ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Designed, not delegated: Human-centred prompt engineering as a promising approach to AI-supported learning design

Rae Mair
University of Central Florida

Generative AI can improve students' immediate performance while reducing the cognitive and metacognitive work required for durable learning. Although recent work on AI learning companions offers a promising response, these systems remain largely beyond the resources available to individual educators. This paper argues that educators can adopt a more modest form of human-centred AI design by encoding pedagogical content knowledge (PCK) into prompt design and sustaining student partnership throughout iterative development. Drawing on human-centred design, PCK, technological pedagogical content knowledge, and Students as Partners, the paper develops a conceptual account of how pedagogical, content, technical, and contextual knowledge can shape what a large language model (LLM) provides, withholds, and asks students to do. An illustrative case of an AI peer tutor for identifying and evaluating peer-reviewed sources demonstrates how this approach was enacted through five major prompt iterations and progressively deeper forms of student involvement. The case shows how educator knowledge of recurring student difficulties can be translated into pedagogical pathways, while student partnership can contribute contextual knowledge and complementary expertise that further develops the educator's technological pedagogical content knowledge (TPACK). Evaluation indicated substantial improvement in peer-tutoring behaviour and favourable student perceptions, while highlighting the need for continued iteration and research examining effects on student learning.

Keywords: generative AI, human-centred design, pedagogical content knowledge, TPACK, Students as Partners, prompt engineering, theoretical review

Introduction

The entrance of generative AI into education has created widespread concern. Generative AI can enhance immediate task performance while undermining the cognitive and metacognitive processes required for durable learning (Soderstrom & Bjork, 2015), a tension Yan et al. (2025) describe as a paradox in the specific context of generative AI in education. For Khosravi et al. (2026), this paradox leads to the conclusion that "AI should be designed and used to support learning rather than merely improve immediate outputs" (p. 1). They propose a solution of developing human-centred, AI learning companions that are adaptive, pedagogically grounded, and designed for longitudinal, responsible engagement with individual learners over time. This type of engagement contrasts with the single-session, transactional interactions most students currently have with out-of-the-box generative AI tools.

Khosravi et al. (2026) also identify why this is a challenging problem to solve well. Intelligent tutoring systems have long achieved genuine adaptivity, but only by relying on manually authored content, feedback rules, and pedagogical pathways; a time- and resource-consuming endeavour. Large language models (LLMs) offer conversational flexibility and scale that earlier systems lacked, but a significant gap remains in combining that flexibility with the learner modelling and pedagogical adaptivity that made intelligent tutoring systems effective. Their AI learning companions aim to bridge this gap, with early exemplars including systems such as Khanmigo, RiPPLE, and Recast (UTS). Prompt-based approaches, in their account, have so far fallen short of closing this gap in three specific ways: they are reactive rather than proactive, largely stateless with respect to a learner's prior knowledge and misconceptions, and pitched at the level of a single task rather than a learning trajectory across a course.

Realising the vision of AI learning companions, however, requires further development of the current systems, especially in their adaptive capabilities. None of their cases have fully developed the longitudinal capability of

following a student across their program or curriculum, adapting to their particular misconceptions or gaps. It would also require significant institutional infrastructure and capital, dedicated development teams, and platform-level governance. These are resources and contextual environments well beyond the reach of an individual instructor or librarian working within an existing course. For educators charged with supporting students' AI literacy now, this presents an immediate and practical design problem. Waiting for institutional-scale solutions to mature is not a viable response when current students require guidance today.

On closer examination, prompt-based approaches have more potential in fulfilling the goal of supporting learning over task completion than Khosravi et al. (2026) give credit. Two of the three limitations they name are not permanent features of prompt-based design. A prompt built around a teacher's wisdom of practice and pedagogical content knowledge, rather than generic instruction-following, can be proactive rather than reactive, encoding a repertoire of anticipated misconceptions before a student ever encounters the tool. And a design that stays open to revision through ongoing partnership with the students who use it is not stateless in any durable sense. Instead, it accumulates and responds to what those students reveal about their own difficulties, iteration by iteration. The third limitation, task-level rather than trajectory-level adaptivity, this paper does not claim to resolve. What it does argue is that this limitation is not essential to the goal of supporting learning. Trajectory-level adaptivity is a genuine enhancement to that purpose, but it is also the aspect that is most out of reach for an individual educator due to both resources and the need for further technological advancement before it is realised.

This paper argues that an individual educator can practice a genuine form of human-centred AI design by encoding pedagogical content knowledge (PCK) (Shulman, 1986) into prompt design, understood as an element of technological pedagogical content knowledge (TPACK) (Mishra & Koehler, 2006), and by treating that design as an ongoing partnership with students rather than a one-time build. Human-centred design, in this sense, has two dimensions. The first is a substantive one concerned with what a tool is built to do for a learner's own thinking. The second is a process one concerned with involving students in the design of AI tools for the classroom. The argument developed here is conceptual, built from established theoretical frameworks and the author's own experience designing, deploying, and revising the prompt discussed as an illustrative case later in this paper.

Human-centred design for educational AI

ISO 9241-210 describes human-centred design as iterative and identifies four linked activities: understanding and specifying the context of use, specifying user requirements, producing design solutions to meet those requirements, and evaluating the design against them (International Organization for Standardization, 2019, Clauses 5.5, 7.1). Sustained user involvement runs through all four activities. Shneiderman (2022) adds a further dimension to this process-focused definition. A system's design choices, what it does and does not do for the person using it, are themselves an ethical commitment. Applied to education, that commitment includes design that prioritises student learning over outputs. Khosravi et al. (2026) extend this ethical commitment into four additional components for AI systems used in education, including security, protecting learner privacy and data through lawful, responsible practices; transparency, making a system's processes and reasoning visible to learners and educators; accountability, allowing users to question and contest AI-supported decisions; and inclusion, ensuring a system is fair, accessible, and responsive to a diverse range of learners.

The first two aspects, sustained user involvement and designing to support learning, are straightforward for a prompt designer, though they take commitment, time, and skill. A prompt's design can build in student involvement from the outset, including considerations of transparency by disclosing to students what a tool is and how it is being designed, gaining their informed consent to participate, and seeking institutional review board review when applicable. A prompt can also be oriented toward learning over output through engineering and iteration. The desired behaviours of the LLM can be listed as criteria against which a designer tests the prompt for fidelity. In the design of AI Companions, however, creators are working with an entire system with more control over each of the additional components of security, transparency, accountability, and inclusion. When considering these same components in prompt design, there arises a distinction between ethical design and ethical practice.

A prompt designer can address security by conducting any portion of a prompt's design and testing that involves sensitive information within a private or institutionally licensed instance of an LLM rather than a public consumer platform, where data retention and terms of use are governed by an institutional agreement rather than a third party. If the designer is also the educator then they can also decide whether they are comfortable with their assignment being entered into an LLM or keep it within a privately licensed instance. Transparency, in the sense already described, is built into the design and testing process itself, but it can also be included as one of the behaviours designed and tested within the prompt. Similarly, inclusion can be addressed to a degree at the level of a single prompt, adapting explanations to a student's level of language proficiency, for instance, rather than assuming a single fixed register for every learner. These are all aspects that fall within the design phase.

However, since prompts are tools that can be used beyond the intentions or environment of their design, there are some aspects of ethics that fall under practice. If the educator who designs the prompt also teaches with it, ethical considerations would be a part of the whole process; however, in cases where another educator is using the prompt, ethical implementation goes beyond the original designer and becomes the purview of the instructor. In practice, an instructor could model ethical AI use, discuss issues of privacy and copyright, point students towards private or institutionally licensed instances of an LLM or private chat options, and encourage critical reflection on bias, hallucinations, output, and related issues An instructor adopting any learning object is responsible for building the pedagogical and ethical scaffolding around its classroom use. Wikipedia offers a useful parallel. Early concerns about its reliability led many educators to ban it outright, until practice shifted toward teaching students to use it critically, checking its sources rather than avoiding the tool altogether. Generative AI prompts used for learning invite the same considerations of supported, reflective, and scaffolded use.

Encoding pedagogical content knowledge in prompt design

Khosravi et al.'s critique of prompt-based approaches targets a specific and shallow category. They describe such approaches as "prompting strategies that cast the model in roles such as a Socratic tutor... [that] remain limited in scope and provide only surface-level support" (2026, p. 5). This describes a great deal of what passes for pedagogical prompting in practice; a single instruction assigning a persona, a few statements of direction or prohibition (don't give the student the answer), and little further development. A widely cited example from Ali et al. (2023) illustrates the pattern even at its most developed: "You are a teacher who facilitates deep construction of knowledge...When I ask questions, guide me in the form of hints, clues, or suggestions one step at a time. Try not to give me a direct answer" (p. 142). The instruction is thoughtful. It is also brief, a role and a prohibition, without the deeper knowledge of learners and subject matter that shapes an expert teacher's actual practice.

Other work treats prompt design as a discipline rather than a few sentences of instruction. Lee and Palmer's (2025) systematic review of prompt engineering in higher education identifies established frameworks, among them CRISPE and CLEAR, built around structured roles, explicit constraints, and iterative refinement. These frameworks treat prompt construction as a design process with its own conventions and stages, not a single instruction improvised in the moment. Wiley (2026), introducing a collection of pedagogically designed prompts, calls for prompts developed through an iterative process that translates an evidence-based teaching practice into a design intended to cause an LLM to enact the practice reliably and with a high degree of fidelity. That fidelity is evaluated against the standards of the teaching practice iteratively. Khodi et al. (2026) draw a related distinction. Pedagogical prompting, in their account, is not the same as technical prompt engineering. It is the redeployment of a teacher's existing instructional expertise at the moment of addressing an AI tool. These accounts share a foundational concept: PCK (Shulman, 1986), later expanded by Mishra and Koehler (2006) to include technological pedagogical content knowledge in the TPACK model.

Shulman (1986) argues that expert teaching draws on a form of knowledge distinct from either content expertise or generic teaching skill, which is PCK. PCK is the capacity to represent a subject in ways that make it accessible to learners and to anticipate where they will struggle with it. This knowledge has two parts. The first is a repertoire of representations, analogies, examples, demonstrations, and explanations a teacher

accumulates through practice, which Shulman calls the "wisdom of practice" (p. 9). The second is a working knowledge of the misconceptions students are likely to bring to a topic and the strategies that address them, built from watching those misconceptions surface repeatedly. Shulman treats the two as inseparable. Knowing a good explanation is intertwined with knowing which specific confusion that explanation needs to correct.

This is the knowledge base this paper argues can be translated into prompt design. A prompt built around PCK does not merely instruct a model to behave like a teacher or a tutor. It encodes a specific repertoire of representations and a specific accounting of where students are likely to get stuck, the same two components Shulman identifies as constituting PCK. What this looks like in a prompt is not a single instruction but a layered design. The repertoire of representations becomes material written directly into the prompt, the specific examples, analogies, and explanations a teacher would reach for at a given point in a lesson, rather than left to whatever the model generates on its own. The knowledge of student misconceptions becomes a set of anticipated branches, points in the interaction where students are likely to arrive with a particular confusion or a particular gap, each given its own planned response rather than a generic instruction to help. Building a prompt this way is a design task in its own right, closer to lesson planning than to writing a request.

What a prompt provides and what it withholds can be tied to the specific learning objective it is designed to support. Using Bloom's revised taxonomy (Anderson & Krathwohl, 2001) is one way to distinguish the level of thinking that objective requires. An objective pitched at recall or comprehension may call for the model to supply information and quiz students directly. An objective pitched at application, analysis, or evaluation calls for withholding a direct answer instead. A prompt built this way embodies a specific pedagogical approach, such as tutoring, think-pair-share, role-playing, or guided reflection, with anticipated points of student difficulty built into it and clear boundaries set for the model's behaviour. It is then tested against those boundaries to confirm the model behaves as designed.

The repertoire of representations and anticipated misconceptions can be encoded in different ways. Sometimes the repertoire of representations is written directly into the prompt as content, an example, an analogy, an explanation the model reproduces as given. Other times it is written as a direction instead. The prompt might instruct the model to walk a student through a multi-step process, or to explain how to apply a specific evaluative criterion. The model generates its own version of that explanation but within bounds the prompt sets. Misconceptions work the same way. Each anticipated point of confusion gets its own planned response, sometimes fixed content, sometimes a bounded direction. Building a prompt this way is a design task closer to lesson planning than to writing a request, and it reflects the human-centred aspect of supporting learning well beyond simply instructing an LLM to withhold answers or evaluative judgment.

Generative AI and TPACK

Encoding PCK into a prompt is not only an act of teaching but also an inseparable act of using technology. Mishra and Koehler (2006) name the knowledge at this intersection TPACK. TPACK is the capacity to use technology in ways that reflect sound pedagogical judgment and serve the teaching of specific content. It is where content, pedagogy, and technology overlap to shape what a teacher can do with a given tool in a given subject. An educator writing a prompt that encodes PCK is exercising TPACK directly. They are applying pedagogical judgment to a technology while building it to teach something specific.

Mishra et al. (2023) argue that generative AI does not simply add a new tool to TPACK's existing categories. Instead, it changes what those categories have to account for. They describe generative AI as protean, unstable, opaque, generative, and social, qualities that set it apart from earlier classroom technology. A word processor behaves the same way every time it is used. A generative AI tool does not. In a later interview, Mishra connects these new aspects of generative AI to Contextual Knowledge (XK), the component added to TPACK to capture what a teacher needs to know about a specific teaching situation (Mishra & Bozkurt, 2026). He talks about XK in terms of "the sociology of the classroom" and teaching "these particular students in this particular moment," built through the same performative and design artistry that years of classroom experience develop (p. 401). He argues that generative AI itself lacks "the Contextual Knowledge (XK) of the specific students in the room," which he treats as essential to effective instruction (p. 405). Institutional and

policy variation, whether a school bans or encourages AI use, or what a given state's standards require, is one further dimension of XK.

TPACK's account of Technological Knowledge (TK) suggests why this challenge is especially pronounced now. Mishra and Koehler (2006) describe TK as unstable, something that becomes quickly outdated, unlike Content Knowledge and PCK, which accumulate reliably over years of practice. A generative AI tool is new enough that educators have not had the same time to build accumulated knowledge about how a specific population of students relates to it, including both misconceptions about the content and misconceptions about the technology itself. And its rapid change makes it unlikely that such knowledge can accumulate in the same way.

Not everyone agrees that TPACK, as originally formulated, still describes teacher knowledge adequately once generative AI enters the picture. Chiu (2026) argues these are differences in kind, not only in degree, and groups the resulting gaps in TPACK into four areas. AI's probabilistic, agentic behaviour requires operational skills, prompt engineering among them, that TPACK does not name. AI's own capacity to generate content shifts teachers from delivering content to validating it, a competency the framework does not cover. Ethical reasoning, only implicit in TPACK, becomes an explicit requirement. And AI's participation in the classroom introduces a form of human-AI collaboration TPACK has no domain for. On this basis, Chiu (2026) proposes Human-Centric AI Pedagogy (HCAP), five interdependent knowledge domains: three that reframe TPACK's original components through an AI lens, and two entirely new ones, for human-AI collaboration and for ethics. Chiu's own account notes a limitation here. HCAP remains a high-level conceptual structure, without yet specifying the actionable knowledge and skills within each of its five domains for classroom use. Along a similar vein, Shenoy and Saarela (2026) describe TPACK as functioning like "a static competency map" (Section 2.3) and propose a generative AI-enhanced professional development framework in its place, arguing the original model "assumes static, deterministic tools" (Section 5.3.1), in contrast to generative AI's conversational, probabilistic behaviour. Demir et al. (2026) take a related but distinct approach, proposing five new knowledge domains layered on top of TPACK's original three rather than replacing the model outright.

The original introduction of TPACK acknowledges these fundamental technology shifts, stating, "But newer technologies often disrupt the status quo, requiring teachers to reconfigure not just their understanding of technology but of all three components" (Mishra & Koehler, 2006, p. 1030). And Mishra still sees TPACK as valid in the new age of generative AI. In a 2026 dialogue with Bozkurt, he states, "Many people have added circles to the TPACK framework (ethics, equity, you name it)…The problem isn't that these elements aren't important; it's that people don't understand what Venn diagrams are" (Mishra & Bozkurt, 2026, p. 402). His account is that generative AI changes what fills TPACK's existing categories, not the categories themselves. "I don't think the framework fundamentally changes with AI," he explains. "What changes is what goes into those circles, the kind of knowledge required gets redefined" (p. 402). Khodi et al. (2026) reach a compatible conclusion. Writing on pedagogical prompting in English-medium instruction, they describe it as "not a domain of knowledge but a situated act in which such knowledge is put to work," one that "presupposes the TPACK base rather than competing with it or extending it" (p. 3). Belkina et al. (2025) synthesise this extension work empirically. Reviewing 21 case studies, they conclude by proposing a generative AI-TPACK diagram meant to guide educators toward effective, responsible use of generative AI. These authors agree that generative AI does not require TPACK to grow new domains to describe how teachers put their existing knowledge to work with it. This paper operates from the assumption that it is still a valid framework for the use of AI in the classroom.

Student partnership in AI design

Generative AI's generative, social, and unpredictable properties mean the wisdom of practice required in this domain is not readily built once and then applied. Contextual Knowledge adds to this instability, since each student may approach the same assignment from a different contextual standpoint in regard to their experience with, feelings about, and ethical understandings of generative AI. Rapid platform change adds a third source of instability. Knowledge of how a specific population of students currently misunderstands and misuses a specific, rapidly evolving generative AI tool cannot be accumulated in the same way, because the technology has not existed long enough, and because it keeps changing. With so much instability, maintaining

TPACK for generative AI-mediated teaching requires sustained engagement with the students using the tool and continued iteration.

Student partnership is one mechanism for maintaining knowledge of students' perspectives and misconceptions. Cook-Sather et al. (2014, pp. 6–7) define student-faculty partnership as "a collaborative, reciprocal process through which all participants have the opportunity to contribute equally, although not necessarily in the same ways." They name five distinct kinds of contribution including curricular or pedagogical conceptualisation, decision-making, implementation, investigation, and analysis. Applied to prompt engineering as a classroom tool, these categories would include how a tool is conceived, how decisions about it are made, how it is built, how it is tested, and how its results are interpreted.

Illustrative case: AI peer tutoring for source evaluation

The following illustrative case draws on a previously published AI peer-tutor project (Mair et al., 2026), reexamining its development and evaluation through the theoretical framework proposed in this paper. The case concerns a prompt developed to behave as a peer tutor in response to a documented difficulty in a senior-level reading practicum. Teacher education students worked with K–12 students to identify an area for improvement, design an action research project around that student, and use a literature review to inform their intervention plan. Students consistently identified finding, identifying, evaluating, and synthesising sources as the most difficult part of the literature review. Although the specific research topics varied, these difficulties reflected recurring problems seen across students working with research, from undergraduate through doctoral levels. This accumulated experience provided a substantial base of PCK about where students tend to become stuck, the misconceptions that contribute to those difficulties, and the different ways those concepts can be explained. The student-identified need therefore provided a specific context in which that knowledge could be encoded into an AI peer tutor.

The learning objective guiding the design was that students would be able to identify and evaluate a peer-reviewed article for a specific assignment. The author therefore designed the interaction to begin after a student had selected a source and uploaded both the source and assignment directions, giving the LLM the specific context needed to support the task. The author specified what the tutor should provide and what it should withhold. The LLM could provide background information about characteristics of peer-reviewed sources and where those characteristics could be verified, but the student remained responsible for determining whether the source was peer-reviewed, met the assignment requirements, and was relevant to the topic, and for explaining the reasoning behind that decision. The author also instructed the LLM to ask Socratic questions that scaffolded the student's evaluation rather than making the evaluation for the student. Likewise, the LLM could help a student communicate reasoning in academically appropriate language, but it was instructed not to add arguments or justifications that the student had not already supplied. The resulting boundaries between what the tutor provided and what it withheld were therefore deliberate design choices tied to the learning objective.

Drawing on PCK and subsequent testing, the author identified recurring points at which students needed more specific support and engineered corresponding pathways into the prompt. For example, students frequently assumed that an article published in a peer-reviewed journal was necessarily a peer-reviewed article. The author therefore instructed the tutor to distinguish among article types and to guide students through investigating a publication's review process for cases such as editorials, commentaries, conference proceedings, and book chapters rather than simply giving them the answer. Another recurring difficulty involved shallow or incorrect reasoning about credibility. A student might, for example, identify a university affiliation as evidence that the research was credible without explaining why the affiliation established credibility. The author designed the tutor to question and clarify such reasoning rather than replace it with a correct judgment. Other anticipated difficulties included students arriving without a source, being unable to access an uploaded file or link, or being unable to locate basic publication information. Rather than relying on the LLM to determine how to handle these situations, the author incorporated specific guidance for each into the prompt along with an overall outline of the steps of the process to guide the student through.

The author developed the prompt through five major iterations, with design and evaluation occurring as an overlapping cycle rather than a linear sequence. Its intended behaviour as a peer tutor was defined by five pedagogically grounded criteria. These were cognitive congruence, scaffolded sequential dialogue, adaptability, metacognitive transparency, and following assignment directions. Here, adaptability refers to responsiveness to student language, knowledge level, and learning preferences rather than Khosravi et al.'s more longitudinal sense of the term (2026). The author used these criteria both to guide design decisions and to evaluate whether the LLM was enacting the intended peer-tutoring behaviours. Testing therefore examined not simply whether students found the tool useful, but whether the engineered design produced the pedagogical behaviours it was intended to produce from both student and educator perspectives.

Student involvement developed across these iterations. Early student workers tested the prompt, provided feedback on the survey instrument, and suggested improvements. Their contributions primarily supported investigation of the tool and informed decisions made by the author and research team. Classroom implementation extended student involvement into authentic use. Students used the prompt with their own assignments and sources, and their interactions and feedback revealed problems that were difficult to anticipate outside the classroom. Some students, for example, asked the tutor to find a source for them or responded with minimal answers that did not demonstrate their reasoning. These experiences led the author to revise the prompt so that it explained the purpose of the interaction, declined requests to complete the research task in a gentle and supportive way while explaining why it could not do so, and explicitly encouraged students to provide and examine their own reasoning. Within the student partnership framework, these phases represented increasing opportunities for students to contribute through investigation and implementation, while the resulting design decisions remained primarily with the author.

A deeper form of partnership emerged from the first phase of evaluation. One computer engineering undergraduate student, Yuqing Li, demonstrated stronger subject expertise and provided more nuanced feedback on the prompt. The author invited her to be more involved in the process of engineering and testing the final prompt for the project. Unlike the student workers, who primarily provided feedback on an existing design, Li had access to the working prompt and chapter files and actively edited them alongside the other authors. She brought technical expertise that differed from the pedagogical expertise underlying the prompt and contributed to the way the prompt itself was engineered. Her comments led to more explicit and specific instructions governing what the LLM could and could not do, reflecting an understanding of the LLM as a system that required highly constrained instructions for reliable behaviour. She also conducted comparative investigation of the prompt across multiple LLM platforms. Her narrative observations of those interactions became part of the analysis in the larger book chapter and informed recommendations about classroom use. Her contribution therefore extended across investigation, analysis, and decision-making, while also contributing directly to the conceptualisation and development of the prompt (Cook-Sather et al., 2014).

One of the clearest examples of this partnership involved metacognitive reflection. The author had already recognised the pedagogical value of asking students to look back over the interaction and reflect on what they had learned, but had decided against including such a feature because it seemed likely to feel excessive or pedantic to students in a single prompt interaction. During her testing, Li independently added an explicit metacognitive look-back question to the prompt instructions. The author embraced that design decision, and the reflective feature became part of the final prompt. Her interactions as an ESL student with the tutor also demonstrated how the system could support a student in communicating her own ideas without replacing them. Observing this process contributed to further attention to accessibility and to the prompt's ability to support students with different levels of language proficiency. Li's role went beyond evaluation of the prompt. Her participation changed what the final design did.

This experience changed the author's understanding of student partnership. Initially, student involvement had been understood primarily as collecting student feedback and observing their difficulties. Those activities remained valuable, but deeper partnership demonstrated that students could contribute forms of knowledge that observation alone did not provide. Li brought technical expertise and a perspective that differed substantially from those of the pedagogical designer. Her participation introduced possibilities and considerations that would not necessarily have emerged through surveys or classroom observation and, in doing so, changed the author's own understanding of what was required to integrate pedagogy and

technology effectively. Student partnership therefore contributed not only contextual knowledge about students, but also complementary expertise that became part of the author's TPACK as the design developed.

Consistent with the ethical dimensions of human-centred design, the project was conducted within defined ethical and institutional boundaries. Participation in all phases of the evaluation was voluntary, and participating students provided informed consent. Interaction transcripts used for analysis were deidentified and were accessible only to the three authors involved in the analysis. Instructors consented to their assignment directions being entered into the LLM. Students had access to institutionally provided Copilot, including access to ChatGPT through that environment, and the project received an institutional determination that it did not constitute human-subjects research. These practices situated the iterative partnership within the broader institutional context for responsible AI use while preserving the distinction between ethical design decisions and ethical classroom practice.

Discussion

This case illustrates both major dimensions of human-centred AI design developed in this paper. The first is substantive. The prompt was deliberately engineered to support a specific learning objective by encoding PCK about where students struggle, how those difficulties can be addressed, and what cognitive work should remain with the student. The second is process-oriented. The design remained open to revision through sustained involvement of the students who used and tested it. Together, these dimensions distinguish human-centred prompt design from simply assigning a pedagogical role to an LLM. The pedagogical approach was designed by an educator with knowledge of the content, learners, and prompt engineering, then tested and revised through actual student experiences.

The case also suggests that this more modest form of human-centred design may provide a mechanism for an educator's TPACK to continue developing as generative AI changes. A pre-packaged AI companion may be able to provide sophisticated adaptive capabilities in the future, but the educator has limited influence over the pedagogical knowledge built into the system or the process by which that knowledge is developed. In contrast, designing and revising a prompt requires the educator to continually integrate content knowledge, pedagogical knowledge, technological knowledge, and knowledge of the particular students and context. Student partnership adds another dimension by bringing students' perspectives and complementary expertise into that process. In this case, Li's contributions both improved the prompt as an artifact and changed the knowledge available to the author for making pedagogical and technological decisions. Human-centred prompt design can therefore function not only as a way of producing a more contextually responsive tool, but also as a way for educators to develop situated TPACK through the design process itself.

The evaluation evidence provides some support for the value of iterative design. The pedagogical behaviour of the prompt showed substantial improvement across iterations, with the later version demonstrating stronger performance across the peer-tutoring criteria, including cognitive congruence, adaptability, and avoiding giving answers. However, these comparisons require caution because the final technical evaluation used a single technically proficient student, a common assignment, and controlled cross-platform testing, whereas earlier evaluation included more diverse students and authentic classroom conditions. The evaluation therefore demonstrates that substantial improvement was possible through iteration, but it does not establish that the later version would perform equivalently in all classroom contexts. Student perceptions were generally favourable regarding the quality and usefulness of the prompt, while also showing the importance of how students were prepared to use AI as a thinking partner.

The case also has important limitations. The evaluation was designed around the requirements of classroom use and the pedagogical goals established by the author rather than criteria independently verified by outside reviewers. The author therefore necessarily shaped both the criteria used to evaluate peer-tutoring behaviour and the interpretation of the resulting evidence. Student feedback provided an important counterpoint, but the relatively small and context-specific samples limit the extent to which the findings can be generalised. Most importantly, the evaluation examined the fidelity of the LLM's peer-tutoring behaviour and students' perceptions of the interaction rather than whether the interaction produced durable learning. The case

therefore supports the feasibility of human-centred prompt design and iterative engineering, not claims about learning gains.

Continued iteration is also necessary because the technology itself is not stable. The behaviour of generative AI platforms can change as models, interfaces, capabilities, and policies change. A prompt that reliably produces a particular pedagogical behaviour at one point may require revision when the underlying model changes. Human-centred design therefore cannot be treated as a one-time process that results in a finished artifact. Sustained evaluation with the students and contexts for which the prompt is intended provides a means of maintaining the fit among pedagogy, technology, and context as those conditions change.

Future research should examine whether this form of human-centred prompt design produces learning outcomes beyond immediate task support. In particular, research could investigate whether students who use such a pedagogically designed prompt become better able to identify and evaluate sources independently, transfer those skills to subsequent assignments, and engage more deliberately in metacognitive reflection. Research across different disciplines, student populations, assignments, and AI platforms would also help determine how much of the design knowledge developed through partnership transfers across contexts. Such work would extend the present argument from the design and evaluation of pedagogical behaviour toward the larger question at the centre of this paper, which is whether educators can use human-centred design to direct generative AI toward learning rather than merely improved performance.

Acknowledgement

The author acknowledges the use of Claude (Anthropic) and ChatGPT in the preparation of this manuscript. Claude was used to assist with drafting and organizing the author's ideas and to critically examine the author's interpretations against source material provided by the author. The author independently evaluated the resulting text, conducted extensive manual editing and rewriting, and verified the accuracy, contextual appropriateness, and evidentiary basis of the final content. ChatGPT was used iteratively to refine the manuscript's grammar and wording. The author retains sole responsibility for the manuscript and its intellectual contributions, interpretations, and conclusions.

References

Ali, F., Choy, D., Divaharan, S., Tay, H. Y., & Chen, W. (2023). Supporting self-directed learning and self-assessment using TeacherGAIA, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice, 9*(2), 135–147. https://doi.org/10.1080/23735082.2023.2258886

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education. Artificial Intelligence*, *8*, Article 100407. https://doi.org/10.1016/j.caeai.2025.100407

Chiu, T. K. F. (2026). Human-Centric Artificial Intelligence Pedagogy (HCAP) framework developed from TPACK through integration of artificial intelligence literacy and competency. *Interactive Learning Environments*, 1–16. https://doi.org/10.1080/10494820.2026.2615818

Cook-Sather, A., Bovill, C., & Felten, P. (2014). *Engaging students as partners in learning and teaching: A guide for faculty*. Jossey-Bass.

Demir, S., Uşak, M., & Kodirov, O. S. (2026). Three inseparable facets and five new knowledge domains: An extended GenAI-TPACK proposal. *Pedagogical Research*, *11*(3), em0268. https://doi.org/10.29333/pr/18714

International Organization for Standardization. (2019). *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems* (ISO 9241-210:2019).

Khodi, A., Curle, S., & Parra-Guinaldo, V. (2026). Pedagogical prompting rather than technical mastery: Generative AI use by English and English-medium instruction teachers. *Frontiers in Education*, *11*. https://doi.org/10.3389/feduc.2026.1901680

Khosravi, H., Gašević, D., Sadiq, S., Yan, L., Lodge, J., Tangen, J., Denny, P., DiCerbo, K., Buckingham Shum, S., & Baker, R. S. (2026). Building AI companions that prioritise learning over performance. *arXiv*. https://doi.org/10.48550/arXiv.2605.04816

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, *22*(1), Article 7. https://doi.org/10.1186/s41239-025-00503-7

Mair, R., Kelley, M., Wenzel, T., Borowczak, A., Li, Y., & Borowczak, M. (2026). Generative AI peer tutoring to support peer-reviewed source identification and evaluation. In D. Wiley (Ed.), *The pedagogical promptbook: Enacting evidence-based teaching and instructional design practices with generative AI*. EdTech Books. https://doi.org/10.59668/2340.26762

Mishra, P. (2019). Considering contextual knowledge: The TPACK diagram gets an upgrade. *Journal of Digital Learning in Teacher Education*, *35*(2), 76–78. https://doi.org/10.1080/21532974.2019.1588611

Mishra, P., & Bozkurt, A. (2026). TPACK in the age of context and artificial intelligence: Punya Mishra on the "wicked problem" of AI, and why algorithms can't replace context. *Open Praxis*, *18*(2), 398–407. https://doi.org/10.55982/openpraxis.18.2.1093

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and Generative AI. *Journal of Digital Learning in Teacher Education*, *39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Shenoy, P., & Saarela, M. (2026). A critical review and actionable framework for integrating generative AI into teacher professional development. *Discover Education*, *5*(1), Article 570. https://doi.org/10.1007/s44217-026-01579-7

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press. https://doi.org/10.1145/3538882.3542790

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Wiley, D. (2026). The power of educational research, generative AI, and openness: Introduction to The pedagogical promptbook. In D. Wiley (Ed.), *The pedagogical promptbook: Enacting evidence-based teaching and instructional design practices with generative AI*. EdTech Books. https://doi.org/10.59668/2340.26874

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology*, *4*, 435–436. https://doi.org/10.1038/s44159-025-00467-5

Mair, R. (2026). Designed, not delegated: Human-centred prompt engineering as a promising approach to AI-supported learning design. In Cochrane, T., Narayan, V., Poole, T., Lai, P., Lake, P., Moore, O., & Li, L. (Eds.), *The Future is Connection: Collaboration, Creativity and Compassion in a Digital Society*. Proceedings ASCILITE 2026. Brisbane (pp. 304–313). https://doi.org/10.65106/apubs.2026.3249

Note: All published papers are refereed, having undergone a double-blind peer-review process.

The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Mair, R. 2006