Empowering Teachers to Design AI Roles for Work: An Empirical Study of Teachers' AI Role Design Practices

Yujin Kim
Department of Computer Sciences
University of Wisconsin–Madison
Madison, Wisconsin, USA
yujinkim@cs.wisc.edu

Yaxuan Yin
The Information School
University of Wisconsin–Madison
Madison, Wisconsin, USA
yaxuan.yin@wisc.edu

Shamya Karumbaiah
Department of Educational Psychology
University of Wisconsin–Madison
Madison, Wisconsin, USA
shamya.karumbaiah@wisc.edu

Devansh Saxena
The Information School
University of Wisconsin–Madison
Madison, Wisconsin, USA
devansh.saxena@wisc.edu

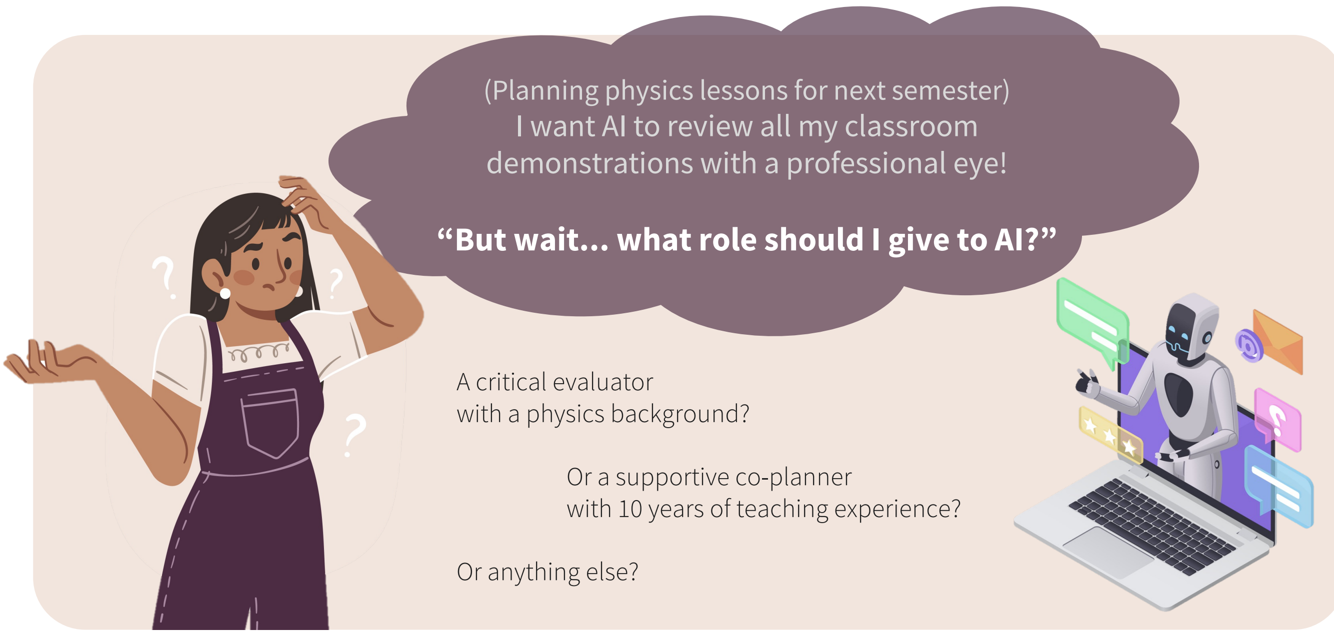


Figure 1: A scenario illustrating a teacher's dilemma in the AI role design process. To support their AI role design process, our empirical study explores teachers' envisioned AI roles, their expectations for each role, challenges they face, and their support needs.

ABSTRACT

As AI agents are increasingly integrated into the workplace, they are gradually acting as partners capable of making independent decisions. To work effectively with these agents, workers must define and assign specific roles to them. However, workers often struggle to translate their complex needs into clear roles that an AI can understand. This challenge is especially significant for teachers, who currently lack organized support to guide them through this process. Developing such support requires a foundational empirical understanding of how teachers conceptualize and adapt AI roles to their specific needs. In this paper, we investigate how teachers envision and design AI roles for lesson planning, a task that encapsulates the complexity of their work. We present findings from an interview study with fifteen K–12 teachers who designed AI roles for lesson-planning tasks using scaffolded worksheets. Our study identifies three categories of AI roles: Educational Practitioners, Professionals outside Education, and Students. We also describe teachers' expectations for these roles and the practical barriers they encounter during the design process. Our contributions include an

CHIWORK '26, June 22–25, 2026, Linz, Austria

ACM ISBN 979-8-4007-2429-9/2026/06
https://doi.org/10.1145/3808045.3808094

empirical foundation for teachers' envisioned AI roles and expectations, the identification of practical challenges in AI role design, and design considerations for tools that facilitate effective teacher-AI collaboration in work.

CCS CONCEPTS

• Human-centered computing → Empirical studies in HCI; **Empirical studies in interaction design**; **• Computing methodologies → Artificial intelligence**.

KEYWORDS

AI Role Design, Teacher, Lesson Planning Task, Worker-AI Collaboration

ACM Reference Format:
Yujin Kim, Yaxuan Yin, Shamya Karumbaiah, and Devansh Saxena. 2026. Empowering Teachers to Design AI Roles for Work: An Empirical Study of Teachers' AI Role Design Practices. In *Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '26), June 22–25, 2026, Linz, Austria.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3808045.3808094

1 INTRODUCTION

Artificial Intelligence (AI) is increasingly integrated into the everyday workplace. Unlike traditional software, AI, such as generative AI, has an agentic nature, capable of independent decision-making aligned with each work context [10, 45]. As a result, this shift has elevated AI from a mere tool to a collaborator for human workers [54, 65]. Within this paradigm, defining and assigning roles to AI has become a critical cognitive process for effective and efficient human–AI collaboration in the workplace [16, 63]. This is because assigned roles determine how labor is divided between workers and AI, which responsibilities each assumes, and what capabilities workers can reasonably expect from AI. When role design fails, risks arise in the workplace. For instance, when AI has a role beyond workers' expectations, workers may lose professional agency, facing risks of job alienation or displacement. Conversely, when AI has a smaller role than expected, workers may be burdened with unexpected post-processing tasks, compromising both work quality and efficiency. Furthermore, since AI is increasingly considered as an independent working agent in the workplace [20, 26], designing AI's role has become not just an individual cognitive exercise, but a shared decision that affects how teams work together [32].

Despite its importance, designing appropriate roles for AI requires careful thought and effort for human workers. To position AI appropriately in their workplace, workers must anticipate task demands, clarify often nuanced expectations, and translate those expectations into a role that AI can meaningfully understand. This process can be especially challenging when workers do not have a clear understanding on AI's capabilities, have limited experience and background with AI, or are uncertain about how different role framings may shape AI outputs. As a result, it can place an additional cognitive and coordination burden on workers.

However, particularly in educational settings, there has been limited systematic support (*e.g.,* guidelines and tools) to help teachers in designing and assigning roles for AI [18, 67]. Based on the Job Demands–Resources model [5], teachers' work involves multiple types of demands arising from their tasks. These include cognitive demands, such as curriculum planning and student assessment; organizational demands, including documentation, meetings, and coordination; emotional demands from interactions with students and parents; and physical demands associated with classroom management and instructional demonstrations [66]. As AI becomes increasingly integrated into teachers' workflows [1, 46, 64], the absence of structured support for AI role design limits its potential to address these diverse demands and function effectively as a collaborative partner. Addressing this gap requires understanding teachers' perceptions of AI roles and the practical challenges they face, to inform the design of effective supports.

To this end, we conducted an initial empirical study examining how teachers envision AI's role in their work, what they expect from various AI roles, what challenges they encounter in designing these roles, and what support they desire to address these challenges. Especially, we focused on a lesson planning task that encapsulates the complexity of teachers' work [75]. This task is a complex, integrative problem-solving process that encompasses elements of assessment design, feedback planning, and administrative work. Teachers must draft official documents aligned with curriculum standards, anticipate students' socio-emotional states, and select instructional strategies based on situated classroom realities. We addressed the following research questions:

- RQ1. What AI roles do teachers envision for lesson planning, and what expectations do they have for each role?
- RQ2. What barriers and opportunities shape teachers' ability to design and adapt AI roles for lesson planning?

We conducted a formative study with fifteen K–12 teachers in the United States. We asked teachers to design AI roles to support their lesson planning tasks through scaffolded worksheet-based exercises and semi-structured interviews. Our study revealed three types of AI roles teachers conceptualize: Educational Practitioners, Professionals outside Education, and Student. In addition, we identified teachers' expectations across three dimensions for each role, which reflect their intentions behind each role design. Furthermore, we investigated novel challenges they encounter in the AI role design process and provided actionable insights to address those challenges. Our contributions are as follows:

- We established an empirical foundation for how teachers conceptualize AI roles for lesson planning and defined their expectations for these roles as collaborative partners.
- We identified practical challenges teachers encounter and the support they desire when designing AI roles for lesson planning.
- We discussed how AI roles are utilized in teachers' lesson planning practice and proposed design considerations to facilitate AI role design in lesson planning contexts.

The remainder of this paper is organized as follows. First, we review empirical research on AI role design in the workplace and its integration into education. Next, we describe our study methodology and materials. Finally, we present our detailed findings and discuss their broader implications for supporting teachers' AI role design.

2 RELATED WORKS

2.1 AI Role Design in Workplace

As AI increasingly supports agentic decision-making and offers free-form prompt-based interfaces, end users are adjusting AI to handle their unique tasks [43]. In this process, the roles that users assign to AI serve as a compressed expression of their underlying perceptions and expectations of it [7, 36, 38]. In other words, a single role label can represent a wide range of specific instructions and behaviors. For instance, Benharrak et al. [7] found that by assigning roles such as 'history professor' or 'strict reviewer', writers could convey their need for specific types of feedback without providing lengthy explanations. In work contexts, defining these roles has been shown to enhance worker control to AI, and improve human-AI collaborative performance [2, 13]. Therefore, AI role design is commonly recommended in contexts where AI is used at work [47].

However, based on the challenges of prompt engineering [77], effectively designing AI roles is difficult for workers. This is because the AI role design requires workers to articulate their expectations in a form that AI can understand. To support this process, we must first understand the specific roles and expectations workers have in mind. Identifying these mental models can provide the necessary scaffolding to help workers formalize their AI roles more easily. Related to this, for instance, Siemon et al. [63] identified four envisioned roles for an AI teammate in the workplace with expert team supervisors: coordinator, creator, perfectionist, and doer, along with accompanying expectations to deepen understanding of AI's responsibilities in teamwork. In addition, Chen et al. [11] explored envisioned AI agents' roles in conflict management for temporary teams, investigating what style (*e.g.,* personality) and intervention modes (*e.g.,* delivery channel) are expected for each role. Furthermore, Johnson et al. [32] examined how workers' design of AI roles is influenced by the AI's embodiment. For example, they found that fantasy characters or animal avatars were expected to bring a sense of whimsy and fun to the workplace.

While these studies focused on investigating AI roles in general work contexts, several works found that practitioners in specific fields (*e.g.,* Tech, Medical) envision AI roles aligned with their field's unique demands. For instance, in software engineering, practitioners designed roles such as senior backend developer, software tester, and Python expert [29]. In addition, in the medical field, AI agents with doctor or patient roles are widely adopted by the nursing community to train nurses in emergency communication [40, 42]. This implies that understanding domain-specific AI role design is essential for effectively supporting workers in creating context-adapted AI roles.

In education, studies have examined how teachers use AI in their work [4, 17, 30]. However, these investigations have primarily focused on the purposes of AI use, rather than on how teachers conceptualize AI roles and what expectations they have for those roles. Without this understanding, it remains unclear what support teachers need to design AI roles effectively in their work.

2.2 AI Integration in Teachers' Work

Teachers handle a wide range of responsibilities in their workplace, from tracking students' academic progress [57] to communicating with parents [28] and attending school staff meetings [44]. Among these, teaching, which is their primary task, is the core task that clearly reflects the complexity of their work. Teaching is fundamentally an ill-structured discipline that takes place in highly complex and dynamic classroom contexts [39]. It requires teachers to translate subject-matter knowledge into forms accessible to students while maintaining pedagogical standards. This involves effectively applying pedagogical content knowledge while adapting to diverse classroom contexts [62, 73, 74]. In this sense, teaching is not merely a matter of classroom management, but a sophisticated practice involving the integration of specialized knowledge domains.

Within teaching, lesson planning constitutes a core task that encapsulates the complexity of this professional work. It serves as the stage where teachers must balance instructional goals, the selection of materials and activities, and the design of assessments [48]. To support such demanding task, AI has increasingly been integrated into teachers' workflows to augment teachers' capabilities rather than replace them [15]. For instance, Fan et al. [23] proposed an interactive system that helps novice teachers create lesson plans using large language models. In addition, AI is expected to help teachers in creating teaching materials and providing personalized guidance during planning [68]. These efforts aim to reduce teachers' workload and save time, which Belloula et al. [6] identified as primary motivations for AI adoption in lesson planning.

However, integrating AI into lesson planning requires teachers to possess not only contextual and pedagogical knowledge but also technical proficiency [39]. As teachers adapt AI to their teaching, this process reshapes their professional knowledge for AI-accompanied classroom practices. In this process, teachers increasingly act as adaptive designers, assigning specific roles to AI to align its capabilities with their work contexts [55]. Despite this growing practice, there has been limited support for teachers in conceptualizing and designing AI roles for lesson planning.

Bridging the Gap. Prior research has examined various AI roles designed by workers in general workplace contexts, highlighting the importance of defining these roles. However, there is still limited understanding of what roles teachers envision for AI, what expectations they have for those roles, and what challenges they face during the role-design process. To bridge this gap, we investigate how teachers conceptualize AI roles, their expectations for each role, the challenges they encounter, and the support they need. We focus specifically on lesson planning because it encapsulates the complexity of teachers' work [75].

3 METHODS

To examine how teachers design AI roles, we adopted Gagné's nine events of instruction as a scaffold [25]. This framework breaks down lesson planning into nine specific events, providing a structured guide for participants' actual practices. Therefore, it ensured that the planning process in our study closely mirrored participants' actual workflows. Based on these nine events, we developed a worksheet to scaffold the AI role design process (see Appendix). Detailed information on the worksheet's content and its use during our interviews is described in Section 3.1.

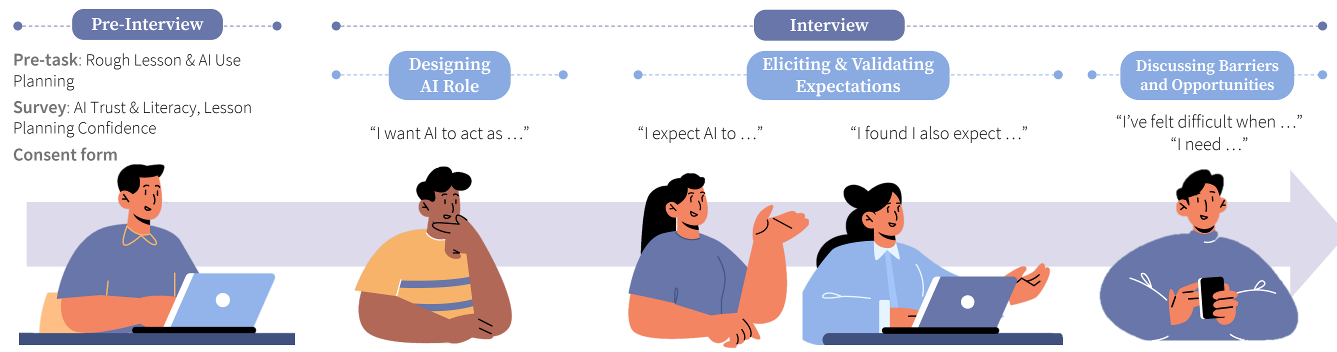


Figure 2: Interview Process. Each task in the main interview sessions is followed by a semi-structured interview.

3.1 Interview Process

Figure 2 summarizes the overall study procedure, which was designed to address our research questions in the context of lesson planning. The process consisted of a pre-interview phase and an interview phase. All study procedures were approved by the university's institutional review board.

3.1.1 Pre-Interview (25min). Once recruited, participants received an email with links to the pre-interview tasks, which included a worksheet for pre-interview tasks, an online survey, and a consent form. Regarding the worksheet, participants were first reintroduced to Gagné's Nine Events of Instruction to help them recall the lesson planning task flow (see Figure 4 in Appendix). Following this, they developed a rough lesson plan based on these nine events (see Figure 5). After that, they identified which of the nine events they would use AI for and described their intended purposes for each. These tasks were designed to externalize participants' initial planning decisions and encourage prior reflection on AI use, allowing the actual interview to focus on AI role design following this initial phase.

To measure AI trust and literacy, we used the scales developed by Jian et al. [31] and Wang et al. [70], both of which are widely adopted metrics for AI trust and literacy in Human-AI Interaction research. Each scale consists of 12 items on a 7-point scale. Finally, lesson planning confidence was measured using a customized question: "How confident are you in the lesson planning task?"

3.1.2 Interview (90min). At the beginning of the interview, the interviewer reviewed the study's purpose and process and obtained verbal consent for audio and video recording. Participants were then provided with a main-task worksheet, into which the interviewer had integrated the participants' pre-interview task responses to help them recall their initial ideas and ensure task continuity. After reviewing their pre-interview responses, participants were introduced to an AI role-prompting orientation. This orientation aimed to establish an understanding of how AI roles can be specified through prompts, using several role-prompt examples.

Designing AI Role. The first interview task was designing AI roles. As shown in Figure 6, participants were asked to select up to three purposes for which they would most likely assign a role to AI in the lesson plan they had designed during the pre-interview task. They then designed specific AI roles for these purposes. During this process, the interviewer also asked follow-up questions on the expected benefits of designing AI roles for those three AI purposes to capture their expectations of AI role design. In addition, the participants were asked how they designed those AI roles.

Eliciting & Validating Expectations. The second task in the interview was eliciting and validating their expectations of the designed AI roles. For that, as shown in Figure 7, participants were asked to clarify what they expected to see in the AI's responses for each role. To check if the provided expectations were comprehensive, we additionally asked them to validate their expectations while using AI. In this process, they were asked to write down prompts they would like to give to AI on the worksheet regarding the task contexts and AI's role (*e.g.,* "You are an endangered species specialist working in the US. Create referent case studies fitting to my high school biology class of different North American endangered species."). Then, they gave the prompts to the AI to see the AI's responses and identify their further expectations. To reflect their actual usage context, participants were asked to use the LLM services they regularly engage with for their work (*e.g.,* Gemini, Claude, ChatGPT).

Discussing Barriers and Opportunities. The interview concluded with a reflective discussion focused on barriers and opportunities related to AI role design. Participants were asked what the challenges were while designing the AI roles. In addition, we asked them what kinds of support they needed in the AI role design process.

3.2 Participants

Fifteen participants were recruited for the study via university mailing lists and snowball sampling [51]. The inclusion criteria required that they be K-12 teachers in the United States, have prior experience using large language models, and be fluent in English. The last criterion was necessary because the interview was conducted in English. Participants were compensated $50 for a 90-minute online interview and 25 minutes of pre-interview tasks. Table 1 shows the characteristics and backgrounds of all participants in our study.

3.3 Analysis

During the interview phase, we collected two types of data: written responses on the worksheets and verbal data from the interviews.

Table 1: Participants' characteristics and backgrounds. Higher scores on a 7-point Likert scale indicate greater confidence in lesson planning, AI trust, and AI literacy.

ID	Current Target Grade	Subject	K-12 Teaching Experience	Confidence in Lesson Planning Task (1~7)	AI Literacy (1~7)	AI Trust (1~7)
P1	12th	Biology	20 years	5	4.25	5
P2	11th	History	3 years	5	4.83	5.08
P3	10th	Digital Media	9 years	7	4.5	3.33
P4	10th	Chemistry	35 years	7	3.83	3.5
P5	9th	English	25 years	6	4	4.33
P6	8th	English	7 years	5	4.67	3.75
P7	7th	History	34 years	7	5.67	4
P8	6th	Spanish	4 years	6	5.5	4
P9	6th	Science	1 year	4	4.25	4.17
P10	5th	Math	27 years	6	5.08	5.5
P11	4th	Social Studies	3 years	5	4	4.17
P12	4th	Math	20 years	5	4.25	5
P13	3rd	Science	28 years	7	4.58	2.5
P14	2nd	English	13 years	4	4.33	3.42
P15	1st	Social Studies	6 years	6	4.08	2.92

The written data were thematically coded [9] by the first and second authors. Regarding the verbal data, the first author transcribed the full interview recordings using Whisper Transcription [27], a locally-run AI-based transcription application. This ensured that participants' audio data was not transmitted to any external cloud service, addressing data privacy considerations. The transcripts were then manually refined for accuracy by cross-checking them with the recordings. The first author then familiarized themselves with the data and developed an initial codebook for each interview question through inductive open coding. To validate the codebook efficiently, the second author independently coded a randomly selected 20% of the responses per question using the codebook, and inter-coder reliability was assessed. Disagreements were resolved through discussion between the two coders, during which the codebook was iteratively revised, and code definitions were clarified until consensus was reached. Once the codebook was finalized, the two authors divided and coded the remaining data. For the survey data from the pre-interview phase, we followed the standardized analysis methods described by Jian et al. [31] and Wang et al.[70]'s work.

3.4 Research team positionality

Our research team has worked closely with K–12 educators. It consists of two faculty and two student researchers, including Korean, Chinese, and Indian members. Three of us are female, and one is male. The first author, who led and conducted the majority of the data collection and analysis activities, is a PhD student with an undergraduate and master's degree in computer science. The second author has worked on responsible AI in education. The third author has been conducting research on K–12 education for over ten years. The fourth author has been conducting public care work for over nine years.

4 RESULTS

4.1 RQ1: What AI roles do teachers envision in practice, and what expectations do they have for each role?

In our study, we identified three categories of roles that participants envisioned for their lesson planning task: *Educational Practitioners*, *Professionals outside Education*, and *Students*. Across these role categories, we also identified three shared expectation dimensions: (1) *Target Audience Calibration*, ensuring whether AI outputs are developmentally and contextually appropriate for students, considering factors such as prior knowledge and grade-level pedagogical requirements; (2) *Practical Implementation*, addressing whether AI outputs are usable and adoptable in real classroom settings under practical constraints such as time, resources, and curricular alignment; and (3) *Content Quality*, concerning whether AI outputs are factually accurate and qualitatively well established. While these dimensions were common across all roles, their relative emphasis and specific expectations varied by role, as summarized in Figure 3. In this section, we describe the participants' envisioned roles and their expectations for each role, with more detailed examples.

4.1.1 *Educational Practitioners: Pedagogical Insiders for My Work.*

Our results showed that in about 55% of the roles designed by the participants, AI was positioned as an educational practitioner for their tasks. In these cases, AI was framed as a pedagogical insider, not merely a generative tool, but a knowledgeable entity. It was expected to have professional expertise in instructional design and teaching. In detail, the following roles were identified in our study: instructional designers who design and develop educational materials, courses, and learning experiences (*e.g., Curriculum instructors*, *Instructional designers at the Wisconsin Department of Instruction*), K-12 teachers who educate students in primary and secondary schools (*e.g., Language arts teachers*), professors who educate at the college

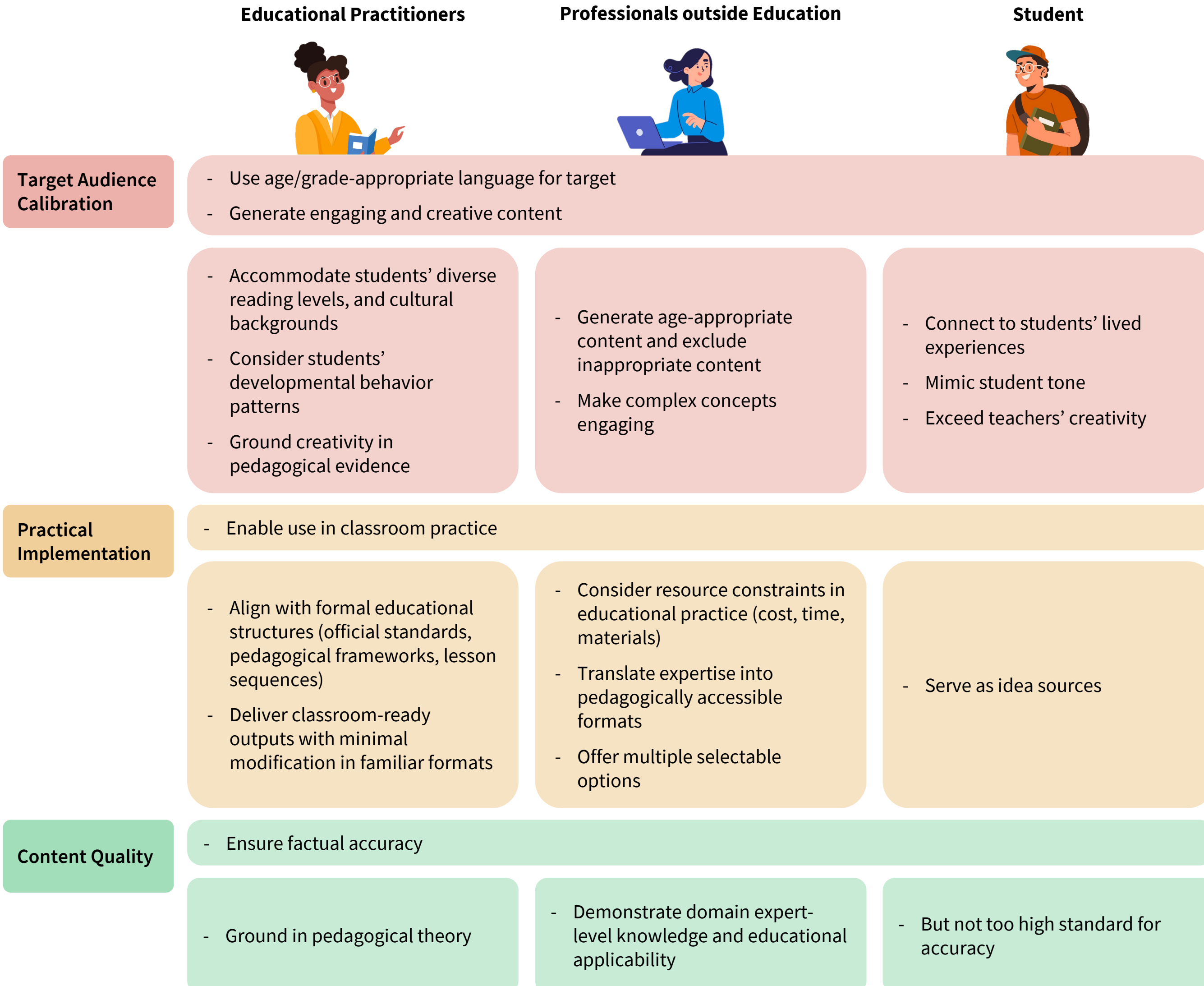


Figure 3: Summary of participants' expectations for each role. This highlights both shared expectations across the three roles and role-specific expectations within each dimension. In total, 44 roles were designed by participants.

level (*e.g., Chemistry professors*), and general educators who teach broader audiences beyond K-12 contexts.

The educational practitioner roles were typically accompanied by two types of role attributes. The most frequently adopted attributes concerned the target students' characteristics and context (79%). These student-related characteristics included students' language backgrounds, academic diversity, and grade level. For instance, P2 designed a history teacher role for AI, specifying that its target students are bilingual. Similarly, P6 designed AI as a language arts teacher for students with various reading and ability levels. Additionally, P14 envisioned AI as a teacher of a mixed-ability class for second-grade students. The other commonly observed role attribute concerned the pedagogical approaches and values the role needs to adopt or consider (38%). This characteristic was only observed in the educational practitioners roles in our study. For instance, P5 designed a high school English educator who favors the GRR model, an instructional model in which the teacher gradually transfers responsibility for learning from the teacher to the students [52]. In addition, P9 envisioned a 6th-grade science teacher committed to providing materials that are accessible and rigorous.

Expectations on Target Audience Calibration. Among expectations for educational practitioners role, the target audience calibration dimension accounts for 34.8% of all expectations. Specifically, participants expected AI-generated content to align with students' developmental needs and contexts (21.1%), while remaining firmly grounded in pedagogical evidence (13.7%).

First, participants emphasized that AI-generated content should utilize kid-friendly language appropriate to students' developmental levels, avoiding overly professional or academic tones. Beyond linguistic simplicity, participants expected AI to reflect developmentally appropriate students' classroom behaviors. For instance, P14 noted that "*[The learning objectives AI generated should] be easy for second graders to read aloud.*", highlighting the expectation that AI recognizes that second-grade learning often occurs through oral engagement, which is prioritized over silent reading. Furthermore, participants required AI to accommodate multiple reading levels simultaneously. P4, for instance, expected AI to provide content at a high school (9th grade) reading level, while providing a parallel version at a 5th–6th grade level. This reflects the reality of classroom heterogeneity, where a single difficulty level cannot adequately serve all learners. Additionally, P15 introduced a cultural dimension, emphasizing that AI-generated content should be potentially culturally relevant for a diverse classroom. This highlights the need for AI to provide context-aware adaptations that mirror the multi-layered reality of actual classrooms.

Meanwhile, participants expected AI to generate creative and engaging content that remains firmly rooted in pedagogical evidence. While for other roles, such as students (see Section 4.1.3), participants prioritize open-ended ideation, the education practitioner roles were required to have innovation to be pedagogically sound. For instance, P13 and P10 rejected traditional instruction formats, such as the fill-in-the-blank, in favor of activities rooted in learning science that encourage students to think outside the box. However, this creativity was expected to operate within strict evidential boundaries. Similarly, P8 stated, "*I don't expect too much creativity, but learning objectives based on language learning research.*", underscoring that creative output must not come at the expense of pedagogical requirements. This emphasizes that AI, in the role of an educational practitioner, is expected to act as a rigorous expert rather than a mere generator of novel ideas.

Expectations on Practical Implementation. This dimension accounts for 36.8% of all expectations within the educational practitioner role. Participants primarily expected AI to provide outputs that are immediately applicable in educational practice by aligning them with formal institutional structures, rather than merely covering relevant topics (21.1%). Additionally, many envisioned that AI in this role could significantly improve work efficiency (15.7%).

First, to facilitate seamless integration into educational practice, participants expected the AI to use official standard terminology, adhere to established pedagogical frameworks, and ensure coherence with existing lesson sequences. For instance, P1 required AI to generate rubrics that align with national science standards, emphasizing the need for consistency with institutional discourse. Beyond standards compliance, participants expected AI to integrate content seamlessly into their specific instructional contexts. For example, P5 expected AI outputs to be consistent with prior lesson plans, while P2 expected AI to utilize standardized frameworks, such as 'SMART', which are commonly employed by them to design clear and effective learning objectives [22, 41, 53]. These expectations highlight the importance of aligning AI-generated content with the actual standards and routines of the educational field.

Meanwhile, participants sought classroom-ready outputs that require minimal modification to enhance work efficiency. For instance, P1 expected deliverables in formats commonly used in education, such as Google Docs or Google Slides. Similarly, P12 preferred printable formats, while P3 and P9 emphasized that materials should be ready for immediate use with little to no editing. They further expected AI to provide multiple versions tailored to diverse student needs, reflecting routine instructional tasks that they must typically perform. This demand for professionally formatted, immediately usable resources distinguished educational practitioners from other roles, whose concerns regarding resource utility were more general and less integrated into specific professional workflows.

Expectations on Content Quality. This dimension accounts for 28.4% of all expectations within the educational practitioner role. Participants primarily expected AI to ensure information accuracy and credibility (23.2%) and maintain pedagogical integrity in its content (5.2%).

For instance, P2 required AI to rely exclusively on provided, verified sources to generate factually precise reading-guide notes, reflecting a cautious stance toward potential AI-generated hallucinations. Beyond such factual correctness, participants expected AI to ensure that information is structured according to validated instructional theories rather than being a mere collection of facts. In other words, for the educational practitioner role, content quality meant that the content must be both factually sound and theoretically robust from a pedagogical perspective. For example, P14 expected AI to create materials that incorporate the think-aloud technique [69] to ensure that the cognitive process of the compare and contrast concept was accurately modeled. Similarly, P15 sought map-related hook questions that correctly activated students' prior knowledge, reflecting the pedagogical principle that new learning is most effective when anchored to existing mental frameworks [3]. This contrasts sharply with the professionals outside education, who prioritized domain-specific credibility over pedagogical accuracy, and the student role, which showed less concern for content accuracy (see Sections 4.1.2 and 4.1.3).

4.1.2 Professionals outside Education: Authentic Domain Experts for My Work. In 25% of the designed roles, AI was envisioned as a professional outside education, bringing specialized expertise from non-pedagogical domains into participants' tasks. Unlike educational practitioners who are pedagogical insiders, these roles contribute disciplinary knowledge or technical skills from fields such as science, business, or technology. Specifically, two types of roles were identified in our study: subject-related domain experts, who provide deep disciplinary knowledge (*e.g., Endangered species specialist* for a biology class, *Mathematician* for a math class), and task-related domain experts, who assist with professional technical skills (*e.g., secretary* for collecting information, *graphic designer* for ideating visual materials).

The professionals outside education roles were typically supplemented with three types of role attributes. The most frequently adopted attribute were specialized fields and topics (45%). For instance, P7 designed a history expert specifically focused on the study of feudalism in medieval Europe to provide depth to her history curriculum. Another commonly observed attribute was specialized tasks (36%). Related to this, P9 envisioned AI as a researcher proficient in data curation, while P3 designed a professional secretary to provide structured context for gathered information. Notably, these task-oriented attributes were often paired with specific workplace affiliations, such as the *U.S. Fish and Wildlife Service (USFWS)*, a characteristic observed exclusively within this role category. By contrast, attributes related to target students' characteristics and context were adopted in only one case, highlighting a clear distinction from the educational practitioner roles where such context was a primary concern.

Expectations on Target Audience Calibration. Among the expectations for the professionals outside education roles, the target audience calibration dimension accounts for 30.4% of all expectations. Specifically, participants expected AI-generated content to be developmentally appropriate in both language and substance (17.4%), an expectation similar to those for the educational practitioner roles. At the same time, they required that domain-specialized knowledge be presented in an engaging and appealing manner to capture student interest (13.0%).

First, participants emphasized that even when AI operates as an external specialist, it must produce age-appropriate material tailored to students' grade levels. For instance, P1 expected AI, acting as an endangered species specialist, to utilize kid-friendly language when presenting biodiversity case studies to students. Beyond linguistic simplification, participants expected AI to perform as a content filter to ensure safety and suitability. P7, for instance, required AI to suggest historical artifacts for middle school students while strictly excluding profanity, overt sexuality, or violence. Furthermore, P14 expected AI to provide several visual materials designed for second graders, ensuring that the depicted differences between those visuals were discernible at that age level.

Meanwhile, participants expected AI to convert complex domain concepts into creative and engaging content that captures student interest. Unlike the educational practitioner roles, which prioritized pedagogical evidence, professionals outside education were expected to prioritize the appeal of their specialized knowledge. P4, for example, asked AI for hydrogen bonding demonstration ideas, expecting them to be interesting enough for students to perform at home for others. Similarly, P12 envisioned AI as an award-winning mathematician who could compose a song about symmetry to make the math class more captivating. These expectations reflect a desire for domain experts to make their specialized knowledge not just accessible but genuinely appealing to young learners, acting as inspirational specialists who bridge the gap between high-level expertise and student curiosity.

Expectations on Practical Implementation. This dimension accounts for 43.5% of expectations for the professionals outside education role. Participants primarily expected AI to consider resource constraints in educational practice (21.7%). Additionally, they envisioned that AI in this role should translate complex technical content into comprehensible formats (13.0%) and provide multiple options for teacher adaptation (8.7%).

First, to ensure that specialized knowledge is applicable in real-world school settings, participants emphasized that AI must account for the practical limitations of both classroom and home environments. For example, P4 expected AI, acting as a purchaser skilled in cost-saving, to recommend materials for chemical experiments that are locally available and reasonably inexpensive. She further required that these demonstrations be doable within a reasonable class period. Similarly, P7 noted that assessment questions suggested by the AI should be easily carried out by students even at home. This highlights an expectation that AI-generated disciplinary content must remain feasible within the material and temporal constraints of actual educational practice.

Meanwhile, participants expected AI to translate complex technical content into a digestible format for non-specialist students. P3, for instance, expected AI to provide metaphors or explanations for programming code while assuming the AI acts as a developer in his digital media class. This illustrates the expectation that domain experts make their knowledge comprehensible to students who lack a background in the field. Similarly, P15 expected AI, acting as a children's literature librarian, to offer ideas for deepening learning on books used in the lesson, reflecting a need for pedagogical scaffolding to accompany domain-specific content.

Furthermore, participants sought multiple options from AI to facilitate teacher choice and adaptation. For instance, P9 expected at least four to five possible responses to potential student questions. This emphasis on quantity and variety, providing teachers with a range of options to select and adapt rather than a single prescriptive solution, distinguishes the practical orientation of the professionals outside education role.

Expectations on Content Quality. This dimension accounts for 26.1% of expectations for the professionals outside education role. Specifically, participants expected AI to demonstrate professional-level knowledge in its specialized fields. For instance, P1 expected AI, as an endangered species specialist, to demonstrate accuracy and credibility by providing precise information on biodiversity. In addition, P3 expected AI to act as a developer by demonstrating knowledge of programming languages with clear code prompts and well-documented code for his class. Similarly, P7 expected AI, acting as a history expert, to provide quality artifacts representing the study of feudalism in medieval Europe. She noted that such examples should demonstrate expertise in both the historical and educational fields. These cases highlight a demand for dual expertise, requiring AI to bridge disciplinary standards with classroom needs. This focus contrasts with the educational practitioner role's priority on pedagogical accuracy and integrity (see Section 4.1.1). It also differs from the student role, which placed limited focus on disciplinary expertise as described in Section 4.1.3.

4.1.3 Student: A Simulated Task Target. In 14% of the designed roles, AI was envisioned as a K-12 student representing the primary target audience for the participants' lesson planning task (*e.g., a current high school student*). This role was typically complemented by specific attributes related to grade levels. For instance, P6 designed AI as an 8th-grade student to investigate the interests and perspectives of that specific age group.

Expectations on Target Audience Calibration. Among the expectations for the student roles, the target audience calibration dimension is the most prominent, accounting for 78.2% of all expectations. Within this dimension, participants primarily expected AI acting as students to adopt behavioral patterns and lived experiences appropriate for specific grade levels (21.7%), while an additional 21.5% of cases focused on matching the tone and language of students. Furthermore, 34.8% of expectations involved generating creative and engaging content for students that exceeds the teacher's own imaginative capacity. Collectively, these requirements reflect a desire for AI to act as a lesson target proxy, which is a simulated learner that mirrors the cognitive and linguistic realities of the target learners.

First, participants expected AI to connect educational content to students' lived experiences by simulating age-appropriate perspectives. For instance, P6 desired AI to suggest introductory examples related to media and trends popular with the specified age group. P8 further expected AI to imitate a typical 6th grader to generate assessment questions that allow students to connect new learning with their own daily lives. By bridging academic content with authentic student viewpoints, these expectations emphasize the role of AI as a learner proxy that validates the relevance of the curriculum from a student's perspective.

In terms of language, participants emphasized that AI must adopt the specific tone and vocabulary typically used by the target grade level. For example, P5 expected AI's tone to match that of a student when generating examples for complex rhetorical concepts like ethos, ensuring the explanation felt familiar to the learners. Similarly, P11 required AI to translate learning objectives into student-friendly language, prioritizing readability and comprehensibility over academic jargon. These cases highlight a need for AI to simplify linguistic complexity, ensuring that even high-level objectives remain clear and accessible to all students.

Finally, practitioners expected AI to serve as a source of fresh creativity to enhance student engagement. P6, for instance, noted that AI should produce a more captivating list of examples than they could brainstorm independently. Unlike the professional roles that relied on disciplinary authority, the student role was expected to provide engaging materials that might elude traditional teacher-led approaches.

Expectations on Practical Implementation. This dimension accounts for 13.0% of expectations, focusing on AI providing student perspectives as useful idea sources for classroom practice rather than fully polished instructional materials. In this context, participants did not necessarily require the AI to generate a final product. Instead, they valued its ability to act as a pedagogical informant by offering a learner's point of view. For instance, in a Spanish class on colors, P8 expected AI to suggest everyday situations where native English-speaking students might have encountered specific colors. By providing such student perspectives, AI serves as a brainstorming partner, offering foundational ideas that teachers can then refine to bridge abstract concepts with students' lived experiences.

Expectations on Content Quality. The remaining 8.7% of expectations focus on the factual accuracy of AI-generated content. Interestingly, expectations for accuracy in the student role were lower than for other roles, such as educational practitioners. Rather than prioritizing pedagogical or disciplinary precision, participants viewed accuracy in this context as a matter of how well it reflects students' characteristics. For instance, while asking AI to generate examples relevant to student interests, P6 noted that "*The only issue with factual correctness would be a matter of opinion, as certain shows may be popular with some subgroups of students but not others, so some examples may not land.*" This suggests that factual precision was less emphasized for the student role, as long as the content remains authentic to the students' experience.

Key Takeaways for RQ1. In summary, these findings demonstrate that teachers envision versatile AI roles in their lesson planning, categorized into educational practitioners, professionals outside education, and students. While each role serves a distinct purpose, our participants consistently evaluate their effectiveness based on target audience calibration, practical implementation, and content quality. These takeaways suggest that for an AI to be truly useful for teachers in their lesson planning tasks, it must transcend simple content generation by balancing disciplinary expertise with students' specific developmental needs and the institutional constraints of the K-12 environment. Ultimately, the diversity of these envisioned roles reflects the multifaceted nature of teachers' professional expertise and their desire for AI to function as a context-aware collaborator.

4.2 RQ2: What barriers and opportunities shape teachers' ability to design and adapt AI roles in practice?

Participants showed clear expectations for designing AI roles in the lesson planning task. They expected designing a role for AI to help with tasks that were difficult to perform alone, took too much time, or required processing a lot of information ($N = 8$). They also hoped that assigning roles to AI would help them break away from routine thinking and generate new ideas ($N = 4$). These expectations were based on the view that well-designed roles could clearly represent the task context and the teacher's intentions. To create these roles, most participants started from scratch ($N = 12$), while five reused roles from previous tasks, and two referenced acquaintances with relevant expertise. However, across all these approaches, participants often faced barriers when trying to design roles that matched their expectations. They described recurring problems alongside opportunities that could make the design process more effective and easier.

4.2.1 AI-Mediated Scaffolding to Surface Implicit Expectations.

Barrier. A primary barrier was the difficulty of recognizing and expressing expectations for AI roles at the start of the design process. Four participants reported that they struggled to identify what they expected from the AI for each task. They were often not consciously aware of their expectations until they observed the AI's output or reflected on conversations with the interviewer. This made it difficult to decide which role to design at the outset. As P1 noted, "*Maybe I don't know exactly what I want, but I know what it will look like when I see it.*" Similarly, P13 explained that expectations that felt obvious internally were hard to clarify explicitly, stating, "*Things that I am thinking about, which to me are obvious, are hard for me to clarify.*" Two participants further reported having only a vague sense of their expectations rather than a clear understanding. P8 reflected on this challenge, saying, "*It was hard when it was something that was more of an implicit expectation. I didn't even realize it.*"

Opportunity. In response to this barrier, five participants suggested opportunities for AI-mediated scaffolding to help surface and clarify expectations during role design. Participants envisioned the AI asking questions to guide them in clarifying their goals and specific requirements for the task. As P8 suggested, "*I could probably ask the AI, 'Can you tell me five things you need me to define?'... and then I'd be like, 'Okay, you're this, you're that.'*" Such interactions were seen as a way to make implicit expectations explicit before committing to a particular role. This would enable more intentional role design rather than reactive adjustment after seeing outputs.

4.2.2 Role Recommendation and Peer Reference Points to Identify Optimal Roles.

Barrier. Even when participants had a general sense of their expectations, many struggled to determine which AI role would be most appropriate for a given task. Seven participants reported difficulty identifying an optimal role and expressed uncertainty about whether the roles they designed would yield the best results. P4 described this uncertainty about selecting appropriate expertise, noting, "*The challenge is whether you actually have any experience in figuring out who you could ask this to. Perhaps I didn't pick the exact best expert, and maybe there is someone else who would know more about this.*" Similarly, P9 questioned whether the roles they chose meaningfully shaped the AI output, stating, "*I'm just thinking whether these are actually the best roles to tell it to get the results that I want. Maybe it would be more valuable to have it be something else. Maybe it would be more valuable to be a scientist who studies phase change.*"

Participants also reported difficulty deciding how much detail to include when defining roles. Four participants struggled to find a balance between being overly general and overloading the AI with too much information. P5 questioned the value of extensive role specification, stating, "*One of the challenges is just knowing how much detail to put into the role and whether there is value in spending a lot of energy and time wordsmithing the role description when it may or may not improve results.*" While P6 expressed frustration with being too general, P14 worried about overwhelming the AI with too much information.

Opportunity. Participants described several opportunities to reduce this uncertainty. Four participants envisioned AI support that could recommend potential roles based on task descriptions and stated expectations. As P12 explained, "*If I was not getting what I wanted, then I would just have the conversation with AI and be like, 'Here's what I'm looking for, what role would you recommend?' AI will be like, 'Here's a list of roles that we recommend,' and then I can just pick from there.*" Similarly, P15 described asking the AI for role suggestions when working in unfamiliar topics, such as "*What kind of person can teach [this topic] really well?*"

In addition to AI support, four participants highlighted opportunities to learn from other teachers' experiences. P9 suggested referring to roles used by peers, such as "*a data table showing which roles teachers assigned for particular expectations and the percentage to which the AI satisfied those teachers in its responses.*" These peer reference points were seen as a way to reduce uncertainty by basing role selection on collective practice rather than individual guesswork.

4.2.3 Structured and Collective Evaluation for Iterative Role Design.

Barrier. To manage uncertainty around role selection, several participants tried to improve their AI roles through trial and error. As P12 explained, "*I'm always trying to think about what would be the best role to put in there. I would probably see what response I get and then switch it if I like it.*" However, this strategy introduced a further barrier related to evaluation. Four participants reported difficulty judging whether AI outputs fulfilled their expectations, particularly when they lacked expertise in the subject. P2 noted the difficulty of judging accuracy when encountering unfamiliar content, stating, "*If there's something I don't know about, being able to judge whether the information is accurate, being able to catch that is tough.*" Similarly, P6 described difficulty evaluating how relevant the output was for students, stating, "*I don't exactly know what's super relevant for kids, so it's hard to evaluate whether or not that is accurate.*"

Opportunity. Participants identified multiple opportunities to help them evaluate how well the AI performed its assigned role. Two participants expressed a desire for structured evaluation criteria, such as checklists or rubrics, to guide their own assessment of whether the AI fulfilled its role. P6 suggested "*Almost like a word bank of certain criteria that I could use to evaluate it*", while P8 suggested "*A worksheet or something where it's like, 'Make sure it includes these three things.'*" These structured supports were seen as particularly helpful when participants lacked the expertise to verify accuracy on their own.

Participants also emphasized the value of shared evaluation experiences through teacher communities. P15 expressed a preference for human feedback, stating, "*I probably would use a human being, to be honest, like another teacher, to say, 'Do you think that this [AI responses] makes sense?'*" P10 similarly noted that feedback from other teachers could help determine whether AI responses were clear and understandable from perspectives other than their own. Beyond human feedback, P2 suggested automated verification tools, stating, "*There could be plugin functions that evaluate whether the information is correct or not, like a Chrome extension, a fact verifier.*"

Key Takeaways for RQ2. In summary, these findings highlight that teachers encounter persistent barriers throughout the entire AI role design process for their lesson planning task, both before and after designing AI roles. From the initial struggle in articulating implicit expectations to the subsequent difficulty in evaluating AI outputs, teachers find themselves navigating a cycle of uncertainty that solitary trial and error cannot easily resolve. To bridge these gaps, they desire a multi-layered support system: interactive AI-mediated scaffolding during the design phase, access to collective peer experiences for role selection, and standardized frameworks or community feedback for structured evaluation. Ultimately, these takeaways emphasize that effective AI role design requires more than individual effort. It calls for a comprehensive infrastructure involving intelligent assistance, teacher communities, and standardized evaluation tools.

5 DISCUSSION

Our study examined how K-12 teachers envision and design roles for AI in lesson planning, including the specific expectations they hold for each role. We also identified barriers that make the AI role design process challenging and the support teachers desire in practice. Teachers cast AI in three kinds of roles: educational practitioners, professionals outside education, and students. Each role is accompanied by distinct expectations regarding target audience calibration, practical implementation, and content quality. However, participants struggled to surface these expectations up front, to decide when a role was appropriate for their intended AI use, and to evaluate whether the AI had actually fulfilled the assigned role. In the following sections, we connect these findings to prior work on AI roles in the workplace and worker-centered AI, translating them into design implications for teacher-centered AI tools.

5.1 AI Roles as Medium for Teacher Expertise and Pedagogical Responsibility Management

This section discusses how AI roles can act as more than just information about the AI's functional capabilities. Based on our findings in the lesson planning context, we argue that for K-12 teachers, role design is a process of projecting their professional knowledge and strategically managing the distribution of pedagogical responsibility between the teachers and AI.

5.1.1 Replicating Internal Expertise through AI Personas. Rather than describing AI roles through abstract functional capabilities such as idea generator, copy editor, or mediator, as in prior creative-writing studies [38, 76], teachers in our study defined AI roles by specifying professional expertise and credentials (*e.g.,* "a professional Spanish teacher who has their K–12 license", "an endangered species specialist working in the US"). We argue that this role-design practice is a process of externalizing teachers' tacit pedagogical expertise into explicit role specifications. Drawing on their domain knowledge, teachers anticipate what expertise a given instructional context demands (*e.g.,* lesson planning). This can include subject standards, student population, and classroom realities. They then encode these anticipations into the AI's role. This pattern is visible in our findings, for instance, 79% of educational practitioner roles included specific student population attributes, and 38% specified pedagogical approaches such as the GRR model or think-aloud that are typically left tacit in classroom practice. Therefore, our findings surface how teachers design these specifications into AI's roles, suggesting that role design is an act of articulating knowledge that is ordinarily left implicit in classroom practice.

This articulation resembles how professionals in other domains embed knowledge into AI. Writer-defined AI personas configure LLMs as imagined readers to elicit audience-specific feedback [7]. Similarly, virtual persona agents let designers brief LLMs on imagined consumers for market-tailored critique [61]. Across these settings, users draw on their professional knowledge to specify who or what the AI should emulate, so that its outputs reflect the standards and constraints of their domain. Our findings extend this pattern to K-12 teachers' work, where teachers configure AI as peer teachers, experts from other domains, or students, depending on the pedagogical task at hand. The AI role thus becomes a medium through which teachers externalize and interact with their own professional standards.

5.1.2 Regulating Pedagogical Responsibility. Through AI Roles Based on the expectations of each role, our findings suggest each role was designed with different expectations of authority and risk inherent in the lesson planning process. For instance, for the educational practitioner roles, teachers expected the role should be thoroughly grounded in pedagogical values and considerations (*e.g.,* Consider students' developmental behavior patterns, Deliver classroom-ready outputs with minimal modification). Meanwhile, for the professionals outside education roles and student roles, teachers anticipated that the role can play a supplementary role (*e.g.,* Make complex concepts engaging, Serve as idea sources). Our further analysis shows this well. As shown in the Table 2, we analyzed for which types of tasks each of the three roles was designed based on Keppler et al.'s four interactional modes [35], specifically *Make for me*, *Find for me*, *Jumpstart for me*, and *Iterate with me* [1]. This supports that teachers do not view these modes as purely technical choices. Instead, they layer roles onto these modes to calibrate the authority and risk inherent in the planning process.

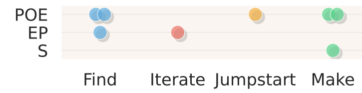
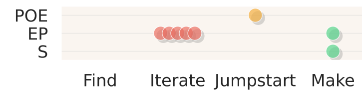
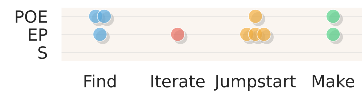
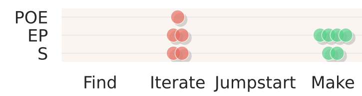
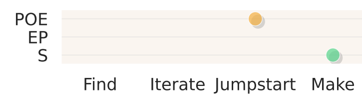
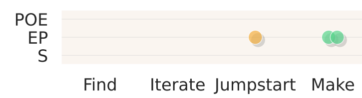
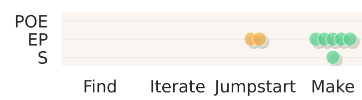
For instance, educational practitioner roles were most frequently assigned to high-stakes lesson-planning events, particularly *Assessing Performance* (23.3%) and *Informing Learners of Objectives* (20%). In the events, this role was frequently paired with the *Make for me* and *Iterate with me*, demonstrating that teachers were willing to delegate writing feedback on students' performance and refine learning objectives with AI when the AI was framed as a professional peer. In contrast, professionals outside education roles were concentrated in *Gaining Attention* (41.7%) and *Presenting Stimulus* (33.3%), accounting for the majority of instances where these roles were adopted in the lesson-planning task. In the events, this role was frequently paired with the *Find for me* and *Make for me*, demonstrating that teachers were willing to utilize AI with domain experts as a creative consultant for lower-stakes brainstorming. These specific combinations of role and mode allow teachers to adjust their level of involvement based on the pedagogical significance of the event. Our findings illustrate how teachers flexibly redistribute pedagogical responsibility across imagined collaborators (through AI roles) to manage the cognitive, organizational, and emotional demands of their work. Ultimately, the choice of an AI role is a strategic decision used to control the level of authority granted to the system.

5.2 Considerations for Developing Supports in Teachers' AI Role Design

5.2.1 AI-Mediated Reflection on Teacher Expectations. Teachers expressed a desire for AI-based assistance in role design for lesson planning task. Specifically, they envisioned AI helping them specify their expectations and suggesting appropriate roles in their work. In other words, rather than requiring complete specifications upfront, the AI should engage in iterative elicitation to surface and refine

[1] **"Make for me"** asks AI to produce fully developed content (e.g., problems, quizzes, essays), using it as a task executor. **"Find for me"** seeks pre-existing facts, quotes, or resources, treating AI like a search engine. **"Jumpstart for me"** asks AI to start complex materials, such as lesson or unit plans, using it as a catalyst. **"Iterate with me"** requests feedback or refinement of ideas and teaching approaches, engaging AI as a sounding board.

Table 2: Frequency of AI interaction mode adoptions across Gagné's Nine Events, categorized by AI role. 'EP' indicates Educational Practitioners, 'POE' indicates Professionals outside Education, and 'S' indicates Student Roles. 'N/A' means that there were no roles designed for Gagné's events.

Gagné's Nine Events	Definition	Distribution of AI Interactions
Gain attention	Present introductory activities that engage learners (*e.g.,* Showing relevant news video)	
Informing Learners of Objectives	Clearly state the learning goals and outcomes (*e.g.,* Stating clear learning goals)	
Stimulating Recall of Prior Learning	Encourage learners to remember and connect previous knowledge (*e.g.,* Reviewing the previous lesson)	**N/A**
Presenting Stimulus	Introduce new content and information (*e.g.,* Introducing core concepts)	
Providing Learner Guidance	Offering instructions and strategies to help learners understand and process the new content (*e.g.,* Offering step-by-step tips)	
Eliciting Performance	Have learners practice what they have learned (*e.g.,* Solving various practice problems)	
Providing Feedback	Give constructive feedback on learners' performance (*e.g.,* Giving instant individual corrections)	
Assessing Performance	Evaluate learners' understanding and skills (*e.g.,* Conducting final exit quiz)	
Enhancing Retention and Transfer	Use activities that help learners retain information and apply it to new situations (*e.g.,* Applying to real-world cases)	**N/A**

expectations. This approach allows the system to suggest optimal roles tailored to specific pedagogical contexts.

This process aligns with studies on intent elicitation in Human-AI Interaction research, particularly in collaborative image generation [56, 60, 79]. These studies suggest that linear conversational interfaces often struggle to capture abstract or vague intentions during initial interactions [56, 80]. In this context, AI serves as a reflective partner, helping users articulate implicit expectations through structured interaction. Designing such AI systems requires exploring alternative modalities beyond conventional text-based dialogue, such as interactive visual sliders, to crystallize vague intents. Furthermore, the AI must demonstrate its evolving understanding to users by providing functions such as real-time summaries of recognized user expectations. The reflection process also needs to ensure that newly surfaced information is dynamically integrated, allowing the AI to adjust previous assumptions and resolve potential contradictions in the teacher's requirements. Such design considerations can be applied to develop an AI-mediated teacher-intent reflection process for AI role design.

However, before applying such techniques to the workplace, two critical factors must be addressed. First, AI must be capable of mapping teachers' expectations for AI behavior and outputs to specific roles. This requires the system to possess a conceptual mapping of pedagogical expectations and an understanding of which roles are effective for specific goals. Our empirical findings on the three envisioned roles can provide the foundational data for such a map.

Second, establishing an appropriate level of teacher trust in AI-driven role suggestions remains a critical challenge. For instance, teachers may be reluctant to adopt suggestions if the underlying reasoning is opaque or misaligned with their professional judgment [34, 49]. Conversely, if teachers over-rely on AI, they might adopt suggestions without sufficient critical consideration [19]. Therefore, future research should investigate how to present clear reasoning for role suggestions based on the established role-expectation map.

5.2.2 Community-Driven Role Libraries for teachers' AI Role Design.

Teachers desired to learn from the role design experiences of their peers by envisioning shared repositories where they could see which roles were used for similar tasks and assess their effectiveness. Even when teachers had a clear sense of their expectations, they often remained uncertain about whether they had chosen the most appropriate role, such as a scientist versus a curriculum coach or a student, and frequently oscillated between roles that felt too vague or over-specified. This indicates a need for community-driven role libraries where teachers can contribute, discover, and adapt role templates proven effective for specific tasks. Unlike generic prompt libraries, these repositories should be contextualized within educational practice and enriched with peer feedback.

To develop such libraries, we propose two primary design considerations. First, role recommendations must be grounded in teachers' specific tasks and lesson structures. For instance, LessonPlanner [23] uses Gagné's events to scaffold novice teachers through lesson planning, improving both plan quality and workload compared to a generic ChatGPT interface. Our role taxonomy builds on this by proposing specific identities for a given task, such as a subject-matter expert for presenting new content or a simulated student for generating practice examples.

Second, to facilitate the search for contextually relevant starting points, roles should be indexed by subject, grade level, and Gagné's event. Role templates should include explicit "slots" for grade levels, standards, and local curriculum terms that teachers can easily adapt. Furthermore, metadata and warnings should flag potential representational issues, such as Western-centric examples or narrow cultural references, to prompt critical review. These needs for templates align with broader research calling for structured ways to define and share AI personas across contexts while addressing the risks they pose [7, 78]. Design research also warns that AI-generated personas can encode stereotypes or overly positive profiles when created without real user data, highlighting the importance of human oversight [58].

These suggestions extend a long tradition of workers turning to online communities of practice to collectively make sense of new tools. For example, software developers rely on Q&A sites to share heuristics and "warning stories" when adopting AI code generators [12], while public servants and health workers use virtual communities to troubleshoot complex, situated problems [8]. Similarly, teachers have long used professional learning networks to share lesson ideas and co-develop practice [21]. By applying this lens to AI role design, we treat role libraries as practitioner-run infrastructures that transform AI configuration into a shared and ongoing professional practice.

5.2.3 Scaffolding Evaluation Checklist on AI Role Fulfillment in Work.

Teachers in our study faced a difficult challenge in evaluating whether the AI actually fulfilled its assigned role. This was especially difficult when the role required expertise outside the teacher's own domain, such as serving as a museum curator. Our findings support growing calls in the field of HCI to move beyond narrow, model-centric metrics toward human-centered evaluations situated in real-world contexts [14, 37, 59, 71, 72]. These researchers argue that, rather than relying on benchmark scores, we should evaluate AI by how well it fulfills specific human goals and needs. In our setting, this means that evaluating role fulfillment requires distinct criteria for each role. For instance, in educational practitioner roles, teachers can assess alignment with pedagogical standards and the potential for time savings. For roles representing professionals outside education, they can verify the practicality of the domain knowledge and its logistical feasibility. For student roles, the focus shifts to the authenticity of the voice and the prevention of social harm. These role-specific criteria can also serve as design guidelines for developing role-based AI systems tailored to each domain [33].

Beyond these functional criteria, a teacher's judgment is also shaped by how errors feel, not just their frequency. Research on perceived accuracy shows that some mistakes undermine trust more than others, even at similar error rates [50]. Our data shows that teachers view certain failures as serious violations of professional standards, even when the factual content is mostly correct. For example, a student persona giving an unrealistic response or a practitioner role ignoring classroom constraints was experienced as a major breach of role expectations. These qualitative mismatches often damaged trust more than simple factual slips.

To address both the functional and qualitative challenges of role evaluation, we propose integrating role-specific evaluation scaffolds directly into AI tools. After the AI produces an output, the interface could display a brief checklist tailored to that specific role. For a student role, it might ask whether the tone sounds authentic or whether the examples are stereotypical, thereby directly addressing the qualitative mismatches that undermine trust. Teachers could then quickly mark specific issues such as "too advanced", "not culturally relevant", or "logistically infeasible". These annotations would help teachers refine their own work while also providing valuable feedback to community role libraries and system audits.

6 LIMITATIONS AND FUTURE WORKS

First, as an exploratory study, our investigation reflects AI role conceptualizations elicited through a structured, scaffolded task rather than teachers' spontaneous, everyday views of AI. Future work could adopt more naturalistic methods, such as ethnographic or longitudinal studies, to complement these findings. Second, we focused on lesson planning because it encapsulates the complexity of teaching practices, but teachers' AI role expectations may differ in other professional contexts. Future work will extend this investigation to areas such as student advising, assessment, or parent communication. Third, our study involved 15 teachers from a specific national and linguistic background who were already experienced with LLMs. Future studies could examine how teachers across more diverse populations, including educators with less AI experience from diverse national and cultural backgrounds, as well as those at different career stages or AI literacy levels, envision AI roles and articulate their expectations. Fourth, we did not examine iterative dynamics such as changes in teachers' lesson-planning confidence after the AI role design task or how AI response quality shaped their subsequent role reformulations. Future studies can explore this feedback loop to better understand how hands-on AI role design shifts teachers' self-efficacy and refines their design strategies over time.

7 CONCLUSION

We investigated how teachers design AI roles for lesson planning through an interview study. In our findings, we identified teacher-envisioned AI roles in their lesson planning task and revealed their nuanced expectations for each. In addition, we identified the challenges they faced and the support they sought to address them in AI role design. Our work provides an empirical foundation and actionable insights to guide teachers to AI role design process. Ultimately, our research underscores the need to support teachers as they translate their expertise into AI roles, ensuring they realize the full benefits of AI integration in their work.

ACKNOWLEDGMENTS

The authors wish to acknowledge the use of Gemini in the writing of this paper. This tool was used to assist with improving the language in the paper. The paper remains an accurate representation of the authors' underlying work and novel intellectual contributions. Figure 1 used vector art asset by `freepik` and `macrovector`, Figure 2 used assets by `upklyak`, and Figure 3 used assets by `freepik` and `pikisuperstar` from Freepik [24].

REFERENCES

[1] Manal A Almuhanna. 2025. Teachers' perspectives of integrating AI-powered technologies in K-12 education for creating customized learning materials and resources. *Education and Information Technologies* 30, 8 (2025), 10343–10371.

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems.* 1–13.

[3] David P Ausubel. 1960. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of educational psychology* 51, 5 (1960), 267.

[4] David Baidoo-Anu and Leticia Owusu Ansah. 2023. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI* 7, 1 (2023), 52–62.

[5] Arnold B Bakker and Evangelia Demerouti. 2007. The job demands-resources model: State of the art. *Journal of managerial psychology* 22, 3 (2007), 309–328.

[6] Sabrina Belloula. 2025. Empowering Educators: Leveraging AI to Revolutionize Lesson Planning. *International Journal of Research in Education and Science* 11, 2 (2025), 264–280.

[7] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2024. Writer-defined AI personas for on-demand feedback generation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* 1–18.

[8] Oliver Blunk and Michael Prilla. 2017. Supporting communities of practice in public administrations: Factors influencing adoption and readiness. In *Proceedings of the 8th International Conference on Communities and Technologies.* 36–45.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Cindy Candrian and Anne Scherer. 2022. Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior* 134 (2022), 107308.

[11] XinHui Chen, Xiang Yuan, Hui Zhang, Ruixiao Zheng, and Wanyi Wei. 2025. Maintaining" Balanced" Conflict: Proactive Intervention Strategies of AI Voice Agents in Online Collaboration of Temporary Design Teams. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems.* 1–19.

[12] Ruijia Cheng, Ruotong Wang, Thomas Zimmermann, and Denae Ford. 2024. "It would work for me too": How online communities shape software developers' trust in AI-powered code generation tools. *ACM Transactions on Interactive Intelligent Systems* 14, 2 (2024), 1–39.

[13] Soumyadeb Chowdhury, Pawan Budhwar, Prasanta Kumar Dey, Sian Joel-Edgar, and Amelie Abadie. 2022. AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research* 144 (2022), 31–49.

[14] Marios Constantinides, Edyta Bogucka, Daniele Quercia, Susanna Kallio, and Mohammad Tahaei. 2024. RAI guidelines: Method for generating responsible AI guidelines grounded in regulations and usable by (non-) technical roles. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–28.

[15] Marios Constantinides and Daniele Quercia. 2025. AI, Jobs, and the Automation Trap: Where Is HCI?. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work.* 1–8.

[16] Marios Constantinides, Himanshu Verma, Shadan Sadeghian, and Abdallah El Ali. 2025. The Future of Work is Blended, Not Hybrid. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work.* 1–13.

[17] Grant Cooper. 2023. Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of science education and technology* 32, 3 (2023), 444–452.

[18] Roula Daher. 2025. Integrating AI literacy into teacher education: a critical perspective paper. *Discover Artificial Intelligence* 5, 1 (2025), 217.

[19] Sander de Jong, Ville Paananen, Benjamin Tag, and Niels van Berkel. 2025. Cognitive Forcing for Better Decision-Making: Reducing Overreliance on AI Systems Through Partial Explanations. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–30.

[20] Alan R Dennis, Akshat Lakhiwal, and Agrim Sachdeva. 2023. AI agents as team members: Effects on satisfaction, conflict, trustworthiness, and willingness to work with. *Journal of Management Information Systems* 40, 2 (2023), 307–337.

[21] Karen Birgitte Dille and Fredrik Mørk Røkenes. 2021. Teachers' professional development in formal online communities: A scoping review. *Teaching and teacher education* 105 (2021), 103431.

[22] George T Doran. 1981. There's a SMART way to write managements's goals and objectives. *Management review* 70, 11 (1981).

[23] Haoxiang Fan, Guanzheng Chen, Xingbo Wang, and Zhenhui Peng. 2024. Lesson-Planner: Assisting novice teachers to prepare pedagogy-driven lesson plans with large language models. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology.* 1–20.

[24] Freepik. 2025. *Website for Freepik.* https://www.freepik.com/

[25] Robert Mills Gagné. 1985. The conditions of learning and theory of instruction. (1985).

[26] Eleni Georganta and Anna-Sophie Ulfert. 2024. Would you trust an AI team member? Team trust in human–AI teams. *Journal of occupational and organizational psychology* 97, 3 (2024), 1212–1241.

[27] Good Snooze. 2025. Whisper Transcription. https://apps.apple.com/us/app/whisper-transcription/id1668083311. iOS/macOS application, Version 4.0.3. Accessed: 2026-04-21.

[28] Susan Graham-Clay. 2005. Communicating with parents: Strategies for teachers. *School Community Journal* 15, 1 (2005), 117–129.

[29] Muhammad Hamza, Dominik Siemon, Muhammad Azeem Akbar, and Tahsinur Rahman. 2024. Human-ai collaboration in software engineering: Lessons learned from a hands-on workshop. In *Proceedings of the 7th ACM/IEEE International Workshop on Software-intensive Business.* 7–14.

[30] Alexandara Harry and Sayudin Sayudin. 2023. Role of AI in Education. *Interdiciplinary Journal and Hummanity (INJURITY)* 2, 3 (2023), 260–268.

[31] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics* 4, 1 (2000), 53–71.

[32] Janet G Johnson, Macarena Peralta, Mansanjam Kaur, Ruijie Sophia Huang, Sheng Zhao, Ruijia Guan, Shwetha Rajaram, and Michael Nebeling. 2025. Exploring collaborative GenAI agents in synchronous group settings: eliciting team perceptions and design considerations for the future of work. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–33.

[33] Ji-Youn Jung, Devansh Saxena, Minjung Park, Jini Kim, Jodi Forlizzi, Ken Holstein, and John Zimmerman. 2025. Making the right thing: Bridging HCI and responsible AI in early-stage AI concept selection. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference.* 2992–3012.

[34] Shamya Karumbaiah, Yaxuan Yin, and Aayush Bharadwaj. 2025. AIBAT: AI Behavior Analysis Tool for Teacher-Driven Contextual Evaluation of Language Models in Education. In *International Conference on Artificial Intelligence in Education.* Springer, 63–76.

[35] Samantha Keppler, Wichinpong Park Sinchaisri, and Clare Snyder. 2025. Making chatgpt work for me. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–23.

[36] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T Hancock, and Michael S Bernstein. 2020. Conceptual metaphors impact perceptions of human-ai collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020),

1–26.
[37] Aman Khullar, Nikhil Nalin, Abhishek Prasad, Ann John Mampilli, and Neha Kumar. 2025. Nurturing Capabilities: Unpacking the Gap in Human-Centered Evaluations of AI-Based Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.
[38] Taenyun Kim, Maria D Molina, Minjin Rheu, Emily S Zhan, and Wei Peng. 2023. One AI does not fit all: A cluster analysis of the laypeople's perception of AI roles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
[39] Matthew Koehler and Punya Mishra. 2009. What is technological pedagogical content knowledge (TPACK)? *Contemporary issues in technology and teacher education* 9, 1 (2009), 60–70.
[40] Chelsea Lebo and Norma Brown. 2024. Integrating artificial intelligence (AI) simulations into undergraduate nursing education: an evolving AI patient. *Nursing Education Perspectives* 45, 1 (2024), 55–56.
[41] Yang Li. 2024. *Mastering the Art of Lesson Planning: A Practitioner Guide for Teachers (Backed by Scientific Evidence)*. CSMFL Publications.
[42] Sok Ying Liaw, Jian Zhi Tan, Siriwan Lim, Wentao Zhou, John Yap, Rabindra Ratan, Sim Leng Ooi, Shu Jing Wong, Betsy Seah, and Wei Ling Chua. 2023. Artificial intelligence in virtual reality simulation for interprofessional communication training: mixed method study. *Nurse education today* 122 (2023), 105718.
[43] Tao Long, Katy Ilonka Gero, and Lydia B Chilton. 2024. Not just novelty: a longitudinal study on utility and customization of an AI workflow. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 782–803.
[44] Onias Mafa. 2016. Perceptions of teachers towards staff meetings. *Journal of Research & Method in Education* 6, 2 (2016), 61–65.
[45] Hannah Mayer, Lareina Yee, Michael Chui, and Roger Roberts. 2025. Superagency in the Workplace. *no. January* (2025).
[46] Ismaila Mounkoro[1], Anton Diaz Uberas, Flordeline A Cadelina, Muhammad Usman Saleem, Shahana Maryam, Wesam Taher Almagharbeh, Azra Abdul Majeed, and Zakir Hidayatallah. 2024. Integrating AI in Classroom Management: Improving Educational Efficiency and Teacher Workflows. (2024).
[47] Trinh Nguyen and Amany Elbanna. 2025. Understanding Human-AI Augmentation in the Workplace: A Review and a Future Research Agenda. *Information Systems Frontiers* (2025), 1–21.
[48] Regina M Panasuk and Jeffrey Todd. 2005. Effectiveness of lesson planning: Factor analysis. *Journal of Instructional Psychology* 32, 3 (2005), 215.
[49] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–9.
[50] Andrea Papenmeier, Dagmar Kern, Daniel Hienert, Yvonne Kammerer, and Christin Seifert. 2022. How accurate does it feel?–human perception of different types of classification mistakes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.
[51] Charlie Parker, Sam Scott, and Alistair Geddes. 2019. Snowball sampling. *SAGE research methods foundations* (2019).
[52] P David Pearson and Margaret C Gallagher. 1983. The instruction of reading comprehension. *Contemporary educational psychology* 8, 3 (1983), 317–344.
[53] Patti Kelly Ralabate. 2016. Your UDL lesson planner. *The Step-by-Step Guide for Teaching all Learners* (2016).
[54] Sam Ransbotham, David Kiro, Shervin Khodabandeh, Sesh Iyer, and Amartya Das. 2025. The emerging agentic enterprise: How leaders must navigate a new age of AI. *MIT Sloan Management Review (Online)* (2025), 0_1–31.
[55] Peter Reimann and Lina Markauskaite. 2018. Expertise. In *International handbook of the learning sciences*. Routledge, 54–63.
[56] Nathalie Riche, Anna Offenwanger, Frederic Gmeiner, David Brown, Hugo Romat, Michel Pahud, Nicolai Marquardt, Kori Inkpen, and Ken Hinckley. 2025. AI-Instruments: Embodying Prompts as Instruments to Abstract & Reflect Graphical Interface Commands as General-Purpose Tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.
[57] Nancy Safer and Steve Fleischman. 2005. Research matters: How student progress monitoring improves instruction. *Educational leadership* 62, 5 (2005), 81–83.
[58] Vanessa Sattele and Juan Carlos Ortiz. 2024. Generating user personas with AI: Reflecting on its implications for design. (2024).
[59] Devansh Saxena, Ji-Youn Jung, Jodi Forlizzi, Kenneth Holstein, and John Zimmerman. 2025. AI Mismatches: Identifying Potential Algorithmic Harms Before AI Development. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.
[60] Xinyu Shi, Yinghou Wang, Ryan Rossi, and Jian Zhao. 2025. Brickify: Enabling Expressive Design Intent Specification through Direct Manipulation on Design Tokens. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.
[61] Donghoon Shin, Daniel Lee, Gary Hsieh, and Gromit Yeuk-Yin Chan. 2025. PosterMate: Audience-driven Collaborative Persona Agents for Poster Design. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–20.
[62] Lee Shulman. 1987. Knowledge and teaching: Foundations of the new reform. *Harvard educational review* 57, 1 (1987), 1–23.
[63] Dominik Siemon. 2022. Elaborating team roles for artificial intelligence-based teammates in human-AI collaboration. *Group Decision and Negotiation* 31, 5 (2022), 871–912.
[64] James Takona. 2024. AI in education: shaping the future of teaching and learning. *International Journal of Current Educational Studies* 3, 2 (2024), 1–25.
[65] Vijayanth S Tummala, Tashieka S Burris-Melville, and Thomas C Eskridge. 2025. AI as a Team Member: Redefining Collaboration. *Journal of Leadership Studies* 18, 4 (2025), 67–80.
[66] Hannah Ulferts. 2021. *Teaching as a Knowledge Profession: Studying Pedagogical Knowledge across Education Systems*. ERIC.
[67] UNESCO. 2023. UNESCO survey: Less than 10% of schools and universities have formal guidance on AI.
[68] Geesje Van den Berg and Elize Du Plessis. 2023. ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences* 13, 10 (2023), 998.
[69] Maarten W Van Someren, Yvonne F Barnard, Jacobijn AC Sandberg, et al. 1994. The think aloud method: a practical approach to modelling cognitive processes. *London: AcademicPress* 11, 6 (1994).
[70] Bingcheng Wang, Pei-Luen Patrick Rau, and Tianyi Yuan. 2023. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology* 42, 9 (2023), 1324–1337.
[71] Zijie J Wang, Chinmay Kulkarni, Lauren Wilcox, Michael Terry, and Michael Madaio. 2024. Farsight: Fostering responsible ai awareness during ai application prototyping. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–40.
[72] Ziang Xiao, Wesley Hanwen Deng, Michelle S Lam, Motahhare Eslami, Juho Kim, Mina Lee, and Q Vera Liao. 2024. Human-centered evaluation and auditing of language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.
[73] Yaxuan Yin, Shamya Karumbaiah, and Shona Acquaye. 2025. Responsible AI in Education: Understanding Teachers' Priorities and Contextual Challenges. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 2705–2727.
[74] Yaxuan Yin, Shamya Karumbaiah, Ju Lim, Shona Acquaye, and Sara Schwartz. 2025. Toward Teacher-Centered AI Design: Exploring the Role of Pedagogical Values and Contextual Factors in K-12 Teachers' Perceptions of Responsible AI. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 196-204*. International Society of the Learning Sciences.
[75] Robert J Yinger. 1980. A study of teacher planning. *The elementary school journal* 80, 3 (1980), 107–127.
[76] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 841–852.
[77] J Diego Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.
[78] Nima Zargham, Mateusz Dubiel, Smit Desai, Thomas Mildner, and Hanz-Joachim Belz. 2024. Designing AI Personalities: Enhancing Human-Agent Interaction Through Thoughtful Persona Design. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*. 490–494.
[79] Xingchen Zeng, Ziyao Gao, Yilin Ye, and Wei Zeng. 2024. IntentTuner: an interactive framework for integrating human intentions in fine-tuning text-to-image generative models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
[80] Yifei Zhang, Lin-Ping Yuan, Yuheng Zhao, Jielin Feng, and Siming Chen. 2025. KinemaFX: A Kinematic-Driven Interactive System for Particle Effect Exploration and Customization. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

APPENDIX

Task Background

Gagné's nine events of instruction are known as a guiding framework for teachers when planning lessons.
Below, you will find a list of the nine events, definitions for each, and an example of a rough lesson plan.
* The lesson topic in the example is the introduction of the concept of force in a 7th grade physics class.

Event name	Definition of event	Example of a rough lesson plan for each event
Gain attention	Present introductory activities that engage learners	Start with open-ended questions or simple case scenarios to spark curiosity about force
Informing Learners of Objectives	Clearly state the learning goals and outcomes	Provide a student-friendly list of learning goals
Stimulating Recall of Prior Learning	Encourage learners to remember and connect previous knowledge	Ask questions to help students recall concepts from the previous lesson, such as motion or speed
Presenting Stimulus	Introduce new content and information	Provide a clear and age-appropriate definition of force for 7th grade students
Providing Learner Guidance	Offer instructions and strategies to help learners understand and process the new content	Share real-life examples to activate students' prior understanding
Eliciting Performance	Have learners practice what they have learned	Facilitate a discussion session where students connect the concept of force to real-life situations
Providing Feedback	Give constructive feedback on learners' performance	Offer sample student answers and provide feedback using a simple checklist while walking around the classroom
Assessing Performance	Evaluate learners' understanding and skills	Distribute a worksheet or administer a short quiz with relevant questions
Enhancing Retention and Transfer	Use activities that help learners retain information and apply it to new situations	Assign a homework activity to reinforce and extend learning

Figure 4: Introduction to lesson planning via Gagné's nine events of instruction.

Pre-Task introduction

Lesson topic : Introducing the concept of biodiversity and species conservation to 12th grade students for a biology class.

- **Task(1) :** **For each event, please write a rough lesson plan in 'one sentence'.** It doesn't need to be perfect, just a rough plan is enough.
- **Task(2) :** Suppose you are going to turn each rough plan into a more detailed one. You can include what exactly you will say or do, what kinds of examples, questions, and content you will use, how you will make your explanations student-friendly, and how students will participate in each event. **As you develop your rough plans into detailed ones, list one to two purposes for which you plan to use an AI (even if only once or for a simple task) for each event.**

  Example of purposes)

  To create a creative and effective force-related example for 7th grade students, To get ideas about anticipated student answers to force-related questions, To identify difficult words for 7th grade students in written learning objectives

  You don't need to write a purpose for every event. Leave the cell blank for any event where you do not plan to use an AI during this process. But across all nine events, aim to list at least five purposes in total.

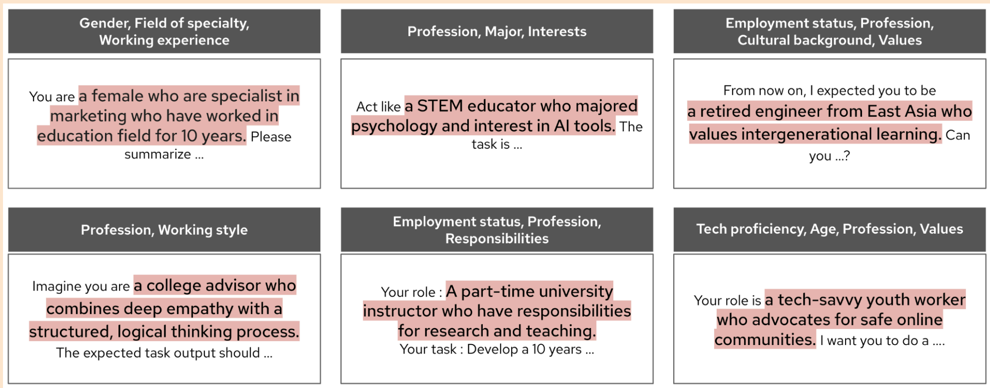
Event name	A rough lesson plan for each event (Task 1)	For what purpose would you use AI? (Task 2)
Gain attention		
Informing Learners of Objectives		
Stimulating Recall of Prior Learning		
Presenting Stimulus		
Providing Learner Guidance		
Eliciting Performance		
Providing Feedback		
Assessing Performance		
Enhancing Retention and Transfer		

Figure 5: Pre-Interview tasks on rough lesson planning and reflection on AI usage purposes.

Main Task

Lesson topic : Introducing the concept of biodiversity and species conservation to 12th grade students for a biology class.

- **Task : <u>Write down the role(s) you will assign to the AI</u>**. You may refer to any sources when designing the role (e.g., another AI, Google, sample prompts, or characteristics of acquaintances). You may also refer to the screenshot of the role prompting introduction slide below.



A rough lesson plan for each event	For which purpose would you assign the roles to an AI?	What role(s) will you assign to the AI for the purpose?

Figure 6: AI role design task.

Main Task

- **What do you expect from the AI's response when it's performing an assigned role?**
  ex) How feasible is the suggested idea?, How understandable and engaging is the content for students?, How creative is the response?, How well does the content fit the teaching context?, How factually correct is the content?

For which purpose would you assign the roles to an AI?	What role(s) will you assign to the AI for the purpose?	What do you expect from the AI's response when it's performing an assigned role?

Prompt for AI	
1	
2	
3	

Figure 7: Expectation elicitation and validation task.