From Answer Engines to Learning Partners: A Dual-ZPD Design Framework for AI-Supported Learning

Reinhard Klein
University of Bonn
Bonn, Germany
rk@cs.uni-bonn.de

Daria Benden
University of Bonn
Bonn, Germany
daria.benden@uni-bonn.de

Alexander Schier
University of Bonn
Bonn, Germany
schier@cs.uni-bonn.de

David Stotko
University of Bonn
Bonn, Germany
dstotko@cs.uni-bonn.de

Fani Lauermann
University of Bonn
Bonn, Germany
flau@uni-bonn.de

Abstract

Generative AI's function as a frictionless "answer engine" creates a paradox in educational HCI: the very tools that can enhance intellect may also weaken it by allowing users to circumvent crucial cognitive processes. This risks creating a "hollowed mind"—knowledge that is broad but superficial, and a user experience that diminishes learner agency. The convenience of cognitive offloading introduces a motivational challenge that traditional cognitive scaffolding cannot address. We argue that designing genuine human-AI partnerships in learning requires moving beyond cognitive support to motivation-aware scaffolding. This paper provides a toolkit for building motivation-aware AI systems. At its core is the Dual Zone of Proximal Development (DZPD), a conceptual framework building on foundational work in educational psychology. We introduce an overarching design principle, concrete design principles, illustrative archetypes, and examples of measurable indicators. These conceptual tools offer essential guidance for the next wave of empirical HCI research in education.

CCS Concepts

• **Human-centered computing** → **Interaction design theory, concepts and paradigms**; • **Applied computing** → **Education**; • **Computing methodologies** → *Artificial intelligence.*

ACM Reference Format:
Reinhard Klein, Daria Benden, Alexander Schier, David Stotko, and Fani Lauermann. 2026. From Answer Engines to Learning Partners: A Dual-ZPD Design Framework for AI-Supported Learning. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3772318.3791476

List of Abbreviations

AI Artificial Intelligence
LLM(s) Large Language Model(s)
ITS Intelligent Tutoring Systems
SDT Self-Determination Theory
ZPD Zone of Proximal Development
ZPM Zone of Proximal Motivation
DZPD Dual Zone of Proximal Development
PLZ Productive Learning Zone
OGR Obligatory Generativity and Responsibility

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791476

1 Introduction: The Motivational Gap in Educational AI

The integration of large language models (LLMs) and generative AI into education has been heralded as a paradigm shift. By externalizing memory, reasoning, and creativity, AI systems promise to act as a collaborative partner and extended mind [18], amplifying cognition and broadening problem-solving capacity [45, 67]. However, this very power creates a profound motivational challenge that is unprecedented in the history of educational technology, posing a central question for Human-Computer Interaction (HCI) in education: *How can we redesign the human-AI partnership to foster deep engagement and restore learner agency, rather than reinforcing convenience-driven shortcuts?*

While the importance of motivation in learning is well-established, previous frameworks were not designed for a world where effortless, synthesized answers to complex problems are perpetually available. Older tools like calculators or search engines offloaded discrete tasks, but left the essential, effortful work of synthesis and explanation to the learner. In contrast, today's generative AI can automate this entire generative process, tempting learners to bypass the very cognitive work—the 'desirable difficulties'—that is essential for building robust knowledge [49, 94]. From a Vygotskian perspective, one might say that the AI system increasingly takes on the role of the "more knowledgeable other". Yet unlike a human teacher who tailors support and challenges to a student's current competence, generative AI tends toward frictionless provision of answers, turning scaffolding into short-circuiting.

This creates what we term the *Convenience Paradox*: the more frictionless the AI, the greater the risk it undermines the learner's own engagement, fostering a '*hollowed mind*' [56]—knowledge that is broad but superficial and that lacks the foundation for long-term intellectual sovereignty. Because this convenience continuously tempts learners to prematurely offload effort, it poses a motivational challenge that traditional frameworks are not equipped to address. At stake is nothing less than the core purpose of educational technology. For decades, HCI has sought to design tools that

scaffold cognitive effort, helping learners build durable knowledge through guided struggle. With the rise of generative AI, this purpose is at risk of inversion: tools so powerful that they may render effort itself obsolete. We argue that this calls for reconceiving AI not as a mere information provider but as a pedagogically-aware partner, demanding a new paradigm of motivation-aware scaffolding that actively cultivates students' willingness to engage in effortful work when an easy alternative is always at hand.

To address this challenge, we propose the Dual Zone of Proximal Development (DZPD), a framework built on two complementary zones. The first is Vygotsky's well-established cognitive Zone of Proximal Development, which we will refer to as the c-ZPD [89]. It describes what a learner can do with cognitive support. The second, building on the seminal work of Brophy [12], is the motivational Zone of Proximal Development, which we will refer to as the m-ZPD. It describes the potential for a learner's motivation and engagement to grow with affective and pedagogical support. For the specific context of AI-mediated learning, we operationalize Brophy's concept of the m-ZPD as the Zone of Proximal Motivation (ZPM). The ZPM therefore serves as the concrete, measurable motivational axis in our framework, representing the learner's moment-to-moment willingness to engage. Crucially, learning can only be sustained if motivation remains within a functional range—high enough to resist the temptation of effortless shortcuts yet resilient enough to withstand challenges.

This idea aligns with insights from Self-Determination Theory (SDT) [24, 57, 79, 80], which emphasizes the centrality of autonomy, competence, and relatedness for sustained motivation. A genuine human-AI partnership for learning requires learners in the intersection of c-ZPD and ZPM, where cognitive challenge and motivational engagement reinforce one another.

AI learning companions have evolved in recent years from static scripted agents to dynamic, adaptive partners [77, 96]. Three archetypical roles stand out: *Teachable Agents*, which leverage the protégé effect by requiring learners to explain and instruct [10, 14]; *Collaborative Companions*, which simulate peer-like interactions that maintain social motivation [30]; and *Creative Co-Creators*, which support divergent thinking and creative ownership [3, 59, 66, 76]. Each of these roles illustrates a shift from a transactional tool to a relational partner, redesigning the interaction to foster sustained learning.

The danger of the Convenience Paradox, however, is particularly visible in everyday educational practice. Learners may copy-paste AI outputs without reflection, rely on LLMs for assignments instead of working through problems themselves, or gradually lose confidence in their own problem-solving competence, leading to diminished self-efficacy [58, 93]. In all of these cases, a breakdown in the learning partnership occurs because the ease of access undermines the very processes of learning engagement [38]. Convenience doesn't just bypass cognition; it actively erodes motivation by short-circuiting behavioral engagement (making effort unnecessary) and cognitive engagement (replacing deep thinking with shallow answers). Over time, this robs the learner of the feelings of mastery and satisfaction that fuel affective engagement, thereby weakening the very foundations of sustained, self-directed learning.

The aim of this paper is to contribute a conceptual framework for the redesign of motivating AI learning ecosystems. We argue that to move beyond the risk of "hollowed minds", such ecosystems must be designed around our proposed DZPD. This DZPD framework is built on the intentional fostering of both cognitive readiness (a learner's ability to tackle a challenge) and motivational readiness. By motivational readiness, we refer to a learner's willingness to productively engage with that challenge on three levels [38]: behaviorally, by investing effort; cognitively, by using deep-level learning strategies; and affectively, by finding value and satisfaction in the process.

This paper makes a foundational theoretical and methodological contribution, providing the concepts, design principles, and evaluation toolkit necessary to guide the next wave of empirical HCI research in motivation-aware educational AI. Specifically, this paper makes the following contributions:

- A novel conceptual framework, the *Dual Zone of Proximal Development (DZPD)*, which defines the Productive Learning Zone (PLZ) and integrates cognitive readiness (c-ZPD) and motivational willingness (ZPM) for designing educational AI.
- A generative design principle, *Obligatory Generativity and Responsibility (OGR)*, as a practical antidote to the "Convenience Paradox" of generative AI.
- A concrete set of five actionable design principles (P0–P4) that translate the DZPD framework into prescriptive guidance for pedagogical instructional design.
- A heuristic toolkit of computable indicators to make the DZPD empirically tractable and to generate testable propositions for future research.

To illustrate the practical power of this framework, we demonstrate how OGR is already embodied in three emerging co-constructive AI roles: Teachable Agents, Collaborative Companions, and Creative Co-Creators. Together, these contributions provide a clear pathway from the risk of hollowed minds toward the cultivation of fortified minds.

2 Related Work

Our framework is situated at the intersection of HCI, learning sciences, and motivational psychology. We first ground our work in HCI perspectives on the Convenience Paradox to frame the problem of the "hollowed mind." We then introduce the core psychological theories and existing AI archetypes that provide the foundation for our solution.

2.1 The Promise and Peril of Cognitive Extension in HCI

The concept of the Extended Mind positions cognition as distributed across brains, bodies, tools, and social ecologies [18]. In this view, digital technologies—and now AI systems—serve as external cognitive resources that expand human problem-solving capacity [67]. In education, this perspective suggests AI can function as a cognitive amplifier, helping learners explore ideas more deeply and creatively than unaided effort would allow. Framed as assistants, current AI systems and LLMs are presented as seamless "cognitive extenders" that promise relief from routine effort so that learners can "focus on higher-order thinking" [44].

Yet generative AI differs qualitatively from earlier cognitive tools. A calculator offloads discrete computations; a book provides information for internalization. The ability of LLMs to produce fully synthesized cognitive outputs—essays, plans, solutions—automates the very generative processes [94] of synthesis and argumentation, which is precisely what allows users to bypass the mental effort required to build a resilient internal architecture for deep reasoning.

These risks of cognitive bypassing are not new; decades of work on digital cognition have foreshadowed them. The "Google effect" shows that when information is externally accessible, people store where to find it rather than what it is [81]; searching can also inflate illusions of internal knowledge [37]. A parallel appears in navigation: GPS reliance correlates with weaker spatial maps and poorer transfer to novel routes [50]. By analogy, LLM reliance can yield task success without schema growth, undermining future transfer.

Now, with generative AI, these well-documented patterns are materializing in high-stakes professional work. Empirically, the pattern is nuanced: human-AI teams often underperform the best individual partner, and gains materialize only when the human has stronger expertise [87]. In field settings, novices benefit on routine tasks but falter at the "jagged frontier" where AI is unreliable; relying on AI there produced a 19-point drop for juniors in consulting, while experts succeeded by rejecting flawed suggestions [26]. Similar asymmetries appear in software development and customer support, where productivity gains concentrate among lower-skill workers on routine work, with minimal advantages for experts [72]. These findings provide a robust foundation for the "Expertise Paradox," a term popularized by observers like Wharton's Ethan Mollick [71] to describe how AI is most beneficial not for novices, but for experts who can critically evaluate and integrate its outputs.

Kahneman's dual-process theory helps explain these asymmetries [54]. Because AI outputs arrive with high fluency and confidence, they encourage fast, intuitive System 1 thinking and acceptance. Overriding that impulse requires effortful System 2 monitoring—checking assumptions, tracing logic, and possibly testing counterfactuals. Novices often lack the schemas and error-detection cues needed to know when to trigger System 2.

Taken together, these findings highlight the Convenience Paradox: the more frictionless AI becomes, the greater the risk that learners will bypass effortful cognitive engagement in favor of shortcutting, producing what we call the "Hollowed Mind", a risk foreshadowed in critical analyses of educational automation [58, 63, 93].

2.2 Foundations for a Motivation-Aware Approach

To address the Convenience Paradox, we must re-examine the foundations of educational support, starting with the now-insufficient model of purely cognitive scaffolding. The concept of the Zone of Proximal Development (ZPD) has been successfully operationalized in Intelligent Tutoring Systems (ITS) to manage cognitive load, guiding learners with adaptive feedback and tailored challenges [88, 98]. Yet, the diagnosis of the Hollowed Mind reveals a critical limitation of this approach in the age of generative AI.

This insufficiency arises because the Convenience Paradox introduces a qualitatively new motivational challenge that prior models were not designed to address. While the ITS community has long recognized the need for motivational support [32], these frameworks operated in contexts where effortful engagement was largely unavoidable; they focused on supporting a learner through a "cannot do" problem. Today's frictionless "answer engines," however, create a dominant "will not do" problem, tempting learners to bypass the very generative processes essential for learning [12, 94]. Recent meta-analyses on AI in education confirm this distinction, showing that AI succeeds when designed as a instructional design/cognitive scaffolding engine but fails when reduced to a universal "answer engine" [21, 99]. This latter model is a pedagogical anti-pattern that bypasses "desirable difficulties" [11] and creates extraneous cognitive load [82]. We therefore argue that this new landscape necessitates a framework built on two complementary zones of proximal development. We must complement Vygotsky's well-established cognitive ZPD (what a learner can do with guidance) with a motivational ZPD (the engagement and self-regulation a learner can sustain with support), an idea pioneered by Brophy [12]. For the context of AI-mediated learning, we term this second zone the Zone of Proximal Motivation (ZPM), a construct specifically designed to regulate motivation against the unique psychological pull of cognitive offloading.

What this distinction makes clear is that motivation acts as the critical gatekeeper. A learner may be cognitively capable of mastering a task but, when offered a frictionless alternative, unwilling to invest effort—the classic "won't do" problem. Without motivational scaffolding, even the most advanced cognitive scaffolding is likely to collapse. Scaffolding has always been conceived as "equal parts motivational and cognitive support" [95], and contemporary research confirms the critical role of motivation, both as a prerequisite to initiate cognitive engagement [2] and as a dynamic state that is continuously manifested through that engagement during the learning process [38]. The ITS community has long recognized this challenge, with Del Solato and Du Boulay [25] and du Boulay [32] proposing motivational indicators such as confidence, independence, and effort as targets for adaptive support.

The central threat of the answer-engine model is its insidious effect on motivation. Its convenience does not just bypass cognition; it replaces deep, resilient motivation with a superficial substitute. To be precise, while the ease of use may temporarily boost affective engagement through the "enjoyment" of effortless task completion, it does so by actively eroding the foundations of learning. It systematically undermines behavioral engagement by making effort unnecessary, and cognitive engagement by replacing deep thinking with shallow answers. This trade-off is profoundly damaging: removing the "desirable difficulties" strips learning of the very experiences that build lasting interest and genuine self-efficacy [49, 82]. Second, learners display present bias—preferring immediate answers over the delayed rewards of understanding [61]. Third, overreliance on AI is linked to declines in critical evaluation and active engagement, reinforcing passivity rather than generative work [99]. Together, these dynamics depress the motivational drivers needed to sustain effortful learning and weaken the conditions required to keep learners in a productive zone.

The upshot is clear: in educational contexts—where the goal is knowledge construction, not mere task completion—the prevailing "answer-engine" interaction pattern systematically pulls learners

toward convenience and away from the effort that builds transferable schemas. This justifies a design shift from answer delivery to scaffolded, generative engagement that makes the effortful path the easy choice—preparing the ground for the DZPD framework and the AI scaffolding patterns developed in the next sections.

2.3 HCI Precedents: Co-Constructive AI Archetypes

The integration of LLMs and generative AI between 2020 and 2025 has fundamentally transformed three central domains of educational technology: Teachable Agents, Collaborative Companions, and Creative Co-Creators systems. This transformation marks a paradigm shift from rule-based, rigid interaction systems to flexible, multimodal partners capable of teaching, learning, and creating alongside human learners. Increasingly, the boundaries between these categories are dissolving: modern AI systems can fluidly transition between the roles of teacher, student, and creative collaborator, depending on context and learner needs [87].

2.3.1 Teachable Agents: Rethinking Learning by Teaching. Teachable agents—AI systems that learn through being instructed by students—have undergone perhaps the most dramatic transformation in the LLM era. Early systems such as SimStudent required thousands of lines of Java code and operated in narrow domains. In contrast, modern LLM-based Teachable Agents deliver comparable pedagogical benefits with far lower development costs and vastly broader subject coverage [51]. Controlled studies demonstrate strong learning gains: for example, Jin et al. [51] reported a 71% increase in knowledge-building conversation density when students used the TeachYou/AlgoBo system compared to traditional methods. Other studies show that LLM-based Teachable Agents support learning across diverse domains including music theory [52] and coding [51, 69] with generalization capabilities to various other fields [17].

The pedagogical power of Teachable Agents builds on the well-established learning-by-teaching framework and the protégé effect, which show that students invest more effort and achieve deeper understanding when responsible for teaching others [14]. Modern LLM-based agents enhance this dynamic by simulating realistic learning behaviors, including authentic misconceptions and knowledge gaps, which require students to explain and adapt their teaching. At the same time, the integration of LLMs introduces challenges, as their extensive pre-trained knowledge risks discouraging genuine student teaching unless carefully constrained through prompting architectures and knowledge scaffolding [51].

2.3.2 Collaborative Companions: From Tools to Partners. AI learning companions have evolved from basic question-answering systems to sophisticated peer partners supporting both academic and emotional dimensions of learning. Whereas Teachable Agents position the AI as a novice protégé that the student is responsible for teaching, collaborative companions typically frame the AI as a peer who shares the learner's task rather than being its object. Where pre-2020 systems offered limited, scripted responses, modern companions engage in natural, contextually aware dialogues that adapt to individual learning styles, cognitive states, and emotional needs [47]. Recent studies illustrate diverse interaction patterns: active questioners, responsive navigators, and silent listeners—all yielding comparable learning outcomes through different modes of engagement (see, e.g., [43]). Other studies already report successful applications of such AI companions in STEM courses like physics [34, 62] and medicine [5]. Multi-agent approaches such as the MAIC (Massive AI-empowered Course) system demonstrate the potential of specialized agent teams (teachers, assistants, creative sparkers, note takers, etc.) to provide richer support than any single generalist AI. Beyond cognitive benefits, companions enhance metacognitive awareness, self-regulation, and motivation, with affective support reducing anxiety and fostering persistence. These developments align with social constructivist theories, positioning companions not as tools but as collaborative partners in learning communities.

2.3.3 Creative Co-Creators: Toward Creative Partnerships. Co-creative AI systems represent the newest frontier, made possible by advances in generative AI. Crucially, this distinguishes Co-Creators from the Collaborative Companions described above: while companions primarily leverage social presence to scaffold self-regulation and persistence, Co-Creators focus specifically on scaffolding divergent thinking and the joint production of novel artifacts. In this dynamic, the AI is not just a peer offering support, but an active collaborator in the generative process itself. A systematic review by Urmeneta and Romero [86] identifies three roles of AI in creative education: facilitator (supporting ideation), co-creator (active collaborator), and autonomous generator (independent producer). Empirical evidence highlights robust benefits for creative learning. For example, students using generative AI tools for digital storytelling showed significant improvements across all creativity dimensions—novelty, relevance, and collaboration—with a large effect size [86]. Applications now span writing, visual arts, music, and digital media, democratizing creative expression and enabling multimodal production at scale. Yet co-creative AI raises pedagogical concerns, particularly the risk of cognitive offloading and diminished critical evaluation. Moreover, evidence from programming education indicates that expert users derive greater benefits from AI support than novices [15]. Successful implementations balance efficiency gains with deliberate reflection, ensuring that students remain engaged in essential processes of meaning-making, evaluation, and creative judgment.

2.3.4 Traditional Instructor-Oriented Tutoring Systems. While our framework focuses on learner-centered interaction patterns in which the student is positioned as more knowledgeable than or at least epistemically aligned with the AI system, it is important to acknowledge the long-standing tradition of teacher-like intelligent tutoring systems (ITS), including recent LLM-based tutors [60]. Classic ITS and hypermedia tutors such as MetaTutor [4, 28] and AutoTutor [40, 41, 73] as well as newer LLM-driven agents [9, 17, 75] adopt a fundamentally different pedagogical stance: the system occupies the role of an expert instructor, delivering explanations, diagnosing misconceptions, and regulating learning for the student. Although these systems provide valuable empirical grounding for the broader landscape of AI-supported learning, their hierarchical teacher–student model contrasts with the relational configurations central to our three archetypes—teachable agents, collaborative companions, and creative co-creators—where learners

guide, partner with, or co-create alongside the AI. By differentiating our focus from these traditional teacher-like systems, we clarify how the unified design space extends existing theoretical foundations (ZPD, motivational ZPD, and SDT) toward interaction paradigms that aim to empower and elevate the learner's epistemic agency.

2.3.5 Convergence and Integrated Learning Ecosystems. Perhaps the most significant development is the convergence of Teachable Agents, Collaborative Companions, and Creative Co-Creators into integrated ecosystems. Emerging platforms allow AI to shift roles dynamically, acting as tutor, peer, or creative collaborator depending on the learner's needs [20]. Recent simulations such as *SimClass* showcase multi-agent LLM-powered classrooms where AI simultaneously assumes the roles of teacher, student, and collaborator. Advances in memory management, multimodal interfaces, and external tool integration further enable persistent, evolving relationships with learners across contexts.

This convergence suggests a fundamental evolution in educational technology: the move from static, tool-based systems toward fluid, adaptive partnerships where AI becomes an active participant in the social and creative fabric of learning. The challenge is thus not merely technological but psychological: how can we design AI systems and learning environments that encourage effortful engagement in contexts where a frictionless shortcut is always available? We argue that the answer lies in designing for cognitive sovereignty, ensuring that learners develop the internal capacity to critically engage with AI outputs. The following section introduces the Dual Zone of Proximal Development (DZPD) as a framework for balancing cognitive readiness and motivational willingness in this endeavor.

3 The Productive Learning Zone: A Dual-ZPD Framework for Motivation-Aware Design

3.1 Defining the Dual-ZPD Framework and the Zone of Proximal Motivation (ZPM)

As established in our introduction, the theoretical foundation for the motivational dimension of our framework is Jere Brophy's [12] concept of a motivational Zone of Proximal Development (m-ZPD). He defined this as the range of activities "familiar enough to the learner to be recognizable as a learning opportunity and attractive enough to interest the learner in pursuing it" but not so familiar as to be boring or so alien as to be unapproachable (p. 77). Building directly on his work, we operationalize this concept for the unique challenges of AI-mediated learning under the term Zone of Proximal Motivation (ZPM). As Yung and Tao [97] emphasize, learners can be advanced within this motivational zone through modeling, coaching, and scaffolding, thereby fostering not only cognitive competence but also appreciation and enjoyment of the learning activity. At its core, this motivational zone—which we term the ZPM—reflects the gap between a learner's current valuing, interest, and confidence and their potential to come to value and feel confident about a domain with appropriate scaffolding. Teaching "within" the ZPM involves calibrating the appeal and familiarity of activities (not too foreign, not over-familiar) and using affective supports, such as AI-delivered enthusiasm, appreciation-oriented feedback, and an interaction climate that nurtures interest and self-efficacy. Thus, where the cognitive ZPD is about stretching what students can do with thinking support, the ZPM is about stretching what they want and feel able to pursue, using social-emotional support so that value and confidence "catch up" to the learning opportunities. As Brophy argued, effective instruction must weave both.

Glenn Regehr's scholarship in medical education further clarifies how motivational development can itself be scaffolded through what he and colleagues term the Educational Alliance [83, 84] and the Supported Independence model [48]. This work describes motivation not as a static trait but as a scaffolded process: learners first recognize uncertainty in their knowledge or skills (self-monitoring), then engage in comfort borrowing by relying on psychologically safe supervisory relationships, progress to strategic help-seeking, and finally enter guided problem-solving where support is gradually tapered until independence is achieved. This developmental sequence directly mirrors the progression of Vygotsky's ZPD: just as cognitive scaffolding is withdrawn once it is no longer needed, motivational scaffolding is progressively faded as learners internalize both competence and confidence to act autonomously. In this sense, these models provide an empirical foundation for the ZPM, highlighting that willingness to engage effortfully can be cultivated through carefully designed relational scaffolds—that is, supports grounded in the social-emotional dynamics of the learning partnership, such as trust, credibility, and psychological safety. By situating motivation within processes of supervision, credibility, and trust, this framework underscores that the Productive Learning Zone (PLZ) is not only a cognitive "sweet spot", but also a relational and motivational one, where both ability and willingness to engage are actively constructed.

Drawing on these traditions, we define the PLZ as the intersection of cognitive reachability (the c-ZPD) and motivational readiness (the ZPM). The distinction between what a learner can do (their state relative to the c-ZPD) and their willingness to productively engage (their state relative to the ZPM) is critical, as it highlights the limitations of treating learning as a purely cognitive problem. As conceptualized in Table 1, their overlap in Quadrant 1 defines the PLZ: the true "sweet spot" where learners are both cognitively capable and prepared to engage behaviorally, cognitively, and affectively. At the same time, Table 1 illustrates that the appropriate scaffolding action depends on the learner's specific state. While intervention is always holistic, its primary focus must adapt to the most immediate barrier. This barrier might be purely cognitive (a knowledge gap), or it might be a breakdown in engagement—for instance, a reluctance to invest behavioral effort or apply deep cognitive strategies. Operating within the PLZ is therefore the core design principle for effective educational AI.

3.2 The AI-Specific ZPM: A Control Mechanism for the Convenience Paradox

While the concept of a motivational ZPD has been previously explored in teacher-student contexts (e.g., Brophy [12], Ilgen et al. [48]), this paper is the first to formalize and operationalize it as the ZPM specifically for the unique design tensions of AI-mediated learning, where the risk of frictionless convenience presents an

	Task is in the c-ZPD (Learner can do it with help)	**Task is NOT in the c-ZPD (Learner cannot do it, even with help)**
Task is in the ZPM (Learner is motivated)	**1. The Productive Learning Zone (PLZ):** The learner is both willing and able. This is the sweet spot for growth. *Action: Provide integrated scaffolding with a primary focus on the cognitive task (e.g., hints, feedback, worked examples).*	**3. The Frustration Zone:** The learner is willing but unable. They try but fail, leading to frustration and loss of confidence. *Action: Simplify the task to bring it into the c-ZPD.*
Task is NOT in the ZPM (Learner is unmotivated)	**2. The Motivational Barrier:** The learner is able but unwilling. They have the potential to learn but refuse to engage. *Action: Provide scaffolding with a primary focus on motivation (e.g., reframing the task, goal-setting, restoring agency)*	**4. The Dead Zone:** The learner is both unwilling and unable. The task is irrelevant and too difficult. No learning will occur. *Action: Abandon or completely redesign the task.*

Table 1: The Dual Zone of Proximal Development (DZPD) framework, which defines the Productive Learning Zone (PLZ) as the target state in Quadrant 1. The core design principle for effective educational AI is to guide and maintain the learner within the PLZ.

unprecedented challenge. Unlike relational models framed around human supervision, the AI-specific ZPM is defined by a critical design tension. On one side lies excessive convenience: if AI provides answers too quickly or completely, learners may bypass effort, short-circuiting both cognitive and motivational growth. On the other side lies excessive challenge: if AI support is too sparse or opaque, learners may disengage in frustration.

ZPM defines the narrow corridor where AI assistance must be calibrated: support should lower barriers enough to invite engagement, while leaving sufficient productive struggle to sustain motivation and deepen understanding. This transforms the ZPM from a passive diagnostic concept into an active design control surface for AI systems. Its role is to ensure that learners remain inside the PLZ, thereby preserving their cognitive sovereignty—the capacity to extend one's mind with AI while retaining essential skills of critical thinking and self-regulation. For designers, this specifies the levers through which this balance can be enacted: calibrating challenge levels, embedding motivational scaffolds, and providing explainable feedback that attributes progress to effort and strategy rather than opaque system outputs.

3.3 Consistency with Established Motivational Theories

In situated expectancy-value theory (SEVT), a learner's choices—such as the choice to invest effort or persist—depend on their expectancy and task value beliefs (which include perceived costs) [33]. From this perspective, the PLZ represents the region where this motivational calculus is most favorable: the zone where a learner's expectancy for success and the perceived value of the task are high enough to outweigh the cost of effort, leading to the choice to sustain engagement. In SDT, intrinsic motivation and self-regulation are maximized when autonomy, competence, and relatedness needs are supported [57]; the PLZ integrates these needs as preconditions for remaining "inside" the zone. In social cognitive theory (SCT), self-efficacy and self-regulation are central [8]; the PLZ is intended to capture the balance where efficacy beliefs and regulatory resources align with task demands. In achievement goal theory (AGT), mastery-oriented goals and contextual cues guide motivation [35]; the PLZ situates mastery orientation within a broader framework that accounts for both cognitive readiness and motivational willingness. Finally, in attribution theory, persistence depends on causal beliefs about success and failure [92]; the PLZ is designed to underscore the importance of attributions that sustain willingness to engage even when challenges occur.

In our use, SEVT specifies the entry conditions for the ZPM: engagement becomes a viable option when learners' expectancy for success and perceived task value jointly outweigh perceived costs [12, 33]. This is precisely the region where the Convenience Paradox can derail learners toward low-effort shortcuts, even though the motivational calculus would otherwise support deeper engagement. SDT then specifies the maintenance conditions for staying in the ZPM by emphasizing the ongoing satisfaction of autonomy, competence, and relatedness needs [57]. Together, SEVT and SDT thus define both when the ZPM "opens" and what it takes to remain inside it, motivating the next step where we operationalize these needs and value–expectancy signals as real-time levers in DZPD-aware AI.

While all these theories offer powerful lenses for understanding the PLZ, our framework relies most directly on the combination of SEVT and SDT. Following expectancy–value accounts of motivation [12, 33], SEVT guides how a DZPD-aware AI infers whether a learner is likely to enter the PLZ/ZPM at all by estimating whether perceived value and expectancy are high enough, relative to cost, for deep engagement to be chosen over low-effort alternatives. Both self-report data and behavioral traces can be used for this purpose. SEVT and SDT together give us the main tools for putting the framework into practice in interaction design. SDT's three basic psychological needs—autonomy, competence, and relatedness—and SEVT-informed, value-enhancing messages (e.g., "Why do I need to do this?") function as direct, tunable levers for real-time adaptation (see Figure 1). This is a distinctly AI-centric approach: whereas a

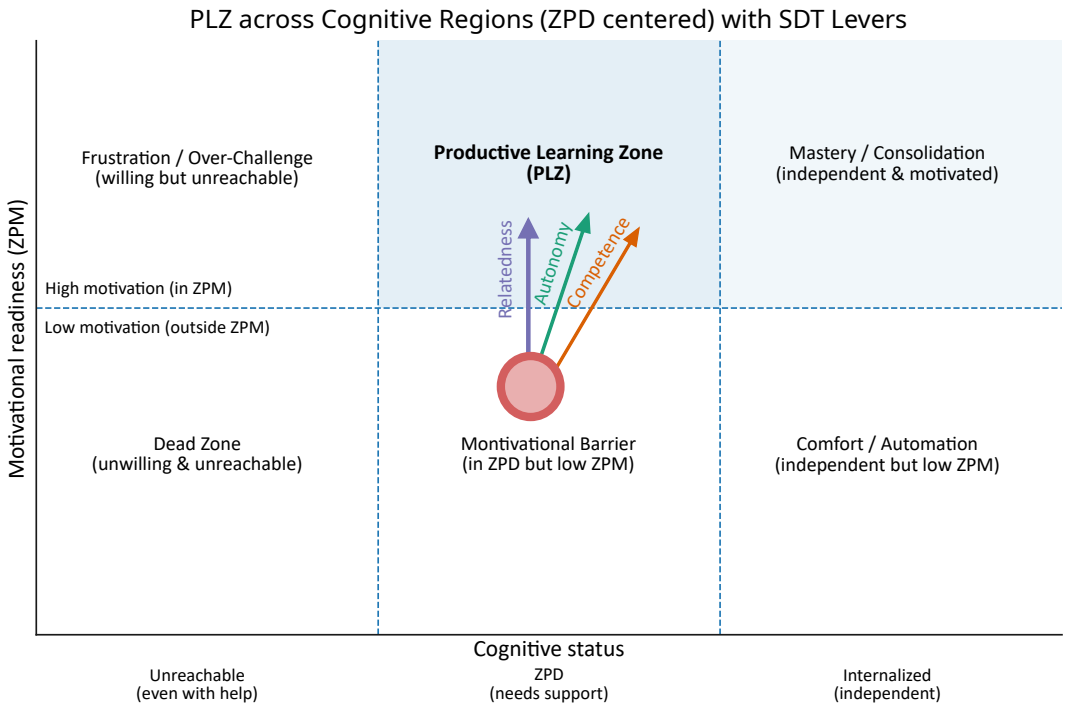

Figure 1: The Dual Zone of Proximal Development (DZPD) framework, which positions the Productive Learning Zone (PLZ) at the intersection of cognitive status (x-axis, ZPD-centered) and motivational readiness (y-axis, ZPM-centered). The red circle marks the learner's current state: cognitively within the ZPD but at risk of low motivation. Arrows represent the three Self-Determination Theory (SDT) needs as regulatory forces: Relatedness (vertical) increases motivational readiness independent of cognitive status, while Autonomy and Competence not only strengthen willingness to invest but also promote mastery gains (slight rightward pull). Together these needs can stabilize the learner within the PLZ. Consolidation processes (knowledge stabilization through generative activities) are not depicted as an axis but are described in the text.

human teacher fosters these motivational conditions intuitively, a DZPD-aware AI can be engineered to explicitly support them through its dialogue and actions. For example, the AI can be designed to infer a dip in the "Dialogue Initiative Ratio" (a proxy for declining autonomy) and trigger system behaviors that (1) cede the generative role back to the learner and (2) provide value-enhancing messages about the importance and relevance of the material to be learned. In doing so, SEVT- and SDT-based principles are transformed from abstract psychological constructs into concrete, manipulable inputs for the AI's adaptive engine.

3.4 Operationalizing the PLZ with Obligatory Generativity and Responsibility (OGR)

3.4.1 Overarching Principle: Obligatory Generativity and Responsibility. The imperative for designing systems that compel generative engagement is not new. Pioneering work like Conati and VanLehn's SE-Coach [19], for instance, demonstrated that AI tutors can successfully scaffold the metacognitive skill of self-explanation, moving learners from passive to active processing. Such systems provided a powerful proof-of-concept for how AI can resist passive learning by prompting for deeper cognitive work.

However, the modern challenge highlighted by the Convenience Paradox—the constant availability of fully synthesized answers—requires a more robust and holistic design principle. It is no longer sufficient to merely invite effortful engagement; we must design systems specifically engineered to anchor learners within their PLZ, the state where they are both cognitively ready (c-ZPD) and motivationally willing (ZPM) to learn. This requires that the interaction itself makes deep engagement not just possible, but obligatory and directed by a clear sense of responsibility.

To this end, we propose an overarching design principle: Obligatory Generativity and Responsibility (OGR). OGR operationalizes this imperative by structuring interactions to encourage learners to engage in generative work—the effortful cognitive acts of explaining, justifying, constructing, and critiquing. This focus on human-led generative work directly answers urgent calls from both learning scientists and industry leaders, who identify critical thinking and the ability to evaluate and build upon AI outputs as paramount skills for the future workforce [6, 29, 39]. Crucially, OGR embeds this generative work within a purposeful social or creative context that establishes responsibility for that knowledge, whether towards an agent, a peer, or a broader audience. It posits that a learning setting designed with OGR is intended to demand both

the cognitive engagement needed to operate in the c-ZPD and the motivational buy-in to sustain the ZPM, thus stabilizing the PLZ.

The three roles of AI learning companions that we focus on—Teachable Agents, Collaborative Companions, and Creative Co-Creators—are archetypical instantiations of this OGR principle. While systems like the SE-Coach masterfully scaffold the "generativity" component, these archetypes add the "responsibility" layer—a social contract that transforms a metacognitive exercise into a meaningful act of teaching, collaborating, or creating. They are not the only ways to stabilize learning in the PLZ, but they powerfully illustrate how OGR can be realized in practice.

3.4.2 Teachable Agents. Research on "learning by teaching" has long demonstrated that students benefit when they are placed in the role of explaining concepts to others [78]. Classical systems such as Betty's Brain [10] operationalize this principle by requiring learners to instruct an artificial protégé. The protégé effect—where students exert more effort and regulate their learning more carefully when they are responsible for teaching—has been consistently observed in empirical work [14].

Recent advances in LLMs have transformed the design of Teachable Agents. Studies show that LLM-based agents such as TeachYou and AlgoBo significantly increase knowledge-building conversational density [51] and can adapt across domains without requiring extensive rule-based engineering. Controlled experiments confirm measurable gains in learning outcomes, including higher post-test scores and reduced cognitive load [68]. These improvements align with the learning-by-teaching framework by fostering metacognition and motivation.

Contribution to DZPD: Teachable Agents compel learners to externalize knowledge in structured ways, activating both their c-ZPD (by engaging with concepts just beyond their current mastery) and their ZPM (by fostering responsibility and a sense of competence). By simulating realistic knowledge gaps and misconceptions, they illustrate how the OGR prevents passive shortcut consumption and reinforces motivational scaffolding.

3.4.3 Collaborative Companions. Collaborative learning has robustly been shown to enhance engagement, persistence, and social motivation [30, 53]. AI-driven companions extend this tradition by assuming peer-like roles in turn-taking dialogues, co-problem solving, and mutual scaffolding [55]. To sustain motivation, it is crucial to avoid over-dominance by the AI, which can diminish autonomy [46].

Recent studies confirm these dynamics. Hu, Hsieh, and Salac [47] demonstrate that AI learning companions improve self-regulation and information literacy, while multi-agent approaches such as MAIC deploy specialized AI peers (teachers, questioners, creative sparkers, note-takers) to diversify interaction modes. Interaction studies with university students show that different engagement styles (active questioners, responsive navigators, and lurkers) all can lead to comparable learning outcomes, suggesting that adaptive companions can flexibly meet individual needs [43].

Contribution to DZPD: Collaborative Companions generate social presence and accountability, sustaining motivation as cognitive demands increase. By situating the learner in a peer-like relationship, these systems operationalize OGR in a social mode, strengthening both competence and relatedness while preventing over-reliance on AI as a provider of ready-made answers.

3.4.4 Creative Co-Creators. Motivation is a central driver of creativity [3]. Research in the LLM era shows that AI can act as a genuine creative partner, supporting ideation, iteration, and sensemaking [22, 59, 66, 76]. Recent systematic reviews confirm that co-creative AI enhances student creativity by acting as facilitator, co-creator, or autonomous generator, depending on the degree of human agency preserved [86].

Experimental evidence is strong: students collaborating with generative AI tools (e.g., ChatGPT, MidJourney, Runway) in digital storytelling achieved large gains in collaborative problem-solving and creativity dimensions, with a reported effect size of 1.18 [86]. These findings suggest that when properly framed, co-creative AI supports ownership, experimentation, and deeper engagement rather than shortcut use.

Contribution to DZPD: Creative Co-Creators expand the motivational landscape by activating curiosity and autonomy, directly linking OGR with exploratory learning. They foster both the competence dimension (through iterative feedback) and autonomy (through open-ended exploration), enabling learners to internalize creative processes instead of outsourcing them to the AI.

3.4.5 Beyond Archetypes: Alternative OGR Methods. While Teachable Agents, Collaborative Companions, and Creative Co-Creators are powerful illustrations, they are not exhaustive. Other designs also implement OGR and support SDT-regulation, for example:

- AI-orchestrated Peer Review: accountability through critique and revision; fosters autonomy (choice of revisions), competence (criteria-based feedback), and relatedness (social exchange).
- AI-Fact-Checking and Red-Teaming: learners must critique AI outputs, preventing shortcut use while supporting autonomy (critical stance) and competence (analytical rigor).
- Learner-Generated Questions and Assessments: students construct exam items; enhances autonomy (choice of focus), competence (taxonomy-based structuring), and relatedness (shared pool).
- Public Portfolios or Artifacts: responsibility to a real audience; supports autonomy (format choice), competence (iteration), and relatedness (audience and peer recognition).
- Learning Contracts and AI Coaching: structured self-regulation; supports autonomy (self-commitment), competence (achievable steps), and relatedness (coach-like interaction).

3.4.6 Synthesis. Across roles and methods, a consistent pattern emerges: OGR designs, when coupled with SDT regulation, prevent learners from slipping into passive shortcut use. Instead, they anchor learners in the PLZ, where cognitive scaffolding (c-ZPD) and motivational scaffolding (ZPM) intersect. Teachable Agents, Collaborative Companions, and Creative Co-Creators thus serve as illustrative archetypes of a more general design principle that emphasizes generativity, responsibility, and motivational alignment.

Principle	Teachable Agents	Collaborative Companions	Creative Co-Creators
P0: Diagnose Before Prescribe	AI elicits a concept map or initial explanation from the learner to surface their prior knowledge and misconceptions.	Companion begins with reflective prompts ("What's your plan?") to gauge the learner's strategy and confidence.	Muse begins by asking about creative goals and constraints (tone, key message) before co-generating content.
P1: Dynamic Role Reciprocity	Agent feigns specific, plausible knowledge gaps, forcing the learner into the "expert" role to correct or explain.	Companion cedes the generative role on a key sub-task, positioning the learner as the leader.	Muse generates a clichéd or flawed starting point, prompting the learner to improve it.
P2a: Constrained Generativity	Goal is a clear, bounded task ("Teach me bubble sort").	Shared, well-defined project ("Let's complete this slide on evaporation").	Joint creative project ("Let's co-draft a storyboard").
P2b: Productive Friction	Protégé simulates misconceptions, forcing iterative explain–revise–correct cycles.	Companion introduces counterarguments or confusion, escalating to hints only if needed.	Muse introduces creative constraints or counterarguments to deepen exploration.
P3: Mandated Articulation	Agent consistently asks "why?" to force learner to justify reasoning.	Companion prompts "walk me through your thinking."	Muse asks for creative rationale behind choices.
P4: Competence-Supportive Feedback	Agent demonstrates flawed result instead of saying "You're wrong."	Companion frames discrepancies as a shared puzzle.	Feedback highlights strengths and process growth opportunities.

Table 2: Mapping of design principles (P0–P4) to the three archetypes of AI learning companions.

4 Strategies for Building Fortified Minds Through AI Learning Companions

We now distill the framework's logic into an actionable toolkit for HCI practitioners. To make our framework actionable, we translate it into a set of principles for pedagogical interaction design—that is, the design of the moment-to-moment dialogue between a learner and an AI system in a way that is guided by established instructional theory. We present this as a prescriptive, hierarchical architecture of actionable design principles (P0-P4) that operationalize our overarching OGR directive. This architecture begins with a crucial diagnostic prerequisite (P0) and is followed by a set of tactical principles (P1-P4) for structuring motivation-aware interactions. Together, these principles provide a clear guide for creating AI-based learning companions that move beyond single examples to a robust and generalizable design pattern.

4.1 Framework: Guiding Principles for DZPD-Aware AI Scaffolding

The successful implementation of AI companions like Teachable Agents, Collaborative Companions, and Creative Co-Creators depends on a set of core design principles. These principles operationalize the central directive of our framework—"AI should keep learners in both zones"—by aiming to ensure every interaction is calibrated to both the learner's ZPD and their ZPM. Taken together, they provide a practical architecture for implementing our overarching principle of OGR. They achieve this by systematically supporting the learner's psychological needs for autonomy, competence, and relatedness, as defined by SDT. The following set of principles details this architecture. To provide a clear overview, Table 2 maps each of the following principles to its concrete instantiation across the three archetypes.

Principle 0: Diagnose before prescribe. AI systems must first diagnose the learner's current motivational state before offering support. This foundational step enables targeted interventions rather than one-size-fits-all approaches. The diagnostic process distinguishes between two primary motivational challenges grounded in expectancy-value theory [13]: expectancy-related problems (the learner doubts their ability to succeed) and value-related problems (the learner doubts the worth of the task). Expectancy problems manifest through linguistic markers of helplessness or perfectionism (e.g., "I can't do this"; excessive frustration over minor errors), while value problems emerge through disengagement signals and challenges to task relevance (e.g., "why do I have to learn this?"). Drawing on conversational logs and behavioral patterns, this diagnostic step is operationalized through existing natural language processing techniques including sentiment analysis, emotion detection, and dialogue act classification. Validated motivational assessments (students' self-report data) will be used to supplement and check the validity of verbal and behavioral trace data. As shown in Table 2, different interaction patterns serve different diagnosed needs—Teachable Agents for expectancy problems, creative co-creation for value problems. This diagnostic prerequisite enables DZPD alignment by ensuring motivational scaffolds address the specific barrier to engagement.

Principle 1: Dynamic Role Reciprocity. The AI's role is not fixed; it must dynamically adjust its own apparent knowledge and capabilities in response to the learner's performance to create optimal challenges. The AI acts as a reciprocal partner whose behavior is designed to ensure the learner remains the "more knowledgeable other" in a key area. This principle directly stabilizes the PLZ by calibrating task difficulty to keep the learner within their c-ZPD.

This successful act of teaching or correcting the AI is a powerful driver for the ZPM, as it reinforces the learner's sense of competence and autonomy. As shown in Table 2, this can be instantiated by a Teachable Agent feigning a plausible misconception or a Creative Co-Creator generating a flawed starting point for the learner to improve.

Principle 2a: Constrained Generativity through Goal-Oriented Scenarios. The interaction must be framed within a clear, purposeful scenario that makes generative engagement the default and most natural path forward. This constraint counters the ambiguity of open-ended prompts and establishes clear responsibility. Goal-oriented scenarios provide cognitive scaffolding (c-ZPD) by defining the problem space into a manageable task. Critically, this is the core implementation of OGR, obliging the learner to generate, explain, or create, which provides the motivational buy-in (ZPM) and fosters relatedness through a shared mission. As detailed in Table 2, this is achieved through explicit goals, such as teaching an agent an algorithm or co-creating a specific artifact.

Principle 2b: Productive Friction with Graduated Support. Rather than minimizing challenge, the AI system should introduce "desirable difficulties" by intentionally calibrating the cognitive costs of the task, while managing frustration through graduated support. This principle is key to building resilience by actively regulating the learner's perceived effort cost to a productive level—not so low that learning is shallow, and not so high that it leads to frustration. The graduated support serves two functions: it provides cognitive assistance to manage the immediate task difficulty, and it provides affective support to manage the ego cost of struggle, framing challenges as opportunities rather than failures. This balance ensures that challenge leads to growth rather than disengagement. Table 2 illustrates how this can be achieved by simulating misconceptions, employing confusion-resolution cycles, or requiring iterative revisions of AI-generated drafts.

Principle 3: Mandated Articulation of Process and Rationale. The AI must consistently prompt the learner to externalize their thinking, demanding not just an answer but also the underlying reasoning. This makes the learner's cognitive process an explicit part of the interaction.

This act of self-explanation is a classic method for inducing cognitive activation; it forces the learner into deliberate, reflective reasoning and is a powerful tool for knowledge consolidation [16]. From a motivational standpoint (ZPM alignment), this is a form of "soft" OGR; by centering the learner's unique thought process, it reinforces their autonomy and competence in a supportive, non-judgmental way. The archetypes in Table 2 demonstrate this through consistent "why?" prompts or requests to "walk me through your thinking."

Principle 4: Fostering Competence and Value through Process-Oriented Feedback. The AI's feedback must be designed to preserve motivational resilience by supporting both the learner's sense of competence and their perception of the task's value. Instead of delivering binary right/wrong judgments, it should support competence by focusing on the learner's process, acknowledging their effort, and providing forward-looking scaffolds to guide self-correction. Crucially for the ZPM, the feedback must also reinforce the value of the effort. This can be achieved by connecting the task to the learner's stated goals (utility value), highlighting an interesting or surprising outcome of their work (interest value), or framing the skill they are developing as an important personal asset (attainment value). This dual focus on competence and value is essential for preventing the demotivation that comes from failure and for reinforcing the learner's choice to invest in a worthwhile endeavor. As seen in the examples in Table 2, this is often achieved by framing errors as a shared puzzle to be solved together, reinforcing the AI's role as a supportive partner.

4.2 Instantiating the Principles: Archetypes and Design Patterns

Furthermore, these principles can guide the creation of new, sophisticated interaction models. For example, a Dual-Agent Teachable System could directly operationalize OGR and the DZPD. In such a system, a learner's primary task would be to teach a concept to a public-facing 'Protégé' agent, which enforces OGR by simulating plausible misconceptions based on the learner's input. Working in parallel, a private 'Scaffolder' agent would monitor the learner's interaction patterns (e.g., response latencies, repeated errors) to keep them within their c-ZPD, offering Socratic hints through a confidential channel. Such a design elegantly separates the motivational pressure of teaching from the cognitive support of tutoring, providing a concrete architecture for an DZPD-aware system.

4.3 Technical Feasibility of the Diagnostic Framework

The diagnostic capabilities required for P0 are well-established through both classical and recent transformer-based approaches. Classical methods demonstrated that sentiment analysis and dialogue act classification can detect learner confidence patterns [36], while behavioral log analysis enables automated motivational state detection [7, 23]. Supplemental self-report data with validated motivational scales based on SEVT and SDT will support the interpretation of these data.

The advent of large language models has dramatically enhanced these capabilities. Recent work shows that GPT-4 and other LLMs can effectively analyze classroom dialogues to detect patterns and trends in educational interactions with high inter-coder reliability compared to human coders [65]. Psychometric evaluations demonstrate that LLMs possess emotional intelligence capabilities that enable understanding of complex emotional scenarios and human affective states [90, 91], while fine-tuned transformer models like BERT achieve substantial improvements in emotion recognition within conversational contexts, specifically detecting confusion and frustration in dialogue [74]. A meta-analysis of 69 experimental studies confirms that ChatGPT interventions demonstrably improve affective-motivational states and reduce mental effort in educational settings [27], providing empirical validation for real-time motivation diagnosis.

For future systems, comprehensive reviews demonstrate that deep learning approaches combining dialogue analysis with multimodal inputs (including eye-tracking and facial expressions) achieve robust emotion detection in educational contexts [1, 31, 42].

The implementation pathway for P0 diagnostics involves three technical components:

(1) Expectancy-related detection through sentiment analysis identifying negative self-talk patterns and behavioral logs revealing rapid guessing or giving-up behaviors;
(2) Value-related detection through intent classification identifying utility challenges and dialogue act classification revealing low ratios of curiosity-driven versus procedural questions; and
(3) Integrated decision logic that synthesizes these signals into actionable diagnostic categories.

This technical foundation transforms P0 from aspiration into a concrete, implementable research program.

5 A Heuristic Toolkit for Designing and Evaluating DZPD-Aware Systems

To complement these design principles, this section provides a heuristic toolkit of measurable indicators for evaluating whether DZPD-aware systems are working as intended. Critically, these indicators are designed to move beyond surface-level metrics (e.g., completion time, score) to capture the deeper dynamics of generative and motivationally resilient learning. We position these not as validated instruments, but as heuristic, computable behavioral proxies that can be used to empirically test whether a system—such as a future prototype of our Dual-Agent exemplar—is successfully maintaining learners in the DZPD.

1. Frequency of Learner Explanations vs. Agent Prompts. This metric measures the proportion of dialogue turns in which the learner formulates a generative statement (e.g., an explanation, a teaching step) versus passively receiving information.

- **Principle Alignment:** A direct measure of P3 (Mandated Articulation) and P1 (Dynamic Role Reciprocity).
- **Archetypal Context:** Producing explanations in conversational tutors like AutoTutor [40], articulating teaching steps in Betty's Brain [10], or proposing ideas in Creative Co-Creators [85].
- **DZPD Connection:** Explains cognitive edge (ZPD) and motivational willingness (ZPM).

2. Dialogue Initiative Ratio (Learner vs. Agent). Tracks learner agency by comparing student-initiated utterances to agent-initiated prompts.

- **Principle Alignment:** Measures P1 (Dynamic Role Reciprocity).
- **Archetypal Context:** Mixed-initiative conversational balance in advanced tutors [40].
- **DZPD Connection:** High learner initiative = autonomy and stronger learning gains.

3. Scenario-Alignment Ratio (On-Task Action Ratio). Assesses the ratio of learner actions relevant to the scenario goal versus off-task actions.

- **Principle Alignment:** Primary measure for P2a (Constrained Generativity).
- **Archetypal Context:** Detects "gaming the system" [7].
- **DZPD Connection:** High ratio = strong buy-in and task engagement.

4. Confusion-Recovery and Affective Transition Cycles. Monitors transitions such as ENG → CON → ENG.

- **Principle Alignment:** Direct evidence of P2b (Productive Friction) and P4 (Competence-Supportive Feedback).
- **Archetypal Context:** Documented in Betty's Brain [10].
- **DZPD Connection:** Successful recovery builds resilient competence (ZPM).

5. Scaffold Fading Rate / Support Dependency Over Time. Measures how quickly AI-provided support diminishes as learners gain competence.

- **Principle Alignment:** Tracks P2b (Productive Friction) and P0 (Diagnose Before Prescribe).
- **Archetypal Context:** Adaptive support and scaffold withdrawal.
- **DZPD Connection:** Shows expansion of independent capacity (c-ZPD) and motivation (ZPM).

6. Persistence Through Failure. Tracks percentage of learners reengaging after failure.

- **Principle Alignment:** Holistic measure of ZPM alignment (P4 + P2b).
- **Archetypal Context:** General outcome metric across all systems.
- **DZPD Connection:** Proof of motivational scaffolding effectiveness.

7. Strategy Monitoring Frequency. Detects how often learners verbalize strategies after challenges.

- **Principle Alignment:** Reflects P3 (Mandated Articulation) and P0 (Diagnose Before Prescribe).
- **Archetypal Context:** Measures metacognitive reflection.
- **DZPD Connection:** Expression of autonomy and ownership.

8. Shortcut-Use Rate (Learner vs. System). Tracks frequency of solution-seeking shortcuts ("just tell me," copy-paste, exploiting hints).

- **Principle Alignment:** Negative measure for P2a (Constrained Generativity) and P2b (Productive Friction).
- **Archetypal Context:** Captures "gaming the system" [7].
- **DZPD Connection:** High rate signals exit from DZPD due to c-ZPD/ZPM mismatch.

6 Illustrative Evidence & Feasibility probe

Rather than a summative evaluation, we provide illustrative evidence that the DZPD framework and the OGR principle are operable and theoretically tractable. Our strategy has two parts: first, synthesizing empirical precedents from prior systems as indirect validation of our design levers and indicators; second, presenting a modest feasibility probe that implements OGR and applies our heuristic indicators. The probe builds on the Dual-Agent Teachable System (§ 4.2), showing how a DZPD-aware interaction can

be prototyped and observed. Together, these strands do not establish causal effects but instead illustrate feasibility, clarify design patterns, and specify testable propositions for future HCI research.

6.1 Indirect Empirical Precedents

As outlined in Section 3.4, our framework is instantiated through design archetypes such as Teachable Agents, Collaborative Companions, and Creative Co-Creators. These archetypes are not hypothetical: prior systems provide converging empirical precedents that demonstrate the mechanisms emphasized in our design levers and indicators. For example, studies of Teachable Agents (e.g., Betty's Brain, SimStudent) show that requiring learners to articulate explanations increases generative effort and self-regulation—directly mapping to Indicator 1 (Learner Explanations) and supporting the OGR principle. Similarly, research on collaborative learning companions and mixed-initiative dialogue systems (e.g., AutoTutor [41, 73], PeerLogic) documents productive confusion cycles and shared initiative, aligning with Indicators 2 and 4. Work on co-creative systems in programming and writing highlights the balance between autonomy and scaffolded support, pointing to Indicators 5 and 8.

Taken together, these strands of evidence operate as *indirect empirical confirmation* of our framework. They show that the mechanisms emphasized in DZPD and OGR—generativity, responsibility, shortcut regulation, and motivational scaffolding—have already been observed across diverse systems and methods. Our contribution is to integrate these isolated effects into a coherent HCI framework, specify them as computable indicators, and tailor them to the distinctive challenges of generative AI.

6.2 Design Probe: Feasibility of OGR in Practice

Unlike Section 6.1's indirect validation via prior systems, here we report a design probe (N = 8) intended as a proof-of-concept for operationalizing OGR and our heuristic indicators in a new domain; it is not a systematic evaluation or a comparative test of effectiveness.

We implemented a prototype of the **Dual-Agent Teachable System** described in Section 4.2 using the SillyTavern[1] framework due to its adjustability and focus on multi-character interactions together with a self-hosted AWQ quantized version of Mistral-Large-Instruct-2407 model[2] [64, 70]. Participants were asked to teach a "Protégé" chatbot the concept of finite differences, including their role in approximating derivatives, approximation behavior, and stability. In the background, a second "Scaffolder" chatbot was present to monitor the interaction, consistent with our design pattern. Pre-questionnaires revealed a moderate level of self-reported prior knowledge in analysis and numerical methods but little or no familiarity with finite differences. Several also reported regular use of AI for explanations or exercises. Building on participants' prior knowledge, we designed a 30-minute "learning phase" during which they watched a tutorial video on finite differences. The video introduced the most essential concepts to provide foundational understanding and to ensure that the subsequent task was situated within the participants' c-ZPD.

[1]https://github.com/SillyTavern/SillyTavern
[2]https://huggingface.co/TechxGenus/Mistral-Large-Instruct-2407-AWQ

A descriptive visualization of post-task responses in Figure 2 shows responses skewed toward agreement that the discussions were motivating, that participants had to engage intensively and think for a long time to answer questions, and that this effort improved their own understanding; disagreement predominated for feeling overwhelmed; and perceptions of an overly "watchful" supervisor were low. Many also agreed they learned more than with passive AI use, consistent with OGR's goal to privilege generativity over answer-taking. These patterns are descriptive, not inferential, and are presented to illustrate feasibility and signal direction rather than to claim effects.

Questionnaire triangulation (N = 8). Self-reported data from the feasibility probe Figure 2 can be directly mapped to the heuristic indicators proposed in Section 5, showing their practical observability.

- Strong agreement with statements such as "I had to engage intensively" and "I had to think for a long time" aligns with **Indicator 1 (Learner Explanations)**, reflecting the system's ability to compel generative effort.
- Low agreement with "I felt overwhelmed" corresponds to **Indicator 4 (Confusion-Recovery Cycles)**, suggesting the probe produced "desirable difficulty" and "productive friction" rather than frustration.
- Perceived added value ("I learned more than if I had simply used AI on my own") supports **Indicator 8 (Shortcut-Use Rate)**, highlighting recognition of the generative process over passive answer-seeking.
- Low perception of being "watched" relates to **Indicator 5 (Support Dependency)**, indicating unobtrusive scaffolding that enables autonomy and fading support.

Together, these mappings illustrate that the proposed *heuristics are not only conceptual but traceable in user experience data*, underscoring the empirical tractability of a DZPD-aware system.

Summary of the Feasibility Probe. While the probe was deliberately modest in scope—with a small sample, brief exposure, and reliance on self-report—it provides hypothesis-generating signals rather than causal claims. More importantly, it demonstrates how our heuristic indicators can be applied in practice. In doing so, it closes the loop from the Dual-Agent design pattern (§4.2) to observable indicators, illustrating the empirical tractability of the framework and pointing toward directions for systematic evaluation.

7 Discussion and Future Work

7.1 Main Contributions and Synthesis

This paper advances a theory-and-design perspective on AI-supported learning. We introduced the Dual-ZPD framework (DZPD), articulated through the Productive Learning Zone (PLZ), and operationalized it via the OGR principle, design principles (P0–P4), and heuristic indicators. These elements together provide both a conceptual lens for understanding how cognitive and motivational scaffolding interact, and a practical toolkit for designing and studying DZPD-aware systems.

Section 6 demonstrated the illustrative operability of these ideas: precedents from prior systems converge on our proposed mechanisms, and the feasibility probe showed how indicators can be

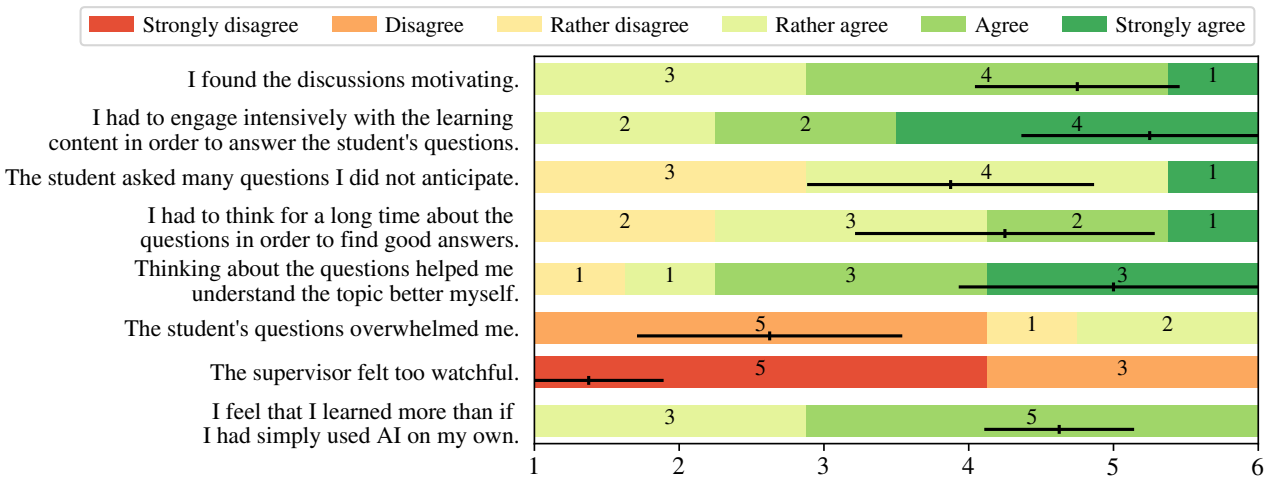

Figure 2: Results of our questionnaire (N = 8, 6-point Likert scale) after participating in the feasibility probe. In addition to the proportions of responses, we compute a mean and standard deviation by converting all possible choices into equally-spaced numerical values between 0 (Strongly disagree) and 5 (Strongly agree) and visualizing these results as black error bars for each question.

applied in practice. This creates a throughline from theory (DZPD and propositions), to design (OGR, principles, archetypes), to evidence (precedents and probe).

In sum, the contributions reframe AI in education from answer engines toward learning partners, equipping HCI with concepts, design patterns, and testable propositions that motivate future empirical research.

7.2 Ethics and Learner Agency

Any system designed to influence motivation must confront the risk of becoming paternalistic or coercive. Our framework addresses this ethical challenge directly through the OGR principle's deep integration with SDT. OGR avoids manipulative "nudging" by framing generative work not as a system demand, but as a meaningful responsibility within a social context (e.g., teaching Clara), allowing the learner to autonomously endorse the effortful path and transform obligation into agency.

However, this autonomy-supportive design is not without risks. The concept of "obligatory" engagement treads a fine line, and its implementation requires careful ethical consideration. For example, what are the consequences if the AI misjudges the learner's ZPM and introduces "productive friction" at a moment of genuine frustration? This could lead to fatigue or premature disengagement. Furthermore, we must ensure transparency and consent regarding the system's motivational goals; learners should understand why the system is not simply providing an answer. Future work must investigate fail-safes and learner overrides to ensure that the system's pursuit of generative engagement does not come at the cost of the learner's well-being and ultimate control over their learning process.

7.3 Limitations

This work is presented as a theory-and-design framework, not a summative evaluation. The indicators are heuristic rather than validated instruments, and the design principles are scaffolds for AI-mediated contexts, not a universal pedagogy. The framework is conceptual, synthesizing prior research without new large-scale user studies; to bridge this gap, we provide a heuristic toolkit (Section 5) to support the future validation of DZPD and its principles.

On a technical level, real-time estimation of ZPM currently relies on approximate behavioral proxies; robust implementation will require advances in prompt engineering and multimodal sensing. On a practical level, while the principles are designed to be general, their application must be adapted to different domains (e.g., mathematics vs. creative writing), learner populations, cultural contexts, and neurotypes

In addition, the principle of "obligatory" engagement raises important ethical considerations. While our framework is designed to be explicitly autonomy-supportive, any system that actively modulates motivation must be carefully designed to avoid becoming coercive or paternalistic. Future work must investigate how to ensure these methods are equitable and empowering for all learners, and we recommend that deployments of such systems include audits to safeguard against potential coercion.

Finally, our review of related work, while broad, is necessarily selective. To fully situate our framework, future iterations would benefit from deeper engagement with several key research areas, including: (i) recent advances in motivation-aware AI tutoring systems; (ii) the application of "desirable difficulties" in AI-supported learning contexts; (iii) learner agency in mixed-initiative educational technologies; and (iv) the interplay of intrinsic versus extrinsic motivation in gamified learning environments.

7.4 Future Work

The limitations of this paper point directly toward a rich agenda for future research and development.

The immediate priority is to enhance the empirical validation of the DZPD framework. This involves building upon the initial feasibility probe by extending and more rigorously testing the Dual-Agent Teachable System described in Section 4.1 and 6.2 and further prototype AI companions based on our design principles (e.g., an agent embodying "Dynamic Role Reciprocity" and "Competence-Supportive Feedback") and testing their impact on learners. Using the proposed indicators—such as the dialogue initiative ratio and confusion–recovery cycles—researchers can quantitatively measure whether these systems are more effective at fostering deep learning and motivational resilience compared to conventional "answer-engine" AI.

Second, the real-time operationalization of the ZPM remains a significant technical challenge. While this paper posits that ZPM can be inferred from "behavioral proxies," we acknowledge that doing so reliably is a non-trivial problem. This constitutes a "black box" that future research must address. Robust implementation will likely require a sophisticated fusion of methods, including model-based inference (e.g., tracking a drop in the Dialogue Initiative Ratio as a proxy for declining autonomy) and potentially multimodal sensing (e.g., affect detection from text or eye-tracking data to infer frustration or disengagement). We position our heuristic toolkit not as a final solution, but as a necessary first step in defining the signals that such systems will need to capture.

A next step is the integration of DZPD and OGR into authentic curricular settings. We are preparing a deployment in our undergraduate Numerics lecture (Computer Science, $N \approx 250$), where students may earn exercise credit by creating small teaching applications that explain a topic of their choice from recent lectures. While this design resonates with the classic "protégé effect" of Teachable Agents, its novelty lies in three aspects: first, motivational scaffolding is embedded directly into curricular incentives rather than offered as an auxiliary activity; second, the setting is at scale and embedded in a core lecture, providing a natural contrast between OGR and non-OGR paths; and third, the analysis will focus on motivational indicators such as persistence and shortcut use rather than solely on cognitive gains. This study will thus extend DZPD and OGR into authentic educational practice.

Furthermore, a particularly promising avenue for this research is to investigate how different modalities can amplify the relational aspects of the DZPD. For example, one could implement the AI Protégé and Scaffolder not as text-based chatbots, but as photorealistic, emotive video avatars using generative video platforms (e.g., Synthesia). This would allow for a much richer test of SDT's need for relatedness. An AI Protégé that can express genuine-seeming confusion through its facial expressions, or an AI Scaffolder that offers a nod of encouragement, could provide a far more powerful motivational scaffold for the ZPM than text alone. This would move the interaction from a purely cognitive–dialogic partnership to a more holistic socio-emotional one, truly testing the principles of motivation-aware design.

A significant practical contribution would be the development of a detailed design pattern library for DZPD-aware EdTech. This would translate our principles into concrete, reusable implementation recipes for AI engineers and learning designers. Such a library would greatly lower the barrier to creating educational tools that actively counter the Convenience Paradox.

Looking forward, we envision AI systems that can dynamically and autonomously regulate a learner's journey within the PLZ. By integrating real-time data from multimodal sensing (e.g., eye-tracking, facial expression analysis, speech prosody) to estimate a learner's cognitive load and motivational state, AI could one day shift seamlessly between the roles of tutor, Teachable Agent, and Creative Co-Creator, providing precisely the right form of cognitive and motivational scaffolding at exactly the right moment. Such a development would mark the true arrival of AI not just as a tool, but as a genuine partner in human learning.

Acknowledgments

Fani Lauermann and Daria Benden acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC2126/1-390838866).

References

[1] Sharmeen M.Saleem Abdullah Abdullah, Siddeeq Y. Ameen Ameen, Mohammed A. M. Sadeeq, and Subhi Zeebaree. 2021. Multimodal Emotion Recognition using Deep Learning. *Journal of Applied Science and Technology Trends* 2, 01 (May 2021), 73–79. doi:10.38094/jastt20291

[2] Elizabeth Acosta-Gonzaga and Aldo Ramirez-Arellano. 2022. Scaffolding Matters? Investigating Its Role in Motivation, Engagement and Learning Achievements in Higher Education. *Sustainability* 14, 20 (2022). doi:10.3390/su142013419

[3] Teresa M Amabile. 1996. *Creativity in context: Update to the social psychology of creativity.* Routledge.

[4] Roger Azevedo, François Bouchet, Melissa Duffy, Jason Harley, Michelle Taub, Gregory Trevors, Elizabeth Cloude, Daryn Dever, Megan Wiedbusch, Franz Wortha, et al. 2022. Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology* 13 (2022), 813632. doi:10.3389/fpsyg.2022.813632

[5] Hongjun Ba, Lili Zhang, and Zizheng Yi. 2024. Enhancing clinical skills in pediatric trainees: a comparative study of ChatGPT-assisted and traditional teaching methods. *BMC Medical Education* 24, 1 (2024), 558. doi:10.1186/s12909-024-05565-1

[6] Leili Babashahi, Carlos Eduardo Barbosa, Yuri Lima, Alan Lyra, Herbert Salazar, Matheus Argôlo, Marcos Antonio de Almeida, and Jano Moreira de Souza. 2024. AI in the Workplace: A Systematic Review of Skill Transformation in the Industry. *Administrative Sciences* 14, 6 (2024). doi:10.3390/admsci14060127

[7] Ryan Shaun Baker, Albert T. Corbett, Kenneth R. Koedinger, and Angela Z. Wagner. 2004. Off-task behavior in the cognitive tutor classroom: when students "game the system". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04).* Association for Computing Machinery, 383–390. doi:10.1145/985692.985741

[8] Albert Bandura. 1997. *Self-efficacy: The exercise of control.* Macmillan.

[9] Arne Bewersdorff, Christian Hartmann, Marie Hornberger, Kathrin Seßler, Maria Bannert, Enkelejda Kasneci, Gjergji Kasneci, Xiaoming Zhai, and Claudia Nerdel. 2025. Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences* 118 (2025), 102601. doi:10.1016/j.lindif.2024.102601

[10] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. 2005. LEARNING BY TEACHING: A NEW AGENT PARADIGM FOR EDUCATIONAL SOFTWARE. *Applied Artificial Intelligence* 19, 3-4 (2005), 363–392. doi:10.1080/08839510590910200

[11] Robert A Bjork. 1994. Memory and metamemory considerations in the training of human beings. *Metacognition: Knowing about knowing* 185, 7.2 (1994), 185–205.

[12] Jere Brophy. 1999. Toward a Model of the Value Aspects of Motivation in Education: Developing Appreciation for Particular Learning Domains and Activities. *Educational Psychologist* 34, 2 (1999), 75–85. doi:10.1207/s15326985ep3402_1

[13] Jere Brophy. 2004. *Motivating students to learn.* Routledge.

[14] Catherine C Chase, Doris B Chin, Marily A Oppezzo, and Daniel L Schwartz. 2009. Teachable agents and the protégé effect: Increasing the effort towards learning.

Journal of science education and technology 18, 4 (2009), 334–352. doi:10.1007/s10956-009-9180-4

[15] John Chen, Xi Lu, Yuzhou Du, Michael Rejtig, Ruth Bagley, Mike Horn, and Uri Wilensky. 2024. Learning Agent-based Modeling with LLM Companions: Experiences of Novices and Experts Using ChatGPT & NetLogo Chat. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, Article 141, 18 pages. doi:10.1145/3613904.3642377

[16] Michelene T.H. Chi, Nicholas De Leeuw, Mei-Hung Chiu, and Christian Lavancher. 1994. Eliciting self-explanations improves understanding. *Cognitive Science* 18, 3 (1994), 439–477. doi:10.1016/0364-0213(94)90016-7

[17] Zhendong Chu, Shen Wang, Jian Xie, Tinghui Zhu, Yibo Yan, Jinheng Ye, Aoxiao Zhong, Xuming Hu, Jing Liang, Philip S. Yu, and Qingsong Wen. 2025. LLM Agents for Education: Advances and Applications. arXiv:2503.11733 [cs.CY] https://arxiv.org/abs/2503.11733

[18] Andy Clark and David Chalmers. 1998. The Extended Mind. *Analysis* 58, 1 (1998), 7–19. http://www.jstor.org/stable/3328150

[19] Cristina Conati and Kurt Vanlehn. 2000. Toward Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. *International Journal of Artificial Intelligence in Education* 11 (2000), 389–415. https://telearn.hal.science/hal-00197335 Part I of the Special Issue on Analysing Educational Dialogue Interaction (editor: Rachel Pilkington).

[20] Diana-Margarita Córdova-Esparza. 2025. AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. *Information* 16, 6 (2025). doi:10.3390/info16060469

[21] Chih-Pu Dai, Fengfeng Ke, Yanjun Pan, Jewoong Moon, and Zhichun Liu. 2024. Effects of Artificial Intelligence-Powered Virtual Agents on Learning Outcomes in Computer-Based Simulations: A Meta-Analysis. *Educational Psychology Review* 36, 1 (2024), 31. doi:10.1007/s10648-024-09855-4

[22] Nicholas Davis, Chih-Pin Hsiao, Yanna Popova, and Brian Magerko. 2015. *An Enactive Model of Creativity for Computational Collaboration and Co-creation*. Springer London, London, 109–133. doi:10.1007/978-1-4471-6681-8_7

[23] Angel de Vicente and Helen Pain. 2002. Informing the Detection of the Students' Motivational State: An Empirical Study. In *Intelligent Tutoring Systems*, Stefano A. Cerri, Guy Gouardères, and Fàbio Paraguaçu (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 933–943.

[24] Edward L. Deci and Richard M. Ryan. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory* 11, 4 (2000), 227–268. doi:10.1207/S15327965PLI1104_01

[25] Teresa Del Solato and Benedict Du Boulay. 1995. Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education* 6 (1995), 337–378. doi:10.1007/s40593-015-0052-1

[26] Fabrizio Dell'Acqua, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24-013 (2023). doi:10.2139/ssrn.4573321

[27] Ruiqi Deng, Maoli Jiang, Xinlu Yu, Yuyan Lu, and Shasha Liu. 2025. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education* 227 (2025), 105224. doi:10.1016/j.compedu.2024.105224

[28] Daryn A. Dever, Megan D. Wiedbusch, Sarah M. Romero, and Roger Azevedo. 2024. Investigating pedagogical agents' scaffolding of self-regulated learning in relation to learners' subgoals. *British Journal of Educational Technology* 55, 4 (2024), 1290–1308. doi:10.1111/bjet.13432

[29] Digital Education Council. 2025. AI in the workplace 2025: Understanding industry needs: What employers expect.

[30] Pierre Dillenbourg. 1999. *Collaborative learning: Cognitive and computational approaches. advances in learning and instruction series*. ERIC.

[31] Sidney K D'Mello and Art C Graesser. 2014. 31 Feeling, Thinking, and Computing with Affect-Aware Learning. *The Oxford handbook of affective computing* (2014), 419.

[32] Benedict du Boulay. 2018. Intelligent tutoring systems that adapt to learner motivation. *Tutoring and intelligent tutoring systems* (2018), 103–128.

[33] Jacquelynne S. Eccles and Allan Wigfield. 2020. From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology* 61 (2020), 101859. doi:10.1016/j.cedpsych.2020.101859

[34] Justin Edwards, Andy Nguyen, Marta Sobocinski, Joni Lämsä, Adelson de Araujo, Belle Dang, Ridwan Whitehead, Anni-Sofia Roberts, Matti Kaarlela, and Sanna Jarvela. 2024. MAI - A Proactive Speech Agent for Metacognitive Mediation in Collaborative Learning. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (CUI '24)*. Association for Computing Machinery, Article 46, 5 pages. doi:10.1145/3640794.3665585

[35] Andrew J. Elliot. 1999. Approach and avoidance motivation and achievement goals. *Educational Psychologist* 34, 3 (1999), 169–189. doi:10.1207/s15326985ep3403_3

[36] Aysu Ezen-Can and Kristy Elizabeth Boyer. 2015. Understanding Student Language: An Unsupervised Dialogue Act Classification Approach. *Journal of Educational Data Mining* 7, 1 (2015), 51–78.

[37] Matthew Fisher, Mariel K Goddu, and Frank C Keil. 2015. Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of experimental psychology: General* 144, 3 (2015), 674.

[38] Jennifer A. Fredricks, Michael Filsecker, and Michael A. Lawson. 2016. Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction* 43 (2016), 1–4. doi:10.1016/j.learninstruc.2016.02.002 Special Issue: Student engagement and learning: theoretical and methodological advances.

[39] Joachim Funke. 2025. *Critical Thinking: A Key Competency in the Twenty-First Century to Deal with Uncertainty and Complexity*. Springer Nature Switzerland, Cham, 109–123. doi:10.1007/978-3-031-82640-5_5

[40] A.C. Graesser, P. Chipman, B.C. Haynes, and A. Olney. 2005. AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education* 48, 4 (2005), 612–618. doi:10.1109/TE.2005.856149

[41] Arthur C Graesser, Shulan Lu, George Tanner Jackson, Heather Hite Mitchell, Mathew Ventura, Andrew Olney, and Max M Louwerse. 2004. AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers* 36, 2 (2004), 180–192. doi:10.3758/BF03195563

[42] Swadha Gupta, Parteek Kumar, and Raj Kumar Tekchandani. 2023. Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications* 82, 8 (2023), 11365–11394. doi:10.1007/s11042-022-13558-9

[43] Zhanxin Hao, Jianxiao Jiang, Jifan Yu, Zhiyuan Liu, and Yu Zhang. 2025. Student-AI Interaction in an LLM-Empowered Learning Environment: A Cluster Analysis of Engagement Profiles. arXiv:2503.01694 [cs.CY] https://arxiv.org/abs/2503.01694

[44] José Hernández-Orallo. 2025. Enhancement and assessment in the AI age: An extended mind perspective. *Journal of Pacific Rim Psychology* 19 (2025), 18344909241309376. doi:10.1177/18344909241309376

[45] Wayne Holmes, Maya Bialik, and Charles Fadel. 2019. *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign.

[46] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2019. Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. In *Artificial Intelligence in Education*, Seiji Isotani, Eva Millán, Amy Ogan, Peter Hastings, Bruce McLaren, and Rose Luckin (Eds.). Springer International Publishing, Cham, 157–171.

[47] Yung-Hsiang Hu, Chieh-Lun Hsieh, and Ellen S.N. Salac. 2024. Advancing freshman skills in information literacy and self-regulation: The role of AI learning companions and Mandala Chart in academic libraries. *The Journal of Academic Librarianship* 50, 3 (2024), 102885. doi:10.1016/j.acalib.2024.102885

[48] Jonathan S Ilgen, Anique BH de Bruin, Pim W Teunissen, Jonathan Sherbino, and Glenn Regehr. 2021. Supported independence: the role of supervision to help trainees manage uncertainty. *Academic medicine* 96, 11S (2021), S81–S86. doi:10.1097/ACM.0000000000004308

[49] Michael Inzlicht, Amitai Shenhav, and Christopher Y. Olivola. 2018. The Effort Paradox: Effort Is Both Costly and Valued. *Trends in cognitive sciences* 22, 4 (April 2018), 337–349. doi:10.1016/j.tics.2018.01.007

[50] Toru Ishikawa, Hiromichi Fujiwara, Osamu Imai, and Atsuyuki Okabe. 2008. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology* 28, 1 (2008), 74–82. doi:10.1016/j.jenvp.2007.09.002

[51] Hyoungwook Jin, Seonghee Lee, Hyungyu Shin, and Juho Kim. 2024. Teach AI How to Code: Using Large Language Models as Teachable Agents for Programming Education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, Article 652, 28 pages. doi:10.1145/3613904.3642349

[52] Lingxi Jin, Baicheng Lin, Mengze Hong, Kun Zhang, and Hyo-Jeong So. 2025. Exploring the Impact of an LLM-Powered Teachable Agent on Learning Gains and Cognitive Load in Music Education. arXiv:2504.00636 [cs.HC] https://arxiv.org/abs/2504.00636

[53] David W Johnson, Roger T Johnson, and Karl A Smith. 2024. Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching* 25, 3&4 (Apr. 2024). https://celt.miamioh.edu/index.php/JECT/article/view/454

[54] Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.

[55] Yanghee Kim and Amy L Baylor. 2016. Research-Based Design of Pedagogical Agent Roles: a Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 160–169. doi:10.1007/s40593-015-0055-y

[56] Christian R. Klein and Reinhard Klein. 2025. The Extended Hollowed Mind: Why Foundational Knowledge Is Indispensable in the Age of AI. *Frontiers in Artificial Intelligence* 8 (2025), 1719019. doi:10.3389/frai.2025.1719019

[57] Martin Klein. 2019. Self-determination theory: Basic psychological needs in motivation, development, and wellness. *Sociologicky Casopis* 55, 3 (2019), 412–413.

[58] Jeremy Knox. 2020. Artificial intelligence and education in China. *Learning, Media and Technology* 45, 3 (2020), 298–311. doi:10.1080/17439884.2020.1754236

[59] Max Kreminski, Melanie Dickinson, Michael Mateas, and Noah Wardrip-Fruin. 2020. Why Are We Like This?: The AI Architecture of a Co-Creative Storytelling Game. In *Proceedings of the 15th International Conference on the Foundations of Digital Games (FDG '20)*. Association for Computing Machinery, Article 13, 4 pages. doi:10.1145/3402942.3402953

[60] James A. Kulik and J. D. Fletcher. 2016. Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research* 86, 1 (2016), 42–78. doi:10.3102/0034654315581420

[61] David Laibson. 1997. Golden Eggs and Hyperbolic Discounting*. *The Quarterly Journal of Economics* 112, 2 (05 1997), 443–478. doi:10.1162/003355397555253

[62] Ehsan Latif, Ramviyas Parasuraman, and Xiaoming Zhai. 2024. PhysicsAssistant: An LLM-Powered Interactive Learning Robot for Physics Lab Investigations. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. 864–871. doi:10.1109/RO-MAN60168.2024.10731312

[63] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, Article 1121, 22 pages. doi:10.1145/3706598.3713778

[64] Ji Lin, Jiaming Tang, Haotian Tang, Shang Yang, Wei-Ming Chen, Wei-Chen Wang, Guangxuan Xiao, Xingyu Dang, Chuang Gan, and Song Han. 2024. AWQ: Activation-aware Weight Quantization for On-Device LLM Compression and Acceleration. In *Proceedings of Machine Learning and Systems*, P. Gibbons, G. Pekhimenko, and C. De Sa (Eds.), Vol. 6. 87–100.

[65] Yun Long, Haifeng Luo, and Yu Zhang. 2024. Evaluating large language models in analysing classroom dialogue. *npj Science of Learning* 9, 1 (2024), 60. doi:10.1038/s41539-024-00273-3

[66] Todd Lubart. 2005. How can computers be partners in the creative process: Classification and commentary on the Special Issue. *International Journal of Human-Computer Studies* 63, 4 (2005), 365–369. doi:10.1016/j.ijhcs.2005.04.002 Computer support for creativity.

[67] Rose Luckin. 2017. Towards artificial intelligence-based assessment systems. *Nature Human Behaviour* 1, 3 (2017), 0028. doi:10.1038/s41562-016-0028

[68] Bailing Lyu, Chenglu Li, Hai Li, Hyunju Oh, Yukyeong Song, Wangda Zhu, and Wanli Xing. 2025. The Role of Teachable Agents' Personality Traits on Student-AI Interactions and Math Learning. *Computers & Education* 234 (2025), 105314. doi:10.1016/j.compedu.2025.105314

[69] Qianou Ma, Hua Shen, Kenneth Koedinger, and Sherry Tongshuang Wu. 2024. How to Teach Programming in the AI Era? Using LLMs as a Teachable Agent for Debugging. In *Artificial Intelligence in Education*, Andrew M. Olney, Irene-Angelica Chounta, Zitao Liu, Olga C. Santos, and Ig Ibert Bittencourt (Eds.). Springer Nature Switzerland, 265–279.

[70] Mistral.ai. 2022. Mistral Large. https://mistral.ai/en/news/mistral-large Accessed: 2025-02-18.

[71] Ethan Mollick. 2024. *Co-intelligence: Living and working with AI.* Penguin.

[72] Shakked Noy and Whitney Zhang. 2023. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science* 381, 6654 (2023), 187–192. doi:10.1126/science.adh2586

[73] Benjamin D Nye, Arthur C Graesser, and Xiangen Hu. 2014. AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education* 24, 4 (2014), 427–469. doi:10.1007/s40593-014-0029-5

[74] Xiangyu Qin, Zhiyu Wu, Tingting Zhang, Yanran Li, Jian Luan, Bin Wang, Li Wang, and Jinshi Cui. 2023. BERT-ERC: fine-tuning BERT is enough for emotion recognition in conversation. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23)*. AAAI Press, Article 1513, 9 pages. doi:10.1609/aaai.v37i11.26582

[75] Jennifer M Reddig, Arav Arora, and Christopher J MacLellan. 2025. Generating in-context, personalized feedback for intelligent tutors with large language models. *International Journal of Artificial Intelligence in Education* (2025), 1–42. doi:10.1007/s40593-025-00505-6

[76] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 67 (Sept. 2023), 28 pages. doi:10.1145/3519026

[77] Ido Roll and Ruth Wylie. 2016. Evolution and revolution in artificial intelligence in education. *International journal of artificial intelligence in education* 26, 2 (2016), 582–599. doi:10.1007/s40593-016-0110-3

[78] Rod D. Roscoe and Michelene T. H. Chi. 2007. Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors' Explanations and Questions. *Review of Educational Research* 77, 4 (2007), 534–574. doi:10.3102/0034654307309920

[79] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[80] Richard M. Ryan and Edward L. Deci. 2020. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology* 61 (2020), 101860. doi:10.1016/j.cedpsych.2020.101860

[81] Betsy Sparrow, Jenny Liu, and Daniel M. Wegner. 2011. Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science* 333, 6043 (Aug. 2011), 776–778. doi:10.1126/science.1207745

[82] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (1988), 257–285. doi:10.1207/s15516709cog1202_4

[83] Susan Telio, Rola Ajjawi, and Glenn Regehr. 2015. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine* 90, 5 (2015), 609–614. doi:10.1097/ACM.0000000000000560

[84] Summer Telio, Glenn Regehr, and Rola Ajjawi. 2016. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 50, 9 (2016), 933–942. doi:10.1111/medu.13063

[85] Alex Urmeneta and Margarida Romero. 2024. *Creative applications of artificial intelligence in education.* Springer Nature. doi:10.1007/978-3-031-55272-4

[86] Alex Urmeneta and Margarida Romero. 2025. AI as a creative partner: a PRISMA review of AI's role in supporting creativity in education. In *Frontiers in Education*, Vol. 10. Frontiers Media SA, 1602151. doi:10.3389/feduc.2025.1602151

[87] Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. 2024. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour* 8, 12 (2024), 2293–2303. doi:10.1038/s41562-024-02024-1

[88] KURT VanLEHN. 2011. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist* 46, 4 (2011), 197–221. doi:10.1080/00461520.2011.611369

[89] Lev S Vygotsky. 1978. *Mind in society: The development of higher psychological processes.* Vol. 86. Harvard university press.

[90] Xuena Wang, Xueting Li, Zi Yin, Yue Wu, and Jia Liu. 2023. Emotional intelligence of Large Language Models. *Journal of Pacific Rim Psychology* 17 (2023). doi:10.1177/18344909231213958

[91] Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, and Wenqiang Zhang. 2022. A systematic review on affective computing: emotion models, databases, and recent advances. *Information Fusion* 83-84 (2022), 19–52. doi:10.1016/j.inffus.2022.03.009

[92] Bernard Weiner. 1985. An attributional theory of achievement motivation and emotion. *Psychological review* 92, 4 (1985), 548.

[93] Ben Williamson and Rebecca Eynon. 2020. Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology* 45, 3 (2020), 223–235. doi:10.1080/17439884.2020.1798995

[94] Merlin C Wittrock. 1974. Learning as a generative process 1. *Educational Psychologist* 11, 2 (Nov. 1974), 87–95. doi:10.1080/00461527409529129

[95] David Wood, Jerome S Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Journal of child psychology and psychiatry* 17, 2 (1976), 89–100.

[96] Beverly Park Woolf. 2009. *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning.* Morgan Kaufmann.

[97] Benny Hin Wai Yung and Ping Kee Tao. 2004. Advancing Pupils within the Motivational Zone of Proximal Development: A Case Study in Science Teaching. *Research in Science Education* 34, 4 (Dec. 2004), 403–426. doi:10.1007/s11165-004-0286-7

[98] Olaf Zawacki-Richter, Victoria I Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International journal of educational technology in higher education* 16, 1 (2019), 1–27. doi:10.1186/s41239-019-0171-0

[99] Chengliang Zhai, Santoso Wibowo, and Ling David Li. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments* 11 (2024), 28. doi:10.1186/s40561-024-00316-7