Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1633

The New Kid on the Tutoring Block: Definitions, Developments, and Future Directions for Hybrid Human-AI Tutoring

Tiffany Wu
Carnegie Mellon University

Ashish Gurung
Carnegie Mellon University

Kenneth R. Koedinger
Carnegie Mellon University

Danielle R. Thomas
Carnegie Mellon University

Human tutoring is among the most effective educational interventions, yet high costs and intensive resource demands make it difficult to provide broadly and at scale. We conceptualize this tension as an "iron triangle" of tradeoffs between effectiveness, affordability, and scalable reach. This essay examines how an emerging genre of tutoring, hybrid human-AI tutoring, reconfigures these tradeoffs by redesigning how the work of tutoring is divided between people and technology. We first synthesize developments in human- and AI-only tutoring models, with a focus on introducing concepts across the fields of learning analytics and AI in education (AIED) for a broader education research audience. We identify four emerging hybrid architectures, compare how tutoring responsibilities are assigned to humans versus AI and coordinated within each architecture, and examine patterns of human involvement. We conclude by highlighting unresolved questions and outlining a research agenda focused on which combinations of human and AI tutoring are most promising for different learners and settings, and what conditions are needed to implement hybrid tutoring effectively, equitably, and sustainably at scale.

VERSION: November 2026

Suggested citation: Wu, Tiffany, Ashish Gurung, Kenneth R. Koedinger, and Danielle R. Thomas. (2026). The New Kid on the Tutoring Block: Definitions, Developments, and Future Directions for Hybrid Human-AI Tutoring. (EdWorkingPaper: 26-1633). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/bph4-wr91

The New Kid on the Tutoring Block: Definitions, Developments, and Future Directions for Hybrid Human-AI Tutoring

Tiffany Wu
Carnegie Mellon University

Ashish Gurung
Carnegie Mellon University

Kenneth R. Koedinger
Carnegie Mellon University

Danielle R. Thomas
Carnegie Mellon University

Authors' note: This study was supported with funding from the Learning Engineering Virtual Institute. Any opinions, findings, and conclusions expressed in this material are those of the authors. The authors thank Carly Robinson, Sidney D'Mello, and Simon Woodhead for their thoughtful and helpful feedback and suggestions in earlier drafts of this work, and Aizat Nurshatayeva and the EdWorkingPapers Review Board (especially Susanna Loeb, Nathaniel Schwartz, and Hiep Ho) for their careful editorial review and helpful corrections on the working paper. ChatGPT (GPT-6 Astra) and Gemini (3.8 Flash) were used to assist with literature discovery, including identifying recent publications and working papers, as well as proofreading, formatting, and references. All substantive arguments, interpretations, and conclusions were developed independently by the authors. Any AI-generated content was critically reviewed and revised to ensure accuracy and consistency with the manuscript's objectives. Correspondence concerning this article should be addressed to Tiffany Wu via e-mail at wutiffa@umich.edu.

Abstract

Human tutoring is among the most effective educational interventions, yet high costs and intensive resource demands make it difficult to provide broadly and at scale. We conceptualize this tension as an "iron triangle" of tradeoffs between effectiveness, affordability, and scalable reach. This essay examines how an emerging genre of tutoring, *hybrid human-AI tutoring*, reconfigures these tradeoffs by redesigning how the work of tutoring is divided between people and technology. We first synthesize developments in *human*- and *AI*-only tutoring models, with a focus on introducing concepts across the fields of learning analytics and AI in education (AIED) for a broader education research audience. We identify four emerging hybrid architectures, compare how tutoring responsibilities are assigned to humans versus AI and coordinated within each architecture, and examine patterns of human involvement. We conclude by highlighting unresolved questions and outlining a research agenda focused on which combinations of human and AI tutoring are most promising for different learners and settings, and what conditions are needed to implement hybrid tutoring effectively, equitably, and sustainably at scale.

The New Kid on the Tutoring Block: Definitions, Developments, and Future Directions for Hybrid Human-AI Tutoring

Few academic interventions in Pre-K to 12 education carry the promise that human tutoring does. For the past 40 years, efforts to scale tutoring have been shaped by Benjamin Bloom's "two sigma problem." In a 1984 *Educational Researcher* article*,* Bloom reported that students receiving one-to-one tutoring performed approximately two standard deviations (SD, or "sigma") better than their peers in traditional classrooms, even as he concluded that such instruction was "too costly for most societies to bear on a large scale" (Bloom, 1984). Despite critiques about the study's methodology (von Hippel, 2024), the "two sigma" finding continues to be invoked as evidence of tutoring's transformative potential (Kraft et al., 2026).

Contemporary empirical evidence points to more modest, though still substantial, positive effects. Recent meta-analyses estimate average effects for tutoring interventions around 0.29 to 0.40 SD, with declines in effect size as program scales increase (Dietrichson et al., 2017; Kraft et al., 2026; Nickow et al., 2024). These findings place tutoring among the most effective interventions for improving student achievement and helped motivate large-scale investments in tutoring in response to pandemic-era learning loss (Kraft et al., 2026). However, access to tutoring has proven difficult to provide broadly and equitably (Kraft & Falken, 2021). In this sense, the central tension Bloom identified 40 years ago persists: human tutoring can be highly effective but sustaining that effectiveness at affordable cost and broad scale remains difficult.

We conceptualize this tension in Figure 1 as an "iron triangle" organized around three interdependent goals: effectiveness, affordability, and scalable reach. Drawing on the project management concept of an "iron triangle"—often quipped as "fast, cheap, or good: pick two"—we use "iron" in this context to emphasize how improvements in one of the tutoring goals often

come at the expense of another due to resource constraints. For example, increasing effectiveness often requires lower student-to-tutor ratios and increased tutor hours, which raises costs, thereby decreasing affordability and limiting the number of students a program can reach. These tradeoffs rest on a foundation of implementation capacity, such as schools' ability to create protected time for tutoring and support student attendance (National Student Support Accelerator [NSSA], n.d.). The shaded overlaps and bidirectional arrows represent the tradeoffs of the "iron triangle," while the question mark at the center highlights the unresolved challenge of whether emerging tutoring models can better align all three goals.

This essay examines how the rise of artificial intelligence (AI) and hybrid human-AI tutoring models can potentially reconfigure tutoring's "iron triangle." We conceptualize hybrid human-AI tutoring as representing a new division of responsibilities between people and technology to enhance student learning, as *reorganizing the work of tutoring* and shifting the question from whether tutoring works to *how* the work of effective tutoring should be redistributed—which functions AI can make more abundant and inexpensive, which functions continue to depend on human involvement, and how these functions can be combined to create new possibilities.

We first synthesize developments in human- and AI-only tutoring, with a focus on introducing and bridging key concepts across the fields of learning analytics, AI in education (AIED), and education policy to establish a common vocabulary for the discussion that follows. We then introduce four emerging hybrid human-AI tutoring architectures and conceptualize how each architecture shapes effectiveness, affordability, and scalable reach. Lastly, we outline an agenda for future directions organized around which divisions of tutoring responsibilities are most effective, for which students, and under what implementation conditions.

Human Tutoring

To understand the potential of hybrid human-AI tutoring, we first identify the "active ingredients" of effective *human tutoring*, which we define as a subject-matter expert working synchronously with a student or a small group of students either in person or virtually. Decades of research have converged on a set of evidence-based features that distinguish high-impact human tutoring (Robinson & Loeb, 2021), including providing tutoring at a high dosage (two to three sessions per week lasting at least 30 minutes each), delivering synchronous instruction during the school day, maintaining student-to-tutor ratios of no more than 4:1, employing professionally trained tutors, fostering tutor-student relationships, and using high-quality instructional materials aligned with classroom curricula (Walker, 2024). These features appear to operate as a mutually reinforcing bundle of ingredients; recent meta-analytic evidence suggests that programs combining these design features experience less attenuation in their effects as they grow, although no individual feature guarantees success at scale (Kraft et al., 2026).

These features create the conditions for five core tutoring functions that contribute to learning. First, tutoring involves diagnosing students' prior knowledge and adapting instruction, thereby reducing the "academic mismatch" that can arise in whole class settings (Bhatt et al., 2024). Second, feedback and scaffolding during tutoring sessions help students move forward when they become stuck, offering timely hints through the next step in problem solving (Chi et al., 2001). Third, tutoring provides students with additional opportunities for practice and time with the learning content (Nickow et al., 2024). Fourth, the formation of student-tutor relationships can increase student motivation along with creating a sense of belonging (Nickow et al., 2024; Robinson et al., 2026). Fifth, tutors foster accountability, monitoring students' progress and following up when they disengage or need additional support.

As illustrated by Figure 1's "iron triangle," however, sustaining the conditions that enable these functions at scale is resource intensive. Effective human tutoring depends on recurring investments in people. Tutors must be recruited, screened, compensated, trained, and supervised, all of which drive up cost. High dosage and low student-to-tutor ratios compound these costs by requiring many tutor hours for each student served, making it difficult to expand access. In Kraft and Falken's (2021) blueprint for scaling high-impact tutoring nationally, personnel expenses accounted for 83% of their estimated tutoring expenditures, with an average annual cost of approximately $1,022 per student. Saga Education's high-dosage human tutoring model cost about $3,100 per student from 2015-2019 (Guryan et al., 2023). A recent four-year randomized controlled trial of a "short burst tutoring" model cost $416 per student annually but produced smaller effects than those typically reported for intensive tutoring models (Jung et al., 2026), again illustrating the "iron triangle" tradeoff that greater affordability may come with more modest learning gains.

AI Tutoring

From the early days of the computer, researchers have been motivated by the premise that not every function of effective tutoring necessarily requires a human instructor (Koedinger & Corbett, 2006; Smith & Sherwood, 1976; VanLehn, 2011). Early *computer-assisted learning (CAL)* programs, a term which predates modern AI and broadly refers to the use of computers to support classroom instruction, sought to automate structured instructional tasks such as selecting a student's next practice problem.

As these technologies became increasingly data-driven and adaptive, they have evolved into what is now often discussed under the broad label of *AI tutors*, particularly in the learning analytics and AIED literatures (Gurung et al., 2026). AI tutors include a range of technologies

under the broader umbrella of CAL programs that vary in how they use AI to personalize instruction. For our purposes, it is useful to distinguish between three overlapping subcategories, but note that these categories are not mutually exclusive, particularly as platforms increasingly combine features associated with multiple approaches.

Larger commercial AI tutoring platforms such as DreamBox, IXL, and i-Ready can be characterized as *adaptive tutoring systems (ATSs).* They use students' performance data to adjust the sequence, pacing, or difficulty of instructional content (Gurung et al., 2026). *Intelligent tutoring systems (ITSs)* can be understood as a more specific subset of ATSs. They use explicit mathematical models of students' knowledge to guide instructional decisions and provide individualized hints and feedback. For example, ALEKS uses Knowledge Space Theory to estimate what each student knows and is ready to learn next, whereas MATHia uses Bayesian Knowledge Tracing along with other machine learning methods to estimate students' mastery of individual knowledge components over time (Anderson et al., 1995; Doignon & Falmagne, 1999).

A newer class of g*enerative AI tutors,* or *LLM-based tutors,* uses large language models (LLMs) to extend personalization beyond the structured responses traditionally available in CAL platforms and ITSs. Khan Academy's Khanmigo is a prominent example (Oreopoulos & Low, 2026). Originally built using GPT-4 and available via a text-to-text chat window, Khanmigo can ask guiding questions and offer hints while using the context of the Khan Academy content a student is working on to provide scaffolding. Another example is Rising Academies' Rori (Henkel et al., 2024), a WhatsApp-based LLM math tutor. These categories are not mutually exclusive. For example, ALEKS and MATHia are both CAL programs, ATSs, and ITSs, while

Rori combines an LLM-based conversational interface with ATS-like features such as structured curricular sequencing and adaptive problem selection.

Importantly, AI tutors can increasingly perform several of the core tutoring functions through which human tutoring supports learning. Depending on the system, AI tutors can adapt the difficulty or sequence of instruction; deliver feedback and scaffolding through just-in-time hints or conversational guidance; and provide additional practice through individualized problem selection (Kasneci et al., 2023; Koedinger & Corbett, 2006; VanLehn, 2011). LLM-based tutors further extend these capabilities through natural language dialogue, allowing more personalized explanations and responses than were typically available in earlier ITSs. This suggests that AI may be particularly suited to scaling the instructional functions associated with effective tutoring.

Because these instructional functions can be delivered with little additional human involvement once the underlying system has been developed, AI tutors can reduce the recurring personnel expenses that make human tutoring expensive, offering a more affordable and scalable approach to tutoring. For example, Khanmigo's programming cost $15 per student annually (Oreopoulos & Low, 2026), which is much lower than the estimated costs listed above for human tutoring programs. This suggests that AI tutors have substantial potential to relax the affordability and reach constraints of the "iron triangle."

The more difficult question is whether these efficiencies can be achieved while preserving the effectiveness dimension of the triangle. Indeed, the evidence base for AI tutoring is less settled than the evidence base for human tutoring (NSSA, 2026). Meta-analyses of ITSs generally find positive effects, ranging from effect sizes of 0.01-0.09 and not statistically significant (Steenbergen-Hu & Cooper, 2013) to 1.07 and statistically significant (Tlili et al.,

2025), but estimates vary across systems, comparison conditions, and implementation contexts, with effects often smaller on independent or standardized assessments than on measures closely aligned with the tutoring system (Kulik & Fletcher, 2016; Létourneau et al., 2025). The evidence base is thinner for the newest generation of LLM-based tutoring systems. A 2026 review by Stanford University's SCALE Initiative found that only 20 of 800 academic papers on LLM-based tutoring in K-12 education provided strong causal evidence, with no high-quality causal studies conducted in U.S. K-12 settings (Fesler et al., 2026).

A missing piece of the effectiveness puzzle is that tutoring efficacy depends not only on the quality of the instructional support, but also on whether students engage with that support consistently and at sufficient dosage. For instance, a randomized evaluation of Khanmigo found that the median student interacted with it on only one-third of practice days and in 17% of sessions in which they made a mistake (Oreopoulos & Low, 2026). Similarly, Robinson et al. (2026) found that many students made little use of an AI literacy platform when given access alone. This pattern echoes the broader "5% problem" in educational technology: only a small share of students use digital learning tools at the recommended intensity, limiting their potential impact at scale (Holt, 2024). In other words, making AI tutors widely available does not ensure that students will use them sufficiently to benefit.

This challenge is especially salient for the relational functions of tutoring. AI has not yet demonstrated (and may not ever demonstrate) that it can reliably reproduce the motivation, relationship-building, engagement, and accountability that human tutors may provide. Emerging evidence suggests that these functions matter. Robinson et al. (2026) found that human tutors' support of student engagement increased students' use of an AI literacy platform by 71-80%, but usage remained low generally and achievement did not improve. Thus, while AI may be able to

make instructional tutoring functions abundant, current evidence suggests that human attention is needed for functions that depend more heavily on motivation and engagement. This possibility has motivated an emerging class of hybrid human-AI tutoring models.

Hybrid Human-AI Tutoring

The term *hybrid human-AI tutoring* emerged around 2023 in the learning analytics and AIED communities to describe approaches in which humans and AI jointly support student learning (Aleven et al., 2023; Saga Education, 2023; Thomas et al., 2024). It refers specifically to tutoring models with an intentional division of tutoring responsibilities and an explicit workflow between human tutors and AI-enabled technologies. Importantly, hybrid human-AI tutoring is not a single architecture. We use *architecture* to mean the organizing design of a tutoring model, how humans and AI perform different tutoring functions and how their contributions are coordinated.

Below, we identify four emerging hybrid architectures through five illustrative examples, labeled A-E to correspond with Figure 2. Figure 2A summarizes each architecture and situates the examples along a conceptual spectrum of overall human involvement in tutoring delivery, adapted from the National Student Support Accelerator (NSSA, 2026). Figure 2B shows how humans and AI contribute to the five core tutoring functions within each example.

AI-delivered, Human-reviewed Tutoring (Example A). This architecture places AI as the primary deliverer of tutoring while retaining humans as a quality-control layer. An example is Eedi + LearnLM, where the LLM generates conversational chat-based tutoring responses. In the version evaluated in 2025, expert human tutors reviewed every message and, when necessary, edited them before they reached students (LearnLM Team, Google, & Eedi, 2025). The current version of Eedi + LearnLM uses a “human-on-the-loop” approach in which a trained

model evaluates every AI-generated message and determines whether to release it, regenerate the response, or escalate it for human review (Eedi, 2026).

AI-delivered, Human-supported Tutoring (Examples B and C). AI provides the main ongoing instruction, while humans support the AI tutor. The role of human support varies in our two illustrative examples, as seen in Figure 2B. In the AI Literacy Tutor model (example B), the human role is strictly relational; human tutors focus on encouraging engagement and productive platform usage while leaving all instructional responsibilities to the AI literacy platform. In PLUS (example C), students work in a CAL platform or ITS independently while virtual human tutors use learning data to intervene selectively when students need additional explanations, instruction, or motivational support (Thomas et al., 2024).

Alternating Human-AI Tutoring (Example D). This architecture divides tutoring work between humans and AI across time. As seen in Bhatt et al.'s (2024) version of Saga Technology (example D), students alternate on a predetermined schedule between working in a CAL platform or ITS and direct small-group instruction with a human tutor (Bhatt et al., 2024).

Human-delivered, AI-augmented Tutoring (Example E). This architecture reverses the student-facing arrangement. One example is Tutor CoPilot, where the human is the primary tutor while AI supports the tutor rather than interacting directly with the student (Wang et al., 2025). During virtual tutoring sessions, an LLM generates real-time pedagogical suggestions that tutors may incorporate into their instruction.

As seen in Figure 2A, these four architectures offer distinct strategies for navigating tutoring's "iron triangle" by changing where human time and expertise are concentrated. In *human-reviewed tutoring* (A), shifting tutors' work from generating responses to reviewing them allows a given workforce to support more students, improving affordability while retaining

oversight of instructional quality. *Human-supported tutoring* (B-C) extends reach and improves affordability by leveraging AI tutors' personalized instruction capacities while concentrating human attention on motivation, accountability, and in models such as PLUS, targeted instructional support. By reducing the need for continuous human-led instruction, this architecture enables human tutors to support more students while sustaining effectiveness through relationships, engagement, and timely assistance at moments when it may add the greatest value. *Alternating human-AI tutoring* (D) seeks to improve affordability by reducing human instructional time per student while retaining recurring human support. Similar to B-C, this architecture allows human tutors to support more students while sustaining effectiveness through complementary opportunities for personalized practice and direct human interaction. *Human-delivered, AI-augmented tutoring* (E) approaches the "iron triangle" by enabling a broader pool of human tutors to deliver effective tutoring. AI reduces preparation time and the need for intensive tutor training and coaching, thus lowering the cost of supporting tutors and helping programs expand their workforce while maintaining instructional quality.

The NSSA's (2026) one-dimensional spectrum, shown at the top of Figures 2A and 2B, provides a useful starting point for conceptualizing overall human versus AI involvement in hybrid models. Figure 2B adds nuance to this spectrum by showing how similar levels of overall involvement can entail different divisions of tutoring responsibilities. For instance, human support in the AI Literacy Tutor model centers on engagement and accountability whereas PLUS also incorporates targeted instruction. Distinguishing these contributions clarifies not only how much humans participate, but which learning and tutoring needs they address. Focusing on tutoring functions may help districts better match hybrid models to students' needs and resource

constraints while helping researchers evaluate why a given model works and which forms of human support should be preserved as programs scale.

Beyond tutoring functions, conceptualizing hybrid models requires examining how human and AI support connect and reach students. Figure 3 offers two-dimensional perspectives of human and AI involvement. Figure 3A distinguishes student-facing and tutor-facing AI architectures as pathways that can coexist within a model. In the architectures reviewed here, examples A-D have students interacting directly with AI platforms, whereas Tutor CoPilot (E) directs AI assistance to human tutors. The newest versions of PLUS and Saga Technology currently being studied combine student-facing AI (e.g., A-D) with AI-enhanced tutor coaching (e.g., E), connecting AI support for both students and tutors in a multilevel framework (e.g., Moulder et al., 2026; Thomas et al., 2026). Such combinations complicate the placement of newer hybrid models on a one-dimensional spectrum of human involvement but reinforce the value of examining how specific tutoring functions are distributed and coordinated between humans and AI. These four architectures should therefore be viewed as a starting point for understanding a fast-evolving field rather than an exhaustive taxonomy.

Figure 3B adds an additional dimension of distinguishing the relative share of tutoring time involving human interaction from its modality. These are design choices that can vary within a given architecture and may shape different sides of tutoring's "iron triangle." For example, virtual human support may broaden access to tutors, whereas in-person support may facilitate engagement but require local staffing. Thus, districts must consider not only how tutoring responsibilities are divided, but also how the delivery of human support affects learning, cost, and scalable reach.

Future Directions

The emerging evidence across these four architectures points to a promising general direction: using AI to make individualized instructional functions of tutoring more scalable, while deploying human attention more for functions where it may add greater value. However, a challenge for the field is that the implications of different hybrid architectures for tutoring's "iron triangle" remain difficult to compare directly. As Figure 2A illustrates, cost estimates use different denominators (e.g., annually, per session, etc.) and include different components (e.g., software costs versus overall cost per student), while effectiveness is measured using outcomes ranging from topic mastery to knowledge transfer. More consistent reporting of effectiveness and cost metrics, context-specific requirements, and the specific functions performed by humans and AI would support more meaningful comparisons across models and help schools make more informed decisions about which designs best fit their students and implementation contexts.

Comparability, however, is only a first step. The next frontier of hybrid human-AI tutoring research lies in understanding which combinations of human and AI support are most effective, for whom, when, and under what conditions. This question is especially important because the optimal configuration is unlikely to be the same across all students, subjects, and contexts. Some students may thrive with predominantly AI-supported instruction subject to light human oversight, while others may require hybrid architectures that encompass more intensive human instructional and motivational support (Thomas et al., 2024). Advances in AI may also alter what is feasible within each architecture; for example, AI tutors with better representations of students' prior knowledge or stronger sensitivity to disengagement may assume functions that currently benefit from human involvement. Conversely, growing concerns about student screen time, data privacy, and overreliance on technology may increase the value placed on human interaction even as AI capabilities improve (Prothero, 2026).

To this end, Table 1 organizes key future directions and unresolved questions across micro-, meso-, and macro-levels. We view this hybrid agenda as complementary to, and bridging, broader efforts to strengthen both human high-impact tutoring, such as the research agenda advanced by the NSSA (Robinson & Loeb, 2021) and emerging efforts to improve AI tutoring, such as the K-12 AI Infrastructure Program's 2026 EDU AI initiative (K-12 AI Infrastructure Program, 2026).

At the micro-level, key questions concern the interaction among students, human tutors, and AI tutors. One priority is identifying the comparative advantages of humans and AI, or which tutoring functions and pedagogical moves AI can perform effectively, when human involvement adds value, and how human roles should change as AI becomes more capable. A second priority is determining how scarce human attention should be targeted. Human support may be especially valuable at particular moments (e.g., after repeated errors or idleness) and the amount of support students need may vary substantially. Related questions concern the role of engagement and developmental relationships, such as examining which aspects of human relationships sustain productive engagement with AI tutoring, through what mechanisms, and why some students remain disengaged even when human support is available. Of course, human relationships have value beyond encouraging AI platform use. Students' sense of belonging, confidence, trust, and feeling known by a consistent adult also deserve attention as outcomes in their own right, especially as the amount and form of human interaction vary across hybrid tutoring implementations.

At the meso-level, the focus shifts from individual interactions to the performance and implementation of tutoring models across settings. Research is needed to establish how well findings generalize across subjects, grade levels, student populations, hybrid architectures, and

implementations. It is important to understand whether sustained participation in hybrid tutoring produces durable benefits, including students' longer-term capacity to learn independently. Because these systems may rely on detailed data about students' behaviors, schools must also navigate questions of data privacy and access to student information. Equally important are questions of school integration. Hybrid tutoring must fit within existing schedules, staffing structures, and technology systems, and its implementation demands may change substantially as programs move from pilots to routine use. These challenges are not unique to hybrid tutoring. School-based human tutoring programs often fall short of intended dosage because of scheduling conflicts, uneven tutor quality, weak tutor-teacher communication, and difficulties reaching targeted students (Shmoys et al., 2026). AI may alleviate some of these constraints while introducing new ones, particularly around data use.

At the macro-level, the central questions concern equity and sustainability. Hybrid tutoring could expand learning opportunities in underserved schools, but it could also create inequitable tiers of tutoring in which students with fewer resources spend a greater share of their tutoring time with AI and receive less human attention, while more advantaged peers retain the benefits of sustained relationships with human tutors. As human attention becomes more selectively allocated, evaluations thus need to examine who receives human support, whether its allocation addresses students' needs or reinforces existing disparities, and how we can create differentiated hybrid models that provide more intensive human support to students who need it without creating inequitable tiers of tutoring. Scaling hybrid tutoring also requires a fuller accounting of affordability. AI costs, tutor time, infrastructure, and funding sources all shape whether a hybrid architecture remains viable beyond a pilot. Affordability must also account for the resources required to preserve meaningful human support as programs scale. The relevant

question is therefore not simply whether hybrid tutoring can reach more students, but whether it can do so effectively, equitably, and sustainably under real-world resource constraints.

In conclusion, we view the four hybrid architectures, frameworks, and research agenda introduced here as a step toward the development and research of new hybrid human-AI tutoring models. By offering a shared vocabulary for describing how the work of tutoring can be divided between humans and AI, we hope to support clearer communication across research, policy, and practice and to accelerate the development and testing of more effective hybrid models. We also expect this framework to evolve as the field does, and invite the interdisciplinary community of researchers, practitioners, and policymakers to challenge, refine, and extend it.

References

Aleven, V., Baraniuk, R., Brunskill, E., Crossley, S., Demszky, D., Fancsali, S., Gupta, S., Koedinger, K., Piech, C., Ritter, S., Thomas, D. R., Woodhead, S., & Xing, W. (2023). Towards the future of AI-augmented human tutoring in math learning. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial intelligence in education: Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (pp. 26–31). Springer. https://doi.org/10.1007/978-3-031-36336-8_3

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2

Bhatt, M. P., Guryan, J., Khan, S. A., LaForest-Tucker, M., & Mishra, B. (2024). *Can technology facilitate scale? Evidence from a randomized evaluation of high dosage tutoring* (NBER Working Paper No. 32510). National Bureau of Economic Research. https://doi.org/10.3386/w32510

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471–533. https://doi.org/10.1207/S15516709COG2504_1

Dietrichson, J., Bøg, M., Filges, T., & Jørgensen, A.-M. K. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research, 87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Springer. https://doi.org/10.1007/978-3-642-58625-5

Eedi. (2026). *Safety*. https://www.eedi.com/safety

Fesler, L., Martinez, J., Agnew, C., & Loeb, S. (2026). *The evidence base on AI in K–12: A 2026 review*. AI Hub for Education of the SCALE Initiative, Stanford University.

Gurung, A., Gao, G., Gutterman, J., Thomas, D. R., Gupta, S., Branstetter, L. G., Brunskill, E., Aleven, V., & Koedinger, K. R. (2026). *Improving hybrid human-AI tutoring by differentiating human tutor roles based on student needs* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2605.11155

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R. G., Jr., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not too late:

Improving academic outcomes among adolescents. *American Economic Review, 113*(3), 738–765. https://doi.org/10.1257/aer.20210434

Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A. (2024). Effective and scalable math support: Experimental evidence on the impact of an AI-math tutor in Ghana. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Artificial intelligence in education: Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (pp. 373–381). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64315-6_34

Holt, L. (2024). The 5 percent problem: Online mathematics programs may benefit most the kids who need it least. *Education Next, 24*(4), 26–31.

Jung, J., Robinson, C. D., Ripma, T., Cortes, K. E., Kortecamp, K., & Loeb, S. (2026). Early literacy tutoring and children's reading trajectories: Evidence from a four-year randomized controlled trial. EdWorkingPapers. https://doi.org/10.26300/g7sg-m495

K-12 AI Infrastructure Program. (2026, June 1). Open source AI model for tutoring (EDU AI): Request for proposals. https://k12-ai-infrastructure.org/edu-ai-rfp/

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). Cambridge University Press.

Kraft, M. A., & Falken, G. T. (2021). A blueprint for scaling tutoring and mentoring across public schools. *AERA Open, 7*. https://doi.org/10.1177/23328584211042858

Kraft, M. A., Schueler, B. E., & Falken, G. T. (2026). What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability. *Review of Educational Research*. Advance online publication. https://doi.org/10.3102/00346543261446660

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

LearnLM Team, Google, & Eedi. (2025). *AI tutoring can safely and effectively support students: An exploratory RCT in UK classrooms* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2512.23633

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. *npj Science of Learning, 10*, Article 29. https://doi.org/10.1038/s41539-025-00320-7

Moulder, R., Sawaya, S., & D'Mello, S. K. (2026). Testing an AI-enhanced coached-tutor professional learning model for scaling high-dosage tutoring. In *Proceedings of the Thirteenth ACM Conference on Learning @ Scale* (pp. 297–307). Association for Computing Machinery. https://doi.org/10.1145/3774398.3811618

National Student Support Accelerator. (n.d.). *District playbook for high-impact tutoring*. Stanford University.

National Student Support Accelerator. (2026, August 20). *AI tutoring is not a monolith: What we actually know*. Stanford University.

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal, 61*(1), 74–107. https://doi.org/10.3102/00028312231208687

Oreopoulos, P., & Low, N. (2026). *One click away: AI tutoring with Khanmigo in a two-year school experiment* (EdWorkingPaper No. 26-1551). Annenberg Institute at Brown University. https://doi.org/10.26300/kner-hv33

Prothero, A. (2026, August 3). Schools brace for growing backlash to classroom screen time. *Education Week*.

Robinson, C. D., Gormley, D., Ribeiro, A. T., & Loeb, S. (2026). *Access is not enough: Human support improves engagement with AI tutoring* (EdWorkingPaper No. 26-1451). Annenberg Institute at Brown University. https://doi.org/10.26300/pz7p-p388

Robinson, C. D., & Loeb, S. (2021). *High-impact tutoring: State of the research and priorities for future learning* (EdWorkingPaper No. 21-384). Annenberg Institute at Brown University. https://doi.org/10.26300/qf76-rj21

Saga Education. (2023, October 30). *Saga leverages human capital and AI technology to accelerate student learning*.

Shmoys, R., McCormick, S., Bretas, S., Ready, D. D., McCarty, G., & Matulewicz, E. (2026). The dosage dilemma: Implementation evidence on the barriers to high-dosage tutoring. *Educational Researcher*. Advance online publication. https://doi.org/10.3102/0013189X261437761

Smith, S. G., & Sherwood, B. A. (1976). Educational uses of the PLATO computer system. *Science, 192*(4237), 344–352. https://doi.org/10.1126/science.769165

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987. https://doi.org/10.1037/a0032447

Thomas, D. R., Cortes, K., Huffaker, E., Koedinger, K. R., Gupta, S., & Borchers, C. (2026). *It takes a village (plus AI): A randomized controlled trial of hybrid human-AI math tutoring at scale* [Trial registration]. AEA RCT Registry. https://www.socialscienceregistry.org/trials/19556

Thomas, D. R., Lin, J., Gatz, E., Gurung, A., Gupta, S., Norberg, K., Fancsali, S. E., Aleven, V., Branstetter, L. G., Brunskill, E., & Koedinger, K. R. (2024). Improving student learning with hybrid human-AI tutoring: A three-study quasi-experimental investigation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 404–415). Association for Computing Machinery. https://doi.org/10.1145/3636555.3636896

Tlili, A., Saqer, K., Salha, S., & Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis. *Information Development, 41*(3), 825–842. https://doi.org/10.1177/02666669241304407

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

von Hippel, P. T. (2024). Two-sigma tutoring: Separating science fiction from science fact. *Education Next, 24*(2), 22–31.

Walker, T. (2024, February 16). The push to scale up high-impact tutoring. *National Education Association*.

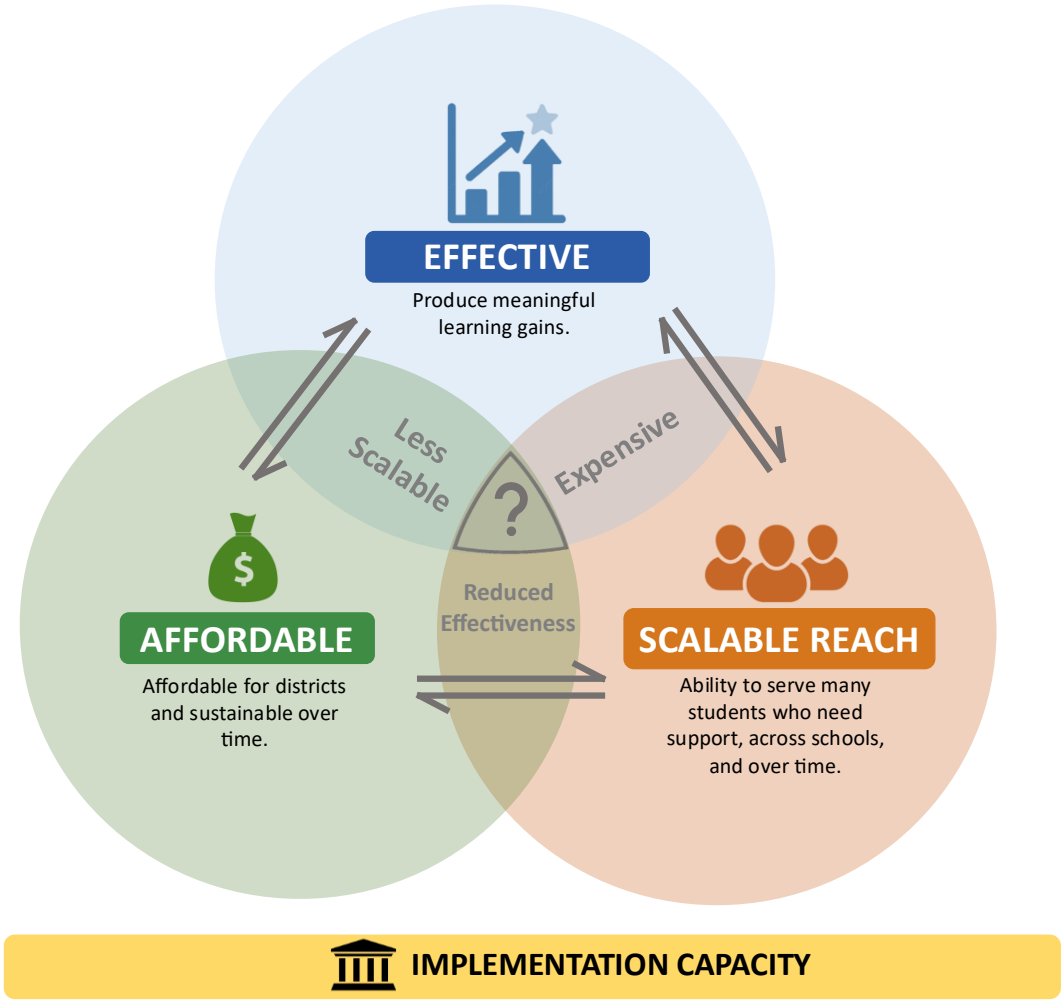
Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2025). *Tutor CoPilot: A human-AI approach for scaling real-time expertise* (EdWorkingPaper No. 24-1056). Annenberg Institute at Brown University. https://doi.org/10.26300/81nh-8262

Tables & Figures

Table 1. Future Directions and Unresolved Questions

	Future Direction	Example Unresolved Questions
Foundational	Making Human-AI Designs Comparable	• Which common effectiveness metrics, cost metrics, and contextual information would allow meaningful comparisons across architectures and help schools make informed decisions about which models best fit their students, resources, and implementation contexts?
Micro-level	Identifying the Comparative Advantage of Humans and AI	• Which tutoring functions and pedagogical moves can AI perform effectively, and when/how does human involvement add value? • How should human roles change as AI capabilities improve?
	Targeting Human Attention: For Whom, When, and How Much?	• At what moments (e.g., after repeated errors, idleness, or disengagement) are human intervention most productive? • Which students thrive with predominantly AI tutoring? Which benefit from AI tutoring paired with periodic human encouragement or accountability? Which require sustained instructional and relational support?
	Supporting Student Engagement and Developmental Relationships	• Which aspects of human relationships are most important for sustaining productive engagement with AI tutoring, and through what mechanisms do they operate? Which students remain disengaged even when human support is available, and why? • How do the frequency, quality, and continuity of human interaction shape students' sense of belonging and academic confidence?
Meso-level	Understanding Generalizability and Long-Term Impact	• How does the effectiveness of hybrid tutoring vary across subjects, grade levels, prior achievement levels, instructional contexts, and student populations? • Which findings generalize across human-AI architectures, schools, and independent implementations, and which are dependent on particular technologies or settings? • How does sustained exposure to hybrid tutoring influence students' longer-term capacity to learn independently, self-regulate, persist through difficulty, and seek help appropriately?
	Integrating Hybrid Tutoring into School Systems	• Which strategies best mitigate hybrid tutoring implementation barriers related to scheduling, tutor training, teacher-tutor coordination, and technology reliability? • How do implementation requirements (e.g., fidelity, dosage, modality), costs, and effectiveness change as programs move from small pilots to routine use across schools and districts? • How should hybrid tutoring models protect student privacy when AI tools rely on detailed data about students' behaviors?
Macro-level	Ensuring Equitable Access to Human Support	• As human attention becomes more selectively allocated, which students receive it, who may be overlooked, and are those patterns equitable? • How can we create differentiated hybrid models that provide more intensive human support to students who need it without creating inequitable tiers of tutoring?
	Building Sustainable Models for Scale	• How do AI costs, human labor, training, coordination, and infrastructure jointly determine an economically viable architecture? • How can recurring funding or shared-cost models sustain effective hybrid tutoring beyond pilots and temporary funding?

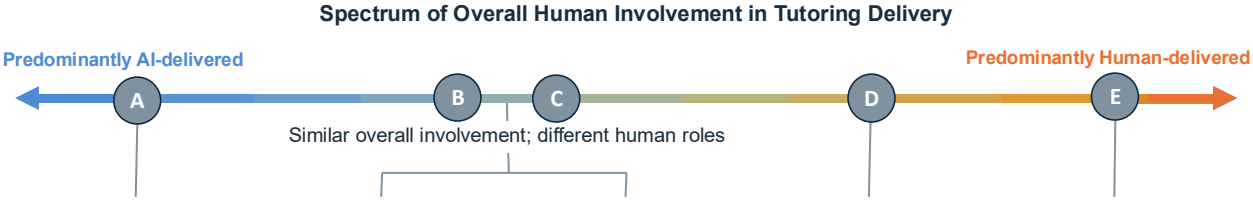
Figure 1. The "Iron Triangle" of High-Impact Tutoring

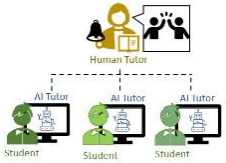
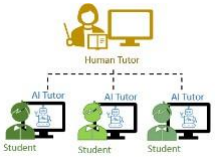
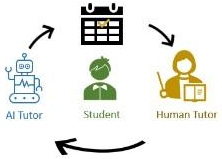


Note. The figure conceptualizes the "iron triangle" of effectiveness, affordability, and scalable reach in high-impact tutoring. We use bidirectional arrows to represent tensions and common tradeoffs between these goals: for instance, increasing effectiveness often requires greater intensity and human capacity, which can increase costs and constrain reach, whereas efforts to improve affordability or expand reach may reduce features associated with tutoring effectiveness. The question mark at the center represents the unresolved challenge of whether emerging hybrid human-AI tutoring models can simultaneously achieve high effectiveness, affordability, and broad reach rather than trading one goal against another. These tradeoffs depend on implementation capacity, including schools' ability to schedule tutoring and support student attendance.

Figure 2. Hybrid Human-AI Tutoring Architectures and Examples

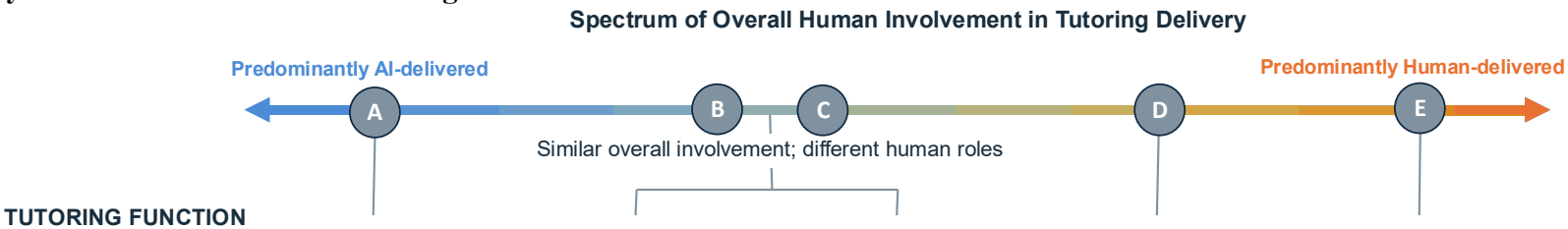
A. Summary of Hybrid Tutoring Architectures



	AI-delivered, Human-reviewed Eedi + LearnLM	AI-delivered, Human-supported AI Literacy Tutor	AI-delivered, Human-supported PLUS	Alternating Human-AI Saga Technology	Human-delivered, AI-augmented Tutor CoPilot
Summary figure					
Summary of human and AI components	LLM generates student-facing tutoring messages; human tutor reviews/edits responses before delivery	AI literacy platform provides instruction and activities; in-person human tutors support motivation and productive use of platform	Adaptive/ITS math platform provides math instruction; virtual human tutors monitor students and intervene selectively for instruction and motivation	Students alternate days between AI-only tutoring using an adaptive/ITS math platform and in-person small-group human tutoring	Human conducts tutoring virtually; AI provides pedagogical suggestions to tutor
Scalability	Increases the number of tutoring interactions one human can oversee while retaining quality control	Targets scarce human attention across more students	Targets scarce human attention across more students	Reduces human tutor-hours per student; same human tutor workforce can serve more students	Scales human tutor expertise/quality
Cost estimates	£0.861 ($1.17 USD) vs. £0.997 ($1.35 USD)/session for human tutoring (operational/ simulation estimates)	Not reported	~$700/student, with a range of $597-$1,170.	~$2,250/student, with a range of $1,900-$2,600.	$20/tutor/year for the LLM API
Effectiveness evidence	Exploratory RCT found +5.5 pp better knowledge transfer to novel problems	Two RCTs found increased platform engagement but no gains in reading achievement	Three QEDs found gains in proficiency and/or usage relative to software-only conditions	RCT found +0.23 SD math gain	RCT found +4 pp for overall topic mastery and +9 pp for students of lower-rated tutors

Note. AI = artificial intelligence; API = application programming interface; LLM = large language model; ITS = intelligent tutoring system; RCT = randomized controlled trial; QED = quasi-experimental design; SD = standard deviation; pp = percentage points. The spectrum of overall human involvement in tutoring delivery is adapted from the National Student Support Accelerator (2026). The ordering of the hybrid tutoring models is conceptual, and the bracket groups PLUS and the AI Literacy Tutor together within the same architecture. Cost estimates are reported as presented in the cited studies.

B. Hybrid Architectures and Tutoring Functions



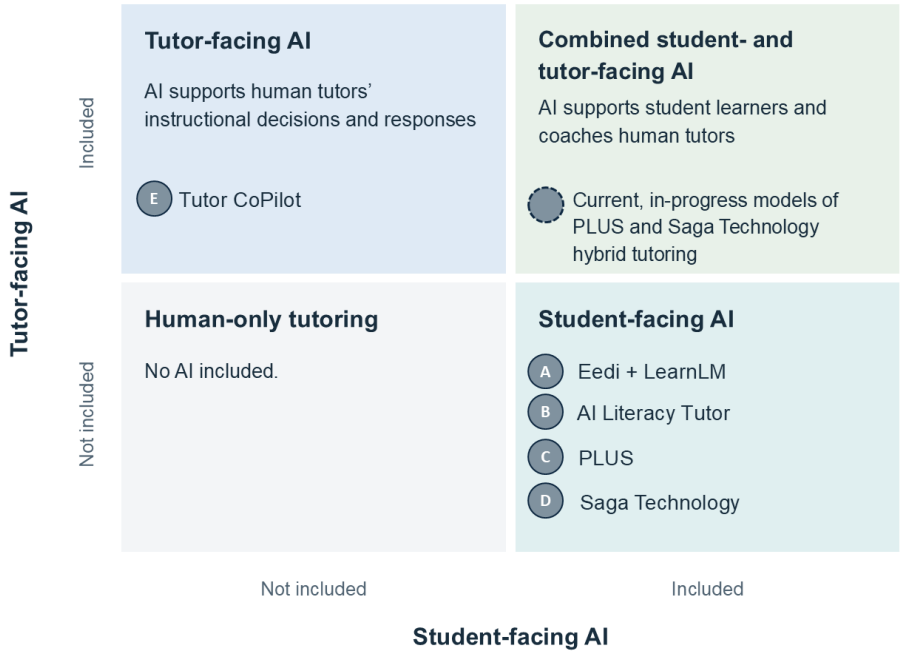
TUTORING FUNCTION	AI-delivered, Human-reviewed Eedi + LearnLM	AI-delivered, Human-supported AI Literacy Tutor	AI-delivered, Human-supported PLUS	Alternating Human-AI Saga Technology	Human-delivered, AI-augmented Tutor CoPilot
Adapting instruction	AI adapts instruction; human tutors review responses	AI adapts instruction	AI adapts instruction; human tutors help diagnose prior knowledge virtually	AI and humans both adapt instruction in their sessions	Human tutor adapts instruction with AI suggestions
Feedback and scaffolding	AI provides feedback; human tutors review	AI provides feedback	AI platform gives feedback; human tutors add explanations virtually if necessary	AI and humans both give feedback and scaffolding in their sessions	Human tutor delivers feedback and scaffolding with AI suggestions
Practice opportunities	AI provides math practice problems	AI provides literacy activities	AI provides math practice problems	Math practice opportunities provided by both AI and human tutors	Human tutor guides student practice virtually with AI suggestions
Relationships and motivation	Human tutors add personal touches and social-emotional nuance in messages	In-person human tutors focus on relationship-building and motivation	Virtual human check-ins encourage relationship-building and motivation	In-person human tutors support relationship-building and motivation	Human tutor maintains student interaction virtually
Accountability and monitoring	Both AI and human tutors monitor student progress	In-person human tutors monitor and support platform use	AI signals progress information; human tutors follow up virtually	In-person human tutors provide accountability and monitoring	Human tutor monitors work and accountability virtually

More of an AI role More of a human role Complementary roles

Note. AI = artificial intelligence. The spectrum of overall human involvement in tutoring delivery is adapted from the National Student Support Accelerator (2026). The ordering of the hybrid tutoring models is conceptual, and the bracket groups PLUS and the AI Literacy Tutor together within the same architecture. Within each tutoring function, colors indicate whether AI or humans primarily perform the function or contribute complementary roles. These assignments represent conceptual descriptions rather than measured proportions of responsibility.

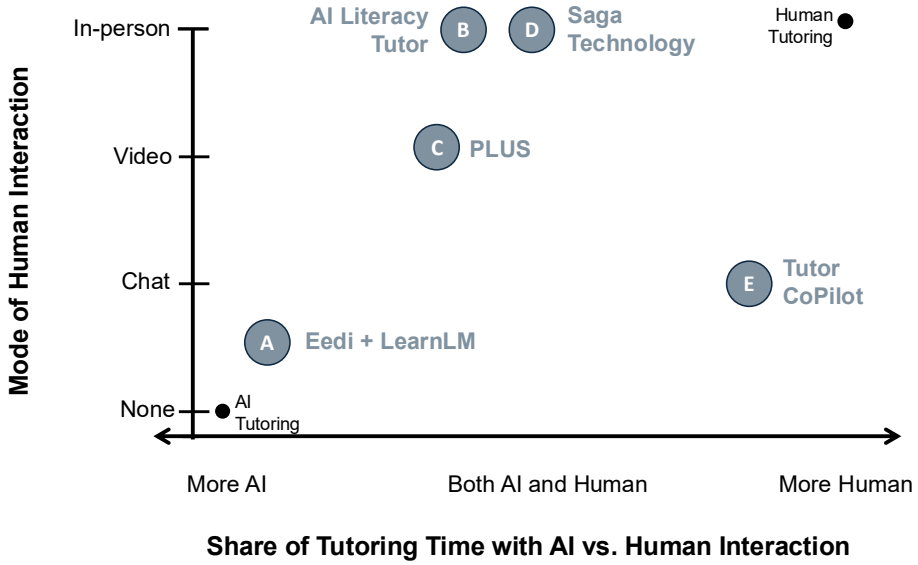
Figure 3. Dimensions of AI and Human Interaction

A. Student-Facing and Tutor-Facing AI



Note. AI = artificial intelligence. The axes distinguish whether a tutoring configuration includes student-facing AI, tutor-facing AI, or both. Student-facing AI provides instructional support to learners; tutor-facing AI supports human tutors' instructional decisions and responses. These pathways can coexist within a single model, as seen in the upper right quadrant of current, in-progress models currently under development, as represented by the dashed circle. Letters correspond with the examples described in Figure 2. Classifications refer to the specific configurations discussed, rather than all versions of a program. Positions within quadrants do not indicate the amount, quality, or effectiveness of AI support.

B. Patterns of Student-facing Human Interaction Across Hybrid Tutoring Architectures



Note. Figure 3B maps the same examples in Figure 2 and Figure 3A along two additional dimensions: the relative share of tutoring time involving AI versus direct human interaction and the modality of that human interaction, from none to text/chat, live audio/video, and face-to-face support. AI-only and human-only tutoring are included as reference points. Positions are conceptual rather than empirical estimates and are intended to illustrate that models with similar overall levels of human involvement (e.g., examples B and C) may differ substantially in what humans do and how students experience human support.