur*, for instance, researchers have piloted AI augmented TAs for an introductory computer science course, where student-to-instructor ratios are often too high for human instructors or TAs to provide individual guidance at scale (Ahmed, 2025_[55]). In response to this situation, the team integrated an AI agent into Prutor, a web-based programming platform used by students to submit solutions for C programming assignments. When a student's programme failed to pass the instructor-defined test cases, they could request assistance by clicking a "Get Help" button within the platform. This action triggered a feedback request that compiled four key inputs and sent them to GPT-4 Turbo via an internal API: the problem description, the student's buggy code, the test case results, and an optional message written by the student to describe their confusion or ask a specific question. Using this information, GPT-4 Turbo generated targeted feedback linked to specific lines in the student's code that highlights the exact locations of potential errors and explaining what may have gone wrong. The output was routed to a centralised dashboard, where human TAs could review the AI's draft response, make edits, add notes, or reject it entirely before sending the final feedback back to the student through the same interface. The AI agent was evaluated against two other conditions: one in which human TAs provided all feedback without any AI support, and another in which feedback generated by GPT-4 Turbo was sent directly to students without human review. Researchers examined how these types of feedback methods affected feedback quality (measured through expert evaluations), TA efficiency

(measured through response times recorded in system logs), and student performance (measured by whether final code submissions passed all instructor-defined test cases). In addition, students rated each piece of feedback based on helpfulness, clarity, and timeliness using built-in rating tools on the platform. Although AI-generated feedback was often rated favourably by students – particularly for its detailed explanations – these positive perceptions did not consistently lead to improved performance. Students receiving AI-assisted or fully automated feedback were not significantly more likely to complete the assignments successfully, and in many cases, manual TA feedback led to faster and more effective problem resolution.

One issue observed by the researchers was that some TAs using AI-generated feedback forwarded it without making necessary corrections, even when the output contained inaccuracies or hallucinations. This appeared to reflect a tendency among certain TAs to rely too heavily on the AI drafts instead of critically evaluating their quality. In contrast, TAs working without AI support often provided responses which highlighted the immediate next step or pinpointed the specific source of the error. As a result, students in the manual TA group were, in some cases, able to resolve issues more efficiently, despite often receiving shorter and less detailed feedback.

Example from Czechia

Another example comes from Czechia, where a GPT-3-based conversational chatbot named Alex was deployed in a university-level English course (Polakova and Klimova, 2024^[56]). Alex is a web-based application that combines multiple AI models: GPT-3 is used to generate natural language responses, while Gramformer and T5 are applied to detect and correct grammatical errors. At the beginning of each chat session, GPT-3 generates a topic-specific opening question based on preselected weekly themes. Students then respond freely in English. Their input is processed by Gramformer and T5 to identify grammatical mistakes. When an error is detected, the chatbot enters a correction phase, during which GPT-3 provides a corrected version of the sentence along with an explanation in natural language. The chatbot also allows users to rate the feedback as either "good" or "bad". Over a four-week period, students engaged with Alex in simulated dialogue sessions and received real-time feedback. Though the topic of each session starts with a fixed opening, users can take the conversation in any direction. To constrain the types of interactions, users were limited to one session per day and three per week, with each session requiring a minimum of 1 000 characters of typed input. After completing the four-week programme, students filled out a questionnaire survey about their experience. Analysis of student feedback revealed that learners responded positively to several aspects of the chatbot. Students noted that the chatbot asked clear, easy-to-understand questions and responded quickly, which helped keep the conversation at a natural pace. Many also appreciated the flexibility to practice outside the classroom. The option to access conversation practice on their own helped them gain confidence. According to the survey, 88% of students reported they did not feel stressed using the chatbot, and several commented that the experience felt like chatting with a real person. In terms of learning gains, pre- and post-tests focusing on grammar and vocabulary showed measurable improvements. Upper-intermediate students improved their tests scores from about 59% to 75%, while advanced students increased from 80% to 90%.

At the same time, students also reported several limitations of the chatbot that affected their overall experience. Some participants encountered technical problems, such as system lags and incomplete replies from the chatbot, which disrupted the flow of conversation. Others pointed out that certain responses felt repetitive or too limited in variation, which reduced the usefulness of later sessions. Survey results also showed that the chatbot failed to increase motivation or encourage continued use: 74% of students did not feel more motivated to learn English as a result of using the chatbot, and 79% said they would rather use other tools like Duolingo or talk to native speakers instead.

Though different in design and pedagogical goals, these examples share JeepyTA's aim of delivering coursealigned, scalable support. Even though these AI-powered teaching assistants differ from JeepyTA in how they function, the courses they support, the tasks they were assigned, and in the regional infrastructure of the learning environment, similar benefits and challenges can be noted.

Reflection on generative AI as teaching assistants: implications and policy recommendations

The emergence of generative AI (GenAI) in education signals a shift in the professional landscape of education, decentralising some traditional instructional roles and prompting a rethinking of what it means to teach. Historically, expertise in education has been concentrated in instructors and (to a lesser extent) human TAs, who provide guidance, feedback, and assessment. However, as GenAI systems become increasingly capable of tutoring, scaffolding learning, and responding to student needs in real time, the boundaries of these roles are being redrawn.

Rather than replacing human expertise, AI TAs built using GenAI demand that we rethink how instructional work is distributed—not just between instructors and human TAs, but across AI-enhanced systems. This shift mirrors what Shaffer, Nash, and Ruis (2015^[57]) describe as the reconfiguration of professional expertise in response to new tools and technologies. As they argue, professionalisation is not static; it evolves when new ways of knowing and working emerge. In the case of education, AI's growing presence means that instructors must develop new competencies – not only in teaching content, but in orchestrating AI-enhanced learning environments, interpreting AI-generated insights, and ensuring alignment between AI feedback and pedagogical goals. As such, policy should emphasise training of instructors and human TAs to work effectively with AI.

It is also critical to ensure that GenAI integration does not drive pedagogy but rather supports it in meaningful ways. The Technological Pedagogical Content Knowledge (TPACK) framework (Koehler, 2009^[58]) provides a useful lens for understanding this challenge. Educators must consider how GenAI interacts with both content knowledge (what is being taught) and pedagogical strategies (how it is taught). Without thoughtful integration, there is a risk that AI could push education further toward efficiency-based models, where rapid feedback and automated assessments replace deeper engagement with complex ideas rather than finding an optimal combination of deeper learning and more automated activities where appropriate (Mishra, Warr and Islam, 2023^[59]). Prioritising efficiency and engagement over meaningful understanding, may increase the quality of student work and student experience in the short-term but may not benefit the student in the longer term.

Mishra and colleagues' work also highlights the need to move beyond mere adoption of AI tools to meaningful integration into learning experiences. The presence of AI TAs does not inherently improve education; their effectiveness depends on how they are aligned with broader learning goals. Educators must take an active role in shaping AI's function within courses, ensuring it complements and enhances human-centred teaching practices rather than supplanting them. Therefore, we recommend against designing of AI Teaching Assistants in ways that replace humans and fully automate all learning activities, and the adoption of cost-cutting measures that create pressure to eliminate human TAs should also be avoided. Beyond reducing the quality of instruction, reduced funding for Teaching Assistants would also decrease opportunities for economically disadvantaged individuals who rely on Teaching Assistant positions as a pathway into academia, ultimately reducing the pipeline of talented scholars into research and scholarship.

Henriksen and Mishra's (2024^[60]) work on practical wisdom further reinforces this perspective, emphasising that experienced educators bring a form of professional knowledge that AI cannot replicate - one rooted in ethical decision-making, contextual understanding, and reflective practice. As AI transforms the nature of knowledge in education, teachers must ensure that human judgment, adaptability, and social-emotional insights remain at the core of teaching. This highlights the need for educators to approach GenAI critically, leveraging its strengths while maintaining the core humanistic elements of teaching and mentorship.

Reprofessionalisation in this context is therefore not just about preparing educators for an AI-integrated classroom—it is about ensuring that humans and AI systems work in sync to advance student learning experiences and outcomes. Just as previous technological shifts reshaped the teaching profession, GenAI requires a reimagining of teacher preparation, assessment design, and professional collaboration. The goal is not merely to integrate AI, but to define new models of expertise in an AI-augmented educational ecosystem—one in which human and AI agents work together to support meaningful learning experiences while maintaining a commitment to equity and effectiveness.

Implications for student assessment

Assessment practices in higher education serve multiple purposes: they inform students about their progress, provide instructors with actionable insights to guide instruction, and certify learners' competence. The emergence of AI Teaching Assistants (AI TAs) such as JeepyTA adds new opportunities for formative assessment, while creating possibilities for summative assessment which merit some caution. In both cases, careful design and policy guardrails are necessary to ensure that AI TAs enhance rather than undermine the educational process.

A chief benefit of AI TAs is in supporting formative feedback, where fast, specific, and individualised guidance can promote deeper learning. This feedback can be provided both immediately and 24/7. Students who work late at night, study remotely, or juggle other responsibilities often cannot attend regular office hours or wait for TAs to become available. An AI TA can fill this gap by providing immediate, round-the-clock feedback, easing the pressure on human TAs and making support more equitable.

Historically, automated assessments (e.g. quizzes, short-answer grading) have helped identify student misconceptions and encouraged targeted practice. Generative AI now expands these possibilities by supporting more complex tasks—from essay drafts to coding projects. In general, an AI TA can assess a broader range of competencies than a typical

human TA, supporting shifts in assessment from assessing what students know to their conceptual understanding and their process of solving problems. Importantly, using AI for assessment also opens the possibility of assessing student learning through a broader range of artifacts – including annotations, reflections, peer feedback, conversations, and other forms of evidence – enabling a more holistic and nuanced view of learning that extends beyond traditional measures.

They can also provide sophisticated, multi-dimensional assessment. As seen in the JeepyTA platform discussed earlier, AI TAs can offer feedback on essay organisation, argument clarity, and conceptual rigor, referencing course rubrics and standards to align with the instructors' goals. This kind of timely, actionable commentary can help learners iterate more quickly, moving from basic correctness checks toward higher-order thinking and reflection (Hattie and Timperley, 2007^[6]; Chi and Wylie, 2014^[7]). If these systems are designed carefully to align with the course-specific materials and rubric (as JeepyTA has been primed to do), it is possible to reduce the risk of misleading feedback that the instructor will disagree with—although it still happens, just as a human TA could also provide feedback that an instructor disagrees with. Furthermore, the use of personas (as discussed above) or carefully designed prompting can offer students feedback from different perspectives, highly relevant in some disciplinary areas but difficult for a single human TA or instructor to provide.

With AI TAs taking on time-consuming tasks such as answering routine questions, reviewing initial drafts, or providing rubric-aligned suggestions, human TAs and instructors are increasingly able to reallocate their time toward more pedagogically meaningful and relationship-centred activities. These include leading in-depth discussions that challenge students to think critically, working directly with individuals or small groups to support their academic progress, meeting individually with students to support their academic and professional growth, and developing activities that promote academic agency and collaboration. Academic staff can now invest more energy in synthesising performance patterns across student submissions, identifying emerging misconceptions, and making ongoing improvements to assignments or assessments based on observed trends and student needs. These higher-order instructional practices remain difficult for AI systems to replicate, though they can support these tasks in various ways. Rather than displacing humans, then, AI TAs can create space for them to focus on tasks that require interpreting student thinking in context, applying subject-matter expertise, and exercising instructional judgment.

Another possible concern with the provision of formative feedback by AI TAs is over-scaffolding, where learners rely so heavily on AI-generated suggestions that their final products no longer represent independent work. Particularly if students can seek several rounds of feedback, or obtain lower-level writing suggestions, the sophistication of today's generative AI can obscure the boundary between a student's own efforts and AI-provided content. To avoid issues of this nature, careful consideration of design is needed, which can be supported by policy funding the establishment of guidelines for how much scaffolding is appropriate in different situations and research on methods for producing the benefits of scaffolding while avoiding over-scaffolding. Nonetheless, on the whole, the benefits of providing formative feedback through an AI TA seem – if carefully designed – do outweigh the risks, and policy should encourage higher education to move forward in using AI TAs for formative feedback in ways that can improve student learning.

Greater concerns are present for more summative assessment, such as assigning final marks on assignments. Automated essay scoring has a history dating back decades (Shermis and Burstein, 2003^[18]), offering consistency and quick turnaround but often criticised for focusing on superficial textual features. Still, the use of automated essay scoring offers many lessons in how automated grading can be psychometrically validated and used appropriately in ways that support higher education institutions in offering high-quality consistent grading at lower human cost. More advanced, generative AI-based approaches can analyse content in greater detail than most earlier methods, but can be prone to systematic biases, a lack of transparency in scoring, treating inaccurate but widely believed misconceptions as true, and even in some cases have “hallucinations” of information not in the original student work. For high-stakes decisions such as course grades, even small error rates can have consequences for student outcomes and perceptions of fairness. Moreover, incorporating AI into summative grading can amplify existing equity concerns. If an AI TA's underlying model has been trained on data reflecting cultural or linguistic biases, students from underrepresented backgrounds may be inadvertently penalised. The presence of generative AI in summative grading therefore necessitates robust validation across diverse student populations and consideration of whose perspectives are centred within assessment (Lee, 1998^[61]) and even in the evaluation of the fairness of assessments (Randall, 2023^[62]). Policymakers and institutional leaders must establish policies that ensure that any summative use of AI-based scoring be supported by transparent procedures, documented reliability metrics (including evidence that there is limited or no algorithmic bias (Baker, Hawn and Lee, 2023^[63]), and the ability for students (and instructors) to question and appeal automated scores.

Decisions about when AI assistance is acceptable – and how much AI-driven contribution is too much – will differ by course context and disciplinary standards, as well as how far along students are in their development of expertise. In fields like computer science or business, where collaborative problem-solving with tools is central, it may be appropriate to evaluate students in more advanced classes based on how effectively they leverage AI to reach correct solutions. By contrast, in courses emphasising individual mastery of foundational skills, unmediated AI assistance could undermine the competencies being assessed. As such, it may be appropriate to develop guidelines, either at a disciplinary level or in the context of specific standard courses, for which tasks should incorporate AI support, and what types of formative assessment and support are warranted.

Overall, there is considerable potential for the use of generative AI-based Teaching Assistants to support formative assessment, and some possibilities in summative assessment as well, if approached with sufficient caution and human oversight is retained. The design of policy to encourage appropriate use has potential benefits for both students and instructors.

Implications for equity and policies to support equity

The simple existence of AI TAs can play an important role in supporting equity. Many students struggle to get academic help when they need it, not because they do not have the motivation but because structural barriers make access difficult. Some students have jobs or caregiving responsibilities that prevent them from attending office hours. Others study in different time zones and cannot reach instructors or TAs outside of regular class hours. AI TAs can straightforwardly remove some of these barriers by providing immediate responses to course-related questions at any time of day, which makes academic support more accessible to students who might otherwise face difficulties getting assistance when they need it. If designed correctly (and if such design effort is made part of the criteria in requests for proposals and other instruments encouraging the development and use of AI TAs), AI TAs can also be more effective at avoiding unconscious bias, where historically underrepresented students may receive different and lower-quality feedback than other students. When AI TAs are built on large language models that are multilingual, it also becomes possible to provide content in multiple languages, increasing the accessibility of feedback and support for some international students.

Similarly, AI TAs can relieve some equity issues impacting human TAs. Large courses create significant demands on Teaching Assistants, who must divide their time between grading, responding to student questions, and assisting with course administration. Many Teaching Assistants take on these responsibilities while managing their own coursework, research, and professional development. The time constraints they face often affect not only their own studies but also the level of detail they can provide in feedback or the number of students they can support individually. AI TAs can reduce some of these pressures by handling routine inquiries and generating structured feedback on assignments. Therefore, policy to adopt AI TAs where appropriate can have fairly rapid and sizable benefits for equity, for both students and Teaching Assistants.

However, the adoption of AI TAs can also create equity concerns. Holstein and Doroudi's (2021^[64]) research suggests that educational technologies often reinforce existing inequities, benefiting students who already possess strong academic skills while leaving others behind. That said, recent studies on large language models indicate a more complex dynamic: LLMs may provide greater relative benefits to less knowledgeable users, a pattern sometimes referred to as the GPS effect (Chiang et al., 2024^[34]). Just as GPS systems support those unfamiliar with a route more than experienced drivers, LLMs can scaffold novices more effectively than they assist experts. This suggests that, once access is secured, AI TAs could disproportionately benefit those with less prior knowledge – potentially narrowing achievement gaps rather than widening them. Of course, access itself remains a critical barrier, particularly in terms of digital literacy, language fluency, and reliable connectivity. Moreover, the risk remains that AI systems might encode cultural, linguistic, or epistemic biases that privilege dominant norms. The effort to incorporate specific cultural perspectives and funds of knowledge through personas, discussed above, is our first attempt to address this concern within JeepyTA. It is also important to ensure that AI TAs do not primarily benefit students with higher digital literacy and stronger academic foundations. Ultimately, then, it is important that policy requires evidence that AI TAs are fair for all learners – not just supporting development but validation of fairness across learners as part of evaluation.

Conclusion

This chapter has examined how AI Teaching Assistants, particularly those driven by generative AI, can expand and support the traditional roles played by human TAs and instructors in higher education. We began by considering the foundational role that human TAs play, along with the logistical and pedagogical challenges they often encounter, such as balancing workload with their own academic commitments. Against this backdrop, we reviewed the emergence of AI TAs—tools that are designed to automate routine administrative tasks, deliver real-time student support, and

provide timely formative feedback at scale in ways that complement and extend human capabilities. Rather than viewing AI TAs as replacements for human TAs, we argue for a model of augmentation—one where AI tools provide support and feedback that is otherwise infeasible for human instructors and TAs alone: support and feedback that is immediate, personalised, detailed, and available around the clock.

To ground these ideas, we discussed JeepyTA – a generative AI system developed by the Penn Center for Learning Analytics at University of Pennsylvania. Deployed across multiple graduate-level courses and multiple institutions and integrated into course discussion forums, JeepyTA supports a variety of instructional needs. These include answering logistical queries, providing formative feedback on essays, assisting students with debugging their code, stimulating creative ideation, summarising discussion posts, and suggesting new discussion prompts for deeper engagement. By embedding course materials and rubrics into an LLM-based system, JeepyTA demonstrates how AI TAs can deliver context-aware and curriculum-aligned responses at scale. Throughout the chapter, we also discussed key design and implementation considerations—such as prompt design, the need for human oversight and review, consideration of ethics and bias, and alignment with policy. These factors are essential for ensuring that AI TAs operate responsibly, transparently, and in service of equitable learning outcomes.

Ultimately, the experiences documented here suggest that well-designed AI TAs can ease pressure on human TAs and academic staff, enhance student engagement, and potentially enhance the quality of learning experiences. Importantly, their effectiveness depends not only on the technical sophistication of generative AI, but on careful integration into pedagogical practices. When thoughtfully deployed, AI Teaching Assistants can help institutions create more scalable, responsive, and personalised educational ecosystems—supporting students, TAs, and instructors in new and meaningful ways.

References

- Ahmad, S., M. Alam, M. Rahmat, M. Mubarak and S. Hyder (2022), "Academic and Administrative Role of Artificial Intelligence in Education", *Sustainability*, Vol. 14/3, p. 1101, <https://doi.org/10.3390/su14031101>. [24]
- Ahmed, U. (2025), *Feasibility Study of Augmenting Teaching Assistants with AI for CS1 Programming Feedback*. [55]
- Aleven, V., B. McLaren, J. Sewall and K. Koedinger (2006), "The Cognitive Tutor Authoring Tools (CTAT): Preliminary Evaluation of Efficiency Gains", in *Lecture Notes in Computer Science, Intelligent Tutoring Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/11774303_7. [22]
- Alsafari, B., E. Atwell, A. Walker and M. Callaghan (2024), "Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants", *Natural Language Processing Journal*, Vol. 8, p. 100101, <https://doi.org/10.1016/j.nlp.2024.100101>. [42]
- Baker, R. (2016), "Stupid Tutoring Systems, Intelligent Humans", *International Journal of Artificial Intelligence in Education*, Vol. 26/2, pp. 600-614, <https://doi.org/10.1007/s40593-016-0105-0>. [23]
- Baker, R., M. Hawn and S. Lee (2023), *Algorithmic Bias: The State of the Situation and Policy Recommendations*, OECD Publishing, Paris, <https://doi.org/10.1787/c74f03de-en>. [63]
- Begley, G., B. Berkey, L. Roe and H. Schuldt (2019), "Becoming partners: Faculty come to appreciate undergraduates as teaching partners in a service-learning teaching assistant program", *International Journal for Students as Partners*, Vol. 3/1, pp. 89-105, <https://doi.org/10.15173/ijps.v3i1.3669>. [10]
- Berta, M. and J. Pembridge (2019), "Exploring Burnout among Graduate Teaching Assistants", *2019 ASEE Annual Conference & Exposition Proceedings*, <https://doi.org/10.18260/1-2--32802>. [13]
- Brooks, H. (2023), *The History of Test Automation*, <https://testrigor.com/blog/the-history-of-test-automation/> (accessed on 2/24/2025). [19]
- Bulut, O. and M. Beiting-Parrish (2024), "The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges", *Chinese/English Journal of Educational Measurement and Evaluation*, Vol. 5/3, <https://doi.org/10.59863/miqj7785>. [29]

- Chiang, C.-H., W.-C. Chen, C.-Y. Kuan, C. Yang and H.-y. Lee (2024), "Large Language Model as an Assignment Evaluator: Insights, Feedback, and Challenges in a 1000+ Student Course", *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 2489-2513, <https://doi.org/10.18653/v1/2024.emnlp-main.146>. [34]
- Chi, M. and R. Wylie (2014), "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes", *Educational Psychologist*, Vol. 49/4, pp. 219-243, <https://doi.org/10.1080/00461520.2014.965823>. [7]
- Dede, C. and J. Richards (eds.) (2018), *Jill Watson: A Virtual Teaching Assistant for Online Education*, Routledge. [21]
- Doshi, A. and O. Hauser (2024), "Generative AI enhances individual creativity but reduces the collective diversity of novel content", *Science Advances*, Vol. 10/28, <https://doi.org/10.1126/sciadv.adn5290>. [52]
- Essel, H., D. Vlachopoulos, A. Tachie-Menson, E. Johnson and P. Baah (2022), "The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education", *International Journal of Educational Technology in Higher Education*, Vol. 19/1, <https://doi.org/10.1186/s41239-022-00362-6>. [27]
- Foster, A. and M. Shah (2020), "Principles for Advancing Game-Based Learning in Teacher Education", *Journal of Digital Learning in Teacher Education*, Vol. 36/2, pp. 84-95, <https://doi.org/10.1080/21532974.2019.1695553>. [47]
- Gao, R., H. Merzdorf, S. Anwar, M. Hipwell and A. Srinivasa (2024), "Automatic assessment of text-based responses in post-secondary education: A systematic review", *Computers and Education: Artificial Intelligence*, Vol. 6, <https://doi.org/10.1016/j.caeai.2024.100206>. [30]
- Gašević, D. and L. Yan (2026), "Generative AI for human skill development and assessment: implications for existing practices and new horizons", in *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*, OECD Publishing, Paris, <https://doi.org/10.1787/062a7394-en>. [1]
- Gonzalez, N., L. Moll and C. Amanti (eds.) (2006), *Funds of Knowledge*, Routledge, <https://doi.org/10.4324/9781410613462>. [48]
- Graham, S. (2023), "Post-Process but Not Post-Writing: Large Language Models and a Future for Composition Pedagogy", *Composition Studies*, Vol. 51(1), pp. 162-168. [38]
- Haaben, F. (2024 (Unpublished Manuscript), *Teaching With AI: Augmenting Professors With LLMs In An Asynchronous Question Center*, <https://doi.org/10.14288/1.0443555>. [43]
- Hattie, J. and H. Timperley (2007), "The Power of Feedback", *Review of Educational Research*, Vol. 77/1, pp. 81-112, <https://doi.org/10.3102/003465430298487>. [6]
- Henriksen, D. (2024), "Teaching, Teacher Education, and Practical Wisdom in the Age of Generative AI", *Society for Information Technology Teacher Education International Conference*, pp. 779-787, <https://www.learntechlib.org/primary/p/224040/>. [60]
- Holstein, K. (2021), *Equity and Artificial Intelligence in Education: Will "AIEd" Amplify or Alleviate Inequities in Education?*, <https://doi.org/10.48550/arXiv.2104.12920>. [64]
- IIITD, A., D. Sethi, N. Gupta, S. Sharma, S. Jain, U. Singhal and D. Kumar (2024), *TAMIGO: Empowering Teaching Assistants using LLM-Assisted Viva and Code Assessment in an Advanced Computing Class*. [33]
- Karim, N., A. Maries and C. Singh (2018), "Exploring one aspect of pedagogical content knowledge of teaching assistants using the Conceptual Survey of Electricity and Magnetism", *Physical Review Physics Education Research*, Vol. 14/1, <https://doi.org/10.1103/physrevphyseducres.14.010117>. [9]
- Karpouzis, K., D. Pantazatos, J. Taouki and K. Meli (2024), "Tailoring Education with GenAI: A New Horizon in Lesson Planning", *2024 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1-10, <https://doi.org/10.1109/educon60312.2024.10578690>. [28]
- Kereluik, K., P. Mishra, C. Fahnoe and L. Terry (2013), "What Knowledge Is of Most Worth", *Journal of Digital Learning in Teacher Education*, Vol. 29/4, pp. 127-140, <https://doi.org/10.1080/21532974.2013.10784716>. [53]
- Khine, M., B. Laszlo and E. and Afari (eds.) (forthcoming), *A Step Towards Adaptive Online Learning: Exploring the Role of GPT as Virtual Teaching Assistants in Online Education*, CRC Press. [16]
- Khosravi, H., S. Shabaninejad, A. Bakharia, S. Sadiq, M. Indulska and D. Gašević (2021), "Intelligent Learning Analytics Dashboards: Automated Drill-Down Recommendations to Support Teacher Data Exploration", *Journal of Learning Analytics*, Vol. 8/3, pp. 133-154, <https://doi.org/10.18608/jla.2021.7279>. [17]
- Koehler, M. (2009), "What is technological pedagogical content knowledge (TPACK)?", *Contemporary Issues in Technology and Teacher Education*, Vol. 9 (1), pp. 60-70. [58]

- Kurdi, G., J. Leo, B. Parsia, U. Sattler and S. Al-Emari (2019), "A Systematic Review of Automatic Question Generation for Educational Purposes", *International Journal of Artificial Intelligence in Education*, Vol. 30/1, pp. 121-204, <https://doi.org/10.1007/s40593-019-00186-y>. [31]
- Ladson-Billings, G. (1995), "Toward a Theory of Culturally Relevant Pedagogy", *American Educational Research Journal*, Vol. 32/3, pp. 465-491, <https://doi.org/10.3102/00028312032003465>. [49]
- Lee, C. (1998), "Culturally Responsive Pedagogy and Performance-Based Assessment", *The Journal of Negro Education*, Vol. 67/3, p. 268, <https://doi.org/10.2307/2668195>. [61]
- Lee, C., J. Myung, J. Han, J. Jin and A. Oh (2023), *Learning from Teaching Assistants to Program with Subgoals: Exploring the Potential for AI Teaching Assistants*. [44]
- Lim, L., S. Gentili, A. Pardo, V. Kovanović, A. Whitelock-Wainwright, D. Gašević and S. Dawson (2021), "What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course", *Learning and Instruction*, Vol. 72, <https://doi.org/10.1016/j.learninstruc.2019.04.003>. [37]
- Liu, M. and F. M'Hiri (2024), "Beyond Traditional Teaching: Large Language Models as Simulated Teaching Assistants in Computer Science", *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, pp. 743-749, <https://doi.org/10.1145/3626252.3630789>. [39]
- Li, Y. and X. Hu (2026), "Learning with dialogue-based AI tutors: implementing the Socratic method with Generative AI", in *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*, OECD Publishing, Paris, <https://doi.org/10.1787/062a7394-en>. [2]
- Maiti, P. and A. Goel (2024), *How Do Students Interact with an LLM-powered Virtual Teaching Assistant in Different Educational Settings?*. [40]
- Marincovich, M. and J. Prostko (eds.) (1998), *Preparing Graduate Students to Teach: Past, Present, and Future*, Anker Publishing Company. [5]
- Marshman, E., R. Sayer, C. Henderson, E. Yerushalmi and C. Singh (2018), "The challenges of changing teaching assistants' grading practices: Requiring students to show evidence of understanding", *Canadian Journal of Physics*, Vol. 96/4, pp. 420-437, <https://doi.org/10.1139/cjp-2017-0030>. [8]
- Miroyan, M., C. Mitra, R. Jain, G. Ranade and N. Norouzi (2025), "Analyzing Pedagogical Quality and Efficiency of LLM Responses with TA Feedback to Live Student Questions", *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, pp. 770-776, <https://doi.org/10.1145/3641554.3701965>. [41]
- Mishra, P., M. Warr and R. Islam (2023), "TPACK in the age of ChatGPT and Generative AI", *Journal of Digital Learning in Teacher Education*, Vol. 39/4, pp. 235-251, <https://doi.org/10.1080/21532974.2023.2247480>. [59]
- Nasiar, N. (forthcoming), "Scaffolding Creativity or Homogenizing Outputs? GPT-based Virtual TAs Impact on Student Ideation in Graduate Education. Manuscript submitted to Special Edition on AI and Education.", *Social Innovations Journal*. [54]
- Nicklow, J., S. Marikunte and L. Chevalier (2007), "Balancing Pedagogical and Professional Practice Skills in the Training of Graduate Teaching Assistants", *Journal of Professional Issues in Engineering Education and Practice*, Vol. 133/2, pp. 89-93, [https://doi.org/10.1061/\(asce\)1052-3928\(2007\)133:2\(89\)](https://doi.org/10.1061/(asce)1052-3928(2007)133:2(89)). [12]
- Nye, B., A. Graesser and X. Hu (2014), "AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring", *International Journal of Artificial Intelligence in Education*, Vol. 24/4, pp. 427-469, <https://doi.org/10.1007/s40593-014-0029-5>. [20]
- Omi, M. and H. Winant (2018), "Racial Formation in the United States", in *Inequality in the 21st Century*, Routledge, <https://doi.org/10.4324/9780429499821-49>. [51]
- Pardo, A., K. Bartimote, S. Buckingham Shum, S. Dawson, J. Gao, D. Gašević, S. Leichtweis, D. Liu, R. Martinez-Maldonado, N. Mirriahi, A. Moskal, J. Schulte, G. Siemens and L. Vigentini (2018), "OnTask: Delivering Data-Informed, Personalized Learning Support Actions", *Journal of Learning Analytics*, Vol. 5/3, <https://doi.org/10.18608/jla.2018.53.15>. [25]
- Paris, D. (2012), "Culturally Sustaining Pedagogy", *Educational Researcher*, Vol. 41/3, pp. 93-97, <https://doi.org/10.3102/0013189x12441244>. [50]
- Park, C. (2004), "The graduate teaching assistant (GTA): Lessons from North American experience", *Teaching in Higher Education*, Vol. 9/3, pp. 349-361, <https://doi.org/10.1080/1356251042000216660>. [3]
- Polakova, P. and B. Klimova (2024), "Implementation of AI-driven technology into education – a pilot study on the use of chatbots in foreign language learning", *Cogent Education*, Vol. 11/1, <https://doi.org/10.1080/2331186x.2024.2355385>. [56]

- Randall, J. (2023), "It Ain't Near 'Bout Fair: Re-Envisioning the Bias and Sensitivity Review Process from a Justice-Oriented Antiracist Perspective", *Educational Assessment*, Vol. 28/2, pp. 68-82, <https://doi.org/10.1080/10627197.2023.2223924>. [62]
- Sajja, R., Y. Sermet, M. Cikmaz, D. Cwiertny and I. Demir (2024), "Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education", *Information*, Vol. 15/10, p. 596, <https://doi.org/10.3390/info15100596>. [14]
- Shaffer, D., P. Nash and A. Ruis (2015), "Technology and the New Professionalization of Teaching", *Teachers College Record: The Voice of Scholarship in Education*, Vol. 117/12, pp. 1-30, <https://doi.org/10.1177/016146811511701205>. [57]
- Shah, M., M. Pankiewicz, R. Baker, J. Chi, Y. Xin, H. Shah and D. Fonseca (2024), "Students' Use of an LLM-Powered Virtual Teaching Assistant for Recommending Educational Applications of Games", in *Lecture Notes in Computer Science, Serious Games*, Springer Nature Switzerland, Cham, https://doi.org/10.1007/978-3-031-74138-8_2. [46]
- Shermis, M. and J. Burstein (eds.) (2003), *Automated Essay Scoring*, Routledge, <https://doi.org/10.4324/9781410606860>. [18]
- Sinha, A. (2024), *BoilerTAI: A Platform for Enhancing Instruction Using Generative AI in Educational Forums*. [45]
- Stahl, M., L. Biermann and A. Wachsmuth (2024), *Exploring LLM Prompting Strategies for Joint Essay Scoring and Feedback Generation*. [35]
- Teng, D., X. Wang, Y. Xia, Y. Zhang, L. Tang, Q. Chen, R. Zhang, S. Xie and W. Yu (2024), "Investigating the utilization and impact of large language model-based intelligent teaching assistants in flipped classrooms", *Education and Information Technologies*, Vol. 30/8, pp. 10777-10810, <https://doi.org/10.1007/s10639-024-13264-z>. [36]
- Tomanek, D. (ed.) (2012), "Does Instructor Type Matter? Undergraduate Student Perception of Graduate Teaching Assistants and Professors", *CBE - Life Sciences Education*, Vol. 11/2, pp. 187-199, <https://doi.org/10.1187/cbe.11-10-0091>. [11]
- Wadams, M. and K. Schick-Makaroff (2022), "Teaching assistant development and contributions in online, MOOC and blended synchronous settings: an integrative review", *Journal of Further and Higher Education*, Vol. 46/8, pp. 1023-1039, <https://doi.org/10.1080/0309877x.2022.2038100>. [4]
- Wise, A. and Y. Jung (2019), "Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making", *Journal of Learning Analytics*, Vol. 6/2, <https://doi.org/10.18608/jla.2019.62.4>. [26]
- Yadav, G. and Y. Tseng (2023), *Contextualizing Problems to Student Interests at Scale in Intelligent Tutoring System Using Large Language Models*. [32]
- Yetişensoy, O. and H. Karaduman (2024), "The effect of AI-powered chatbots in social studies education", *Education and Information Technologies*, Vol. 29/13, pp. 17035-17069, <https://doi.org/10.1007/s10639-024-12485-6>. [15]