A Multi-Theoretical Pedagogical Architecture for Adaptive AI Tutoring in Special Education: Mapping Platform Features to Established Research in Learning Sciences, Educational Psychology, and Disability Studies

Travis Gilly[1]
[1]Real Safety AI Foundation, Cary, Illinois, United States

Intelligent tutoring systems built on large language models now demonstrate learning gains comparable to, or exceeding, in-person instruction in general education settings. These systems rarely address the pedagogical, legal, and psychosocial requirements of students with documented learning needs, which in the United States are governed by the Individuals with Disabilities Education Improvement Act of 2004 and parallel frameworks in other jurisdictions. This article presents a theoretical mapping of the Teacher in the Loop platform, a design-stage adaptive tutoring system built specifically for students receiving legally mandated accommodations, to the established literatures that inform each of its published features. Platform features are grouped into five theoretical clusters: cognitive and learning theory, disability studies and accommodation psychology, affective and behavioral sensing, academic integrity reframed through authorship, and institutional design with human oversight. For each feature the article identifies the seminal tradition on which the feature draws, the recent empirical work that validates the feature or a closely related mechanism, and the specific claims that remain theoretical pending deployment. The contribution is synthesis rather than efficacy demonstration: a resource that allows designers, researchers, and policy makers to see whether a platform built for students with disabilities has actually been grounded in the literatures that govern how those students learn, and where the remaining evidentiary gaps lie. Limitations, including the absence of deployment data and the novelty of population-specific adaptive content strategies, are discussed.

Keywords: artificial intelligence, special education, adaptive learning, Universal Design for Learning, zone of proximal development, productive failure, accommodation, human-in-the-loop

1 Introduction

Recent randomized controlled evidence has begun to settle a question that dominated educational technology discussions in the early 2020s. AI tutoring systems, when designed around established pedagogical principles, produce learning gains at or above those of well-implemented in-person instruction in authentic educational settings (Kestin, Miller, Klales, Milbourne, & Ponti, 2025). These gains hold even when the comparison is to an active-learning classroom rather than a lecture. A separate randomized trial of AI that coaches human tutors in real time found similar patterns, with the largest benefits accruing to less-experienced tutors serving students who were academically behind (Wang, Ribeiro, Robinson, Loeb, & Demszky, 2024).

These findings create a problem for the special education field. The students for whom AI tutoring could plausibly provide the largest marginal benefit, those who receive legally mandated accommodations under

Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Cary, Illinois. Email: t.gilly@ai-literacy-labs.org. See the Acknowledgments and Conflict of Interest Statement sections at the end of this article for disclosures regarding assistive technology use and commercial interest.

the Individuals with Disabilities Education Improvement Act (U.S. Congress, 2004) or parallel frameworks abroad, are rarely the population these systems were designed for. Most adaptive tutoring platforms treat accommodation delivery as an administrative feature rather than a pedagogical commitment. Invisible delivery, machine-verifiable audit trails, differential content-adaptation strategies for heterogeneous neurodivergent populations, and contractual protections for the human teachers who anchor the pedagogy are typically absent from commercial offerings.

The Teacher in the Loop platform (hereafter TITL) was filed as a provisional patent in October 2025 as an adaptive AI tutoring system built from first principles for students with documented learning needs. The platform is currently at design stage, with no deployment data. This article is therefore a synthesis rather than an efficacy study. The question it answers is narrower and, in the author's view, prior to any efficacy question: for each of the platform's published features, what is the pedagogical or disability-studies literature that supports the feature, what is the recent empirical work that validates either the feature itself or a closely related mechanism, and which features remain theoretical until deployment?

This question matters because design-stage platforms that claim evidence-based grounding frequently rely on weak chains of theoretical justification that do not survive close reading. A platform that claims to operate within a student's zone of proximal development (Vygotsky, 1978) must actually implement a mechanism consistent with that claim, and the mechanism must be distinguishable from the adaptive-difficulty baseline common to every intelligent tutoring system since the 1980s. A platform that claims to remove accommodation stigma must address a specific literature on why students under-use their legally available accommodations (Aquino & Bittinger, 2019). A platform that claims to protect teacher agency must be mappable to the teacher-in-the-loop framework as articulated in the European Commission Joint Research Centre report and related work (U.S. Department of Education, Office of Educational Technology, 2023; Fajardo-Ramos, Chiappe, & Mella-Norambuena, 2025).

The present article carries out this mapping across the published features of TITL, organized into three groups. Twelve pedagogical features currently implemented or in active development are treated in the main synthesis. Five features listed in the platform's public documentation but not yet implemented are treated in a dedicated aspirational features section, with the literature that would ground each when built. Five structural commitments (hybrid privacy architecture, locally deployable model option, contractual anti-displacement clause, compliance-as-byproduct, and default opt-out for non-essential sensing) are treated in a separate section as institutional preconditions under which the pedagogical features can operate without triggering the implementation failures that have characterized prior special-education technology rollouts. Grouping the features in this way allows the reader to distinguish what the platform does from what the platform plans to do, and to weigh each claim against the research base appropriate to its implementation status.

1.1 Contribution and Scope

The contribution is a feature-to-literature mapping that allows three distinct audiences to evaluate the platform on terms they care about. For pedagogical researchers, the article identifies which TITL features rest on well-supported theoretical and empirical foundations, which rest on novel combinations of established elements, and which are genuinely theoretical and await empirical validation. For special education practitioners, the article identifies the literatures from which each feature derives, allowing practitioners to assess fit with their own professional judgment. For policy makers and funders evaluating AI tutoring investments under the Individuals with Disabilities Education Improvement Act compliance framework, the article identifies the legal and research basis for each design choice.

The article does not claim that TITL is superior to alternative platforms. It does not present efficacy data, which do not yet exist. It does not resolve empirical questions that the platform's differential content-adaptation strategy for autism spectrum and attention-deficit populations is designed to test. The article does claim that the platform's feature set, taken as a whole, represents a defensible synthesis of available research, and that the gaps in that synthesis are identifiable and acknowledgeable rather than hidden.

2 Methodology

The mapping presented here is a narrative synthesis rather than a systematic review. Each published feature of TITL, as presented on the platform's public documentation at the Real Safety AI Foundation website, was traced back through three layers. First, the feature was restated in the language of the pedagogical or disability-studies tradition that most directly addresses it. Second, the seminal theoretical source or sources on which that tradition rests were identified. Third, recent empirical work from 2023 through the first quarter of 2026 was surveyed to identify validation, non-validation, or cautionary findings relevant to the feature.

Citations included in this article were subjected to a three-source verification protocol before inclusion. Where a citation could be independently confirmed across at least two sources during drafting, it is treated as verified. Where a citation is plausible and grounded in the established literature but could not be independently re-verified during drafting, it is flagged in the accompanying bibliography as needing verification before final submission. Where a citation carries a known discrepancy (for example, a recent preprint that has not yet been peer-reviewed), the caveat is stated in the text.

The article deliberately errs toward conservatism in its claims. Where the mapping between feature and literature is clean, this is stated directly. Where the mapping is theoretically supported but empirically thin, the gap is identified rather than papered over. Where a feature does not belong in a pedagogical synthesis, the article says so and addresses the feature elsewhere.

3 Cognitive and Learning Theory Foundations

3.1 Socratic Teaching Engine with Struggle-Aware Routing

The core instructional engine of TITL never provides direct answers to student queries. All instruction proceeds through strategic questioning, analogies, counterexamples, and scaffolded hints calibrated to the student's current understanding. This design draws on two traditions that have converged in recent AI tutoring research.

The first tradition is the Socratic and constructivist line that runs from Bloom's two-sigma problem (Bloom, 1984) through the human tutoring analyses of Chi, Siler, Jeong, Yamauchi, and Hausmann (2001) and the intelligent tutoring meta-analyses of VanLehn (2011). Bloom's original finding, that one-to-one tutoring produced learning gains approximately two standard deviations above conventional classroom instruction, motivated four decades of attempts to approximate individual tutoring at scale. The core mechanism, repeatedly identified in this literature, concerns the instructor's moment-to-moment calibration of question, hint, and feedback to the individual learner's current state rather than the content being delivered. Systems that fail to implement this calibration, regardless of content quality, consistently underperform.

The second tradition is the post-2022 empirical validation of this mechanism in large language model tutors. Kestin et al. (2025) conducted a randomized controlled trial in an undergraduate physics course at Harvard, comparing a custom-designed AI tutor (TeachGPT) built on explicit pedagogical best practices to a well-implemented active learning classroom. Students using the AI tutor learned significantly more in less time, reported higher engagement, and reported higher motivation. The design of the tutor is the critical point. Kestin et al. (2025) did not deploy a generic conversational model. They engineered a tutor around the same pedagogical best practices that informed their in-class lessons, including growth mindset framing, cognitive load management, critical thinking support, and personalized feedback. The finding is therefore conditional: AI tutoring engineered around established pedagogical principles works, which is a narrower claim than "AI tutoring works."

TITL's Socratic engine operationalizes this finding in a special education context. The engine does not simply refuse to give direct answers. It calibrates question difficulty and scaffolding type to the student's current state, using non-invasive behavioral signals rather than performance on high-stakes items. The engine's design also draws on productive failure research, discussed in Section 3.3, which argues that premature scaffolding can suppress the generative struggle that produces deep conceptual change (Kapur, 2008, 2016).

The mapping strength here is high. The theoretical foundation is canonical, the empirical foundation is a rigorously designed 2025 RCT, and the specific operationalization (Socratic engine calibrated to individual state) is the mechanism Kestin et al. (2025) explicitly identified as responsible for the observed gains.

The novelty of TITL lies not in the engine itself (Socratic tutoring is prior art across most major AI tutoring platforms) but in the special-education-first design constraints imposed on the engine.

3.2 Dynamic Cognitive Load Adjustment

TITL adjusts difficulty and pacing continuously based on the student's current cognitive state, with the explicit goal of maintaining the student within the zone of proximal development. The theoretical foundation for this feature draws on two distinct traditions that are frequently conflated in educational technology discourse.

The first is Vygotsky's zone of proximal development, defined as the cognitive space between what a learner can do independently and what the learner can do with assistance from a more capable peer or instructor (Vygotsky, 1978). The second is the scaffolding metaphor introduced by Wood, Bruner, and Ross (1976), which Vygotsky himself never used. The distinction matters for a pedagogical mapping article because the two concepts, though complementary, describe different mechanisms. The zone of proximal development describes a region; scaffolding describes the support structures that allow a learner to traverse that region. A platform that claims to operate within a student's zone of proximal development must specify both the region (how it is detected) and the scaffolding (how the support adjusts).

The third theoretical contribution comes from cognitive load theory as developed by Sweller (1988) and extended in Sweller, van Merrienboer, and Paas (1998) and Paas and van Merrienboer (2020). Cognitive load theory distinguishes intrinsic load (the inherent difficulty of the material), extraneous load (the difficulty added by the instructional presentation), and germane load (the productive effort of schema construction). Adaptive systems that reduce only intrinsic load produce easier learning experiences but not necessarily better learning. Systems that reduce extraneous load while preserving intrinsic and germane load produce better learning outcomes at similar or lower cognitive cost. TITL's adjustment logic is designed to reduce extraneous load (format, presentation, distracting elements) while preserving intrinsic load (the actual cognitive challenge of the material).

Recent empirical work supports this distinction in AI tutoring contexts. Chung, Zhang, Kung, Bastani, and Bastani (2026), in a preprint reporting on an LLM-guided reinforcement learning approach to personalized problem sequencing, found that personalized difficulty sequences produced post-test gains approximately equivalent to six to nine months of additional schooling relative to fixed difficulty sequences, in a five-month intervention. The preprint status of this work requires caution in how strongly the finding is claimed. The direction of the effect, however, is consistent with a large body of prior work on adaptive difficulty in intelligent tutoring systems.

The mapping strength for this feature is high on the theoretical side and moderate on the empirical side. The theoretical foundations are canonical. The empirical validation in AI-specific contexts is recent and in some cases still preprint. The TITL implementation is consistent with both layers but has not itself been validated.

3.3 Productive Struggle versus Destructive Frustration Routing

TITL distinguishes between two categories of student struggle and routes them differently. Productive struggle, in which the student is engaged, making progress, and generating intuitive approaches to a problem, is supported with minimal intervention. Destructive frustration, in which the student is disengaging, spinning, or showing signs of shutdown, triggers a three-tier intervention cascade: AI hint reframing, then (in future versions) peer matching, and finally direct teacher notification with full context of the student's attempts.

This feature draws on productive failure research as developed by Kapur (2008); Kapur and Bielaczyc (2012); Kapur (2016, 2024). The core finding of productive failure research is that students who collaboratively attempt to solve complex novel problems without instructional support before receiving instruction develop deeper conceptual understanding than students who receive the same instruction first. The mechanism at work concerns the activation of prior knowledge and the generation of intuitive solution ideas during the failure phase, which prepares the learner to assimilate canonical instruction more deeply when it arrives. Failure itself is the setting; prior knowledge activation is the engine. The assistance dilemma, articulated by Koedinger and Aleven (2007), describes the related question of when to give and when to withhold instructional support. Premature support suppresses productive failure; delayed support allows destructive frustration.

TITL's routing is designed to solve the assistance dilemma in a special education context, where the cost of destructive frustration is higher than in general education populations. Students receiving accommodations have, on average, lower tolerance for unresolved cognitive load and lower confidence in their own problem-solving capacity. An intervention system that treats all struggle identically will produce either chronic over-scaffolding (which suppresses learning) or chronic under-scaffolding (which produces shutdown). The three-tier cascade attempts to detect the transition from productive to destructive in real time, using behavioral signals, and escalate support in graduated rather than binary form.

The theoretical foundation of this feature is strong. The specific operationalization (three-tier cascade, particular behavioral signals, specific thresholds) is novel and has not been empirically validated. The feature is therefore presented as theoretically grounded but empirically pending.

4 Disability Studies and Accommodation Psychology

4.1 Invisible Accommodation Delivery

When a TITL student logs in, the platform automatically configures their learning environment to deliver accommodations specified in the student's individualized education program, Section 504 plan, or jurisdictionally equivalent document, without any visible indication to the student or their peers. Extended time, modified presentation, alternative response formats, reading supports, and other accommodations activate as system-level defaults rather than as student-triggered actions.

This feature responds to one of the most robust findings in disability studies. Students with documented accommodations frequently decline to use them, and this non-use is particularly pronounced in adolescence. The driver is stigma rather than ignorance or access. Aquino and Bittinger (2019) documented that a substantial proportion of students with legally established disability status decline to self-identify in higher education contexts, citing social cost as a primary reason. Toutain (2019) reviewed barriers to accommodation use in higher education and identified stigma and visibility as among the most significant. Clouder et al. (2020) examined neurodiversity in higher education more broadly and found that neurodivergent students consistently report accommodation-related social cost as a barrier to help-seeking.

The theoretical framework within which invisible delivery most naturally fits is Universal Design for Learning as developed by Meyer, Rose, and Gordon (2014) and extended in the Universal Design for Learning Guidelines version 2.2 (CAST, 2018). UDL argues that variability in how students perceive, engage with, and demonstrate learning is the norm rather than the exception, and that learning environments should be designed to accommodate this variability from the outset rather than retrofitted for individual students. Burgstahler (2015) extended this argument to higher education and institutional practice. UDL has historically been implemented at the curricular level: multiple modalities of representation, multiple means of engagement, multiple means of expression. TITL extends UDL to the system level by making accommodation delivery invisible and automatic rather than requested and visible.

This extension is the platform's strongest novel contribution in the disability studies space. UDL as a framework has typically stopped short of recommending that individual accommodations be delivered without the student's explicit engagement, partly because of concerns about student autonomy and informed consent in traditional educational contexts. In an AI-mediated context, where system-level defaults can be configured transparently in coordination with the student, family, and individualized education program team, the autonomy concern is reframed rather than eliminated. The student retains the right to know which accommodations are active; the student loses only the social cost of being visibly marked as an accommodation recipient.

A recent 2026 Frontiers in Child and Adolescent Psychiatry paper (Marras & Pasqualotto, 2026) raises a concern relevant to this design. The authors, reporting a qualitative study of teachers' professional reasoning about assistive technology and AI in inclusive classrooms, note that adaptive and generative AI systems may blur the boundaries between individualized accommodation and universal design, and that algorithmic inference intersects with pre-existing classificatory systems in ways that can amplify categorization effects. The concern is serious and should be addressed directly rather than deflected. TITL's invisible delivery architecture does not eliminate the classification; the classification persists in the individualized education program and in the platform's delivery logic. What the architecture eliminates is the peer-visible signal of that classification during instruction. This is a narrower claim than full de-categorization and is, in the author's view, the defensible claim.

The mapping strength for this feature is very high. Accommodation non-use due to stigma is among the best-documented findings in disability studies. UDL is a mature theoretical framework with broad empirical support. The specific extension to system-level invisible delivery is novel but traceable to both literatures.

4.2 Differential Content Adaptation for Heterogeneous Neurodivergent Populations

TITL implements different content-adaptation strategies for students with autism spectrum presentations and students with attention-deficit presentations. The design logic is that autism spectrum presentations, particularly those characterized by strong preferences for predictability and sameness, are better served by content that maintains consistent structure across sessions, while attention-deficit presentations are better served by varied content and presentation modalities that sustain engagement through novelty.

The theoretical foundation draws on executive function research as articulated by Barkley (1997) and extended in Diamond (2013), which frames attention-deficit hyperactivity disorder presentations as a set of executive function deficits affecting inhibition, sustained attention, and working memory. It also draws on research on autism spectrum presentations as characterized by different sensory processing profiles and stronger systemizing preferences (Baron-Cohen, 2009). The two profiles, while overlapping on many dimensions, have distinct implications for instructional design.

Recent neurophysiological work provides cautionary context. Le Cunff, Giampietro, and Dommett (2024) conducted a scoping review of cognitive load measurement in attention-deficit, autism spectrum, and dyslexia populations and found that the measurement techniques themselves (electroencephalography, functional magnetic resonance imaging, eye tracking) have significant methodological limitations in neurodivergent populations, including pupil foreshortening error in eye tracking and motion artifact in imaging. The implication for TITL is that any automated content-adaptation system that infers cognitive load from behavioral signals will inherit the measurement limitations of its underlying signal layer.

The mapping strength for this feature is the weakest in the paper, and the article presents it as such. The underlying neuroscience supports differential profiles. The specific operationalization, a single platform that switches content-adaptation strategies based on the student's dominant profile, does not yet have direct empirical validation. The feature is presented in this article as a theoretically grounded hypothesis that the platform is designed to test, not as an already-validated design choice. A reviewer would be correct to push back on any stronger claim, and the platform's public documentation should be consistent with the theoretical-pending-empirical framing rather than overclaiming.

5 Affective and Behavioral Sensing

5.1 Multi-Modal State Detection

TITL infers student cognitive and affective state from multiple behavioral signals: typing rhythm, mouse movement patterns, pause patterns, and, where students and guardians have opted in, facial expression analysis. The platform does not require any biometric sensor beyond a standard laptop or Chromebook camera, and the platform's default configuration operates entirely without camera input.

The theoretical foundation draws on affective computing as founded by Picard (1997) and extended in a substantial literature on learner affect and engagement detection. S. D'Mello and Graesser (2012) and S. K. D'Mello (2013) analyzed the dynamics and relative incidence of discrete affective states during complex learning. Baker, D'Mello, Rodrigo, and Graesser (2010) identified that learners spend substantial portions of time in non-engaged states during interaction with computer-based learning environments, and that the persistence and impact of those states varies by specific affective state. The finding that bears most directly on TITL's design is that frustration, counterintuitively, is often more productive than boredom, but only within specific bounds.

Keystroke dynamics as a signal for cognitive state has a mature empirical foundation. Epp, Lippold, and Mandryk (2011) demonstrated that keystroke dynamics could identify emotional states in working contexts. More recent work has extended this to cognitive fatigue and load detection, with reported accuracies in the range of 72 to 91 percent depending on the specific classification task. The signals are passively captured, require no additional hardware, and operate without raising the privacy concerns associated with camera-based affect detection.

Camera-based facial analysis is more contested. Pupil dilation measures are confounded by luminance and gaze position (Le Cunff et al., 2024). Facial expression classifiers trained on non-representative populations exhibit known accuracy degradation on children, non-white subjects, and neurodivergent populations. TITL's design default is therefore camera-free operation, with facial analysis available only as an explicit opt-in with consent. The article recommends that the platform's documentation consistently foreground the passive-signal layer rather than the camera layer, both for privacy reasons and for accuracy reasons.

The mapping strength for the passive-signal layer is strong. The mapping strength for the camera layer is weaker and should be treated as optional rather than core.

5.2 Behavioral Biometric Verification

TITL includes a behavioral biometric verification layer that provides continuous passive authentication during instructional sessions. The purpose of the layer is to support academic integrity claims (distinguishing authentic student work from work produced by another person) without introducing intrusive proctoring or keystroke logging that would raise privacy and disability law concerns.

The theoretical foundation draws on behavioral biometric authentication research. A recent 2026 paper by Dutta, Roy, and Roy (2026) proposes a quantum kernel anomaly detection framework for behavioral biometric authentication that explicitly addresses the problem of intra-user temporal drift arising from fatigue, cognitive load, device handling, and contextual variation. The article notes that behavioral biometric signals are non-stationary and should be modeled as one-class anomaly detection problems rather than fixed-distribution classification tasks. This finding is directly relevant to the special education context, where intra-session and across-session variation in student signals is frequently greater than in neurotypical populations, and where a biometric system that cannot tolerate this drift will produce high false-rejection rates for the students the platform is designed to serve.

The mapping strength here is strong on the technical foundation, moderate on the pedagogical application. The feature is primarily an infrastructural protection against identity substitution in AI-assisted assessment contexts. Its pedagogical relevance is indirect: it supports the larger claim that AI-assisted work can be verifiably authentic, which in turn reduces the pressure to ban AI assistance outright in special education contexts where AI is a legally meaningful accommodation.

6 Academic Integrity Reframed Through Authorship

6.1 Provenance Tracking

Every AI-generated suggestion in TITL is marked and tracked. Teachers and parents receive a complete record of which suggestions were offered, which were accepted, and which were rejected. The purpose is auditability rather than surveillance: AI assistance becomes a documented component of the student's learning process, particularly in contexts where AI assistance is a legally established accommodation.

The theoretical foundation here is thinner than in other sections because the specific problem, provenance of AI-generated content in student work, is new enough that the literature is still forming. The closest established tradition is cognitive writing process research as developed by Flower and Hayes (1981), which frames writing as a recursive, planning-translating-reviewing process rather than a linear output operation. If writing is a process, then the traces of that process (drafts, revisions, suggestions accepted and rejected) are a legitimate artifact of the work, and AI-generated suggestions integrated into that process are a component of the artifact rather than a contamination of it.

This reframing matters for the special education case specifically. Students with writing-related accommodations frequently use assistive technology as part of their legally mandated supports. Traditional academic integrity frameworks, which tend to treat AI use as binary (present or absent, permitted or prohibited), cannot accommodate this reality. Provenance tracking allows a more granular record: the student used AI in these specific ways, the AI made these specific suggestions, the student accepted some and rejected others, the teacher reviewed the pattern. This is the student's own proof of authorship rather than the institution's proof of non-cheating.

The mapping strength for the underlying theoretical claim (writing as process, traces as artifact) is high. The mapping strength for the specific operationalization in an AI assistance context is emerging rather than established.

6.2 Process Capture as Student's Proof of Authorship

Closely related to provenance tracking, TITL captures the student's revision process: drafts, edits, time between revisions, and the sequence of changes across a session. The framing is deliberately inverted from typical surveillance-based academic integrity tools. The process capture does not exist to catch the student in dishonesty. It exists to provide the student with evidence that the work is genuinely theirs when questioned.

This inversion is one of the platform's most defensible pedagogical moves. Surveillance-based integrity systems place the burden of proof on the student in adversarial framing: the student must not appear to have used prohibited assistance. Process-capture-as-proof places the evidence in the student's hands in supportive framing: the student has documentary proof of their own engagement with the work. The pedagogical function is the same (establishing authorship), but the psychological and relational function is different.

The theoretical support here is indirect. Flower and Hayes' process model of writing (Flower & Hayes, 1981) supports the claim that the process is a legitimate artifact. Research on assessment and student agency supports the claim that students who have evidence of their own effort exhibit higher academic self-efficacy. The specific reframing from surveillance to empowerment has not, to the author's knowledge, been empirically tested. The feature is presented as theoretically coherent and normatively defensible rather than empirically validated.

7 Institutional Design and Human-in-the-Loop Architecture

7.1 Complete Teacher Control and the Teacher Dashboard

TITL is not a replacement for a teacher. The platform's central governance principle is that the human educator retains pedagogical authority at every level of the system, with the AI acting as an assistive layer that extends rather than substitutes for the teacher's judgment. The teacher dashboard provides real-time visibility into student activity, struggle classification, intervention history, and accommodation delivery.

The theoretical and policy foundation for this design is explicit and well-developed. The U.S. Department of Education's Office of Educational Technology report (U.S. Department of Education, Office of Educational Technology, 2023) identifies keeping humans in the loop as a top policy priority in educational AI, grounded in the Blueprint for an AI Bill of Rights and in educational tradition. Fajardo-Ramos et al. (2025), in a 2025 Frontiers in Education paper on human-in-the-loop assessment with AI in teacher education across Ibero-American universities, documents the practical implementation of this principle. Marras and Pasqualotto (2026), in the 2026 Frontiers paper already cited, examines teachers' professional reasoning about assistive technology and AI in inclusive classrooms specifically, and finds that teachers' interpretive work is the critical variable in whether AI-based systems produce inclusive outcomes or categorization effects.

The 2026 Marras and Pasqualotto study raises concerns that TITL must address directly rather than ignore. The authors note that even minor algorithmic misclassifications can have disproportionate consequences for students' autonomy, perceived competence, and trajectories of inclusion. The mitigation in TITL is architectural: the AI operates as a suggestion layer rather than a decision layer, teachers receive full context on each algorithmic recommendation, and teachers retain authority to override any system-generated classification. This does not eliminate the concern. It constrains the concern to the zone where teacher judgment is available as a correction mechanism.

The mapping strength is very high. Teacher-in-the-loop is literally the named approach in the European Commission Joint Research Centre framework, the U.S. Department of Education report, and the Fajardo-Ramos et al. and Marras and Pasqualotto papers.

7.2 Three-Tier Escalation with Full Context Handoff

When the platform detects that a student has crossed from productive struggle into destructive frustration, it initiates a three-tier escalation. Tier one is AI-generated hint reframing. Tier two, planned for future versions, is peer matching with a classmate who has recently mastered the relevant material. Tier three is direct teacher notification with full context: what the student has attempted, what signals triggered the escalation, and what interventions have already been tried.

The empirical foundation for AI-assisted escalation to human support is now available. Wang et al. (2024), in a randomized controlled trial involving approximately 900 tutors and 1,800 elementary and secondary school students, found that AI coaching of human tutors produced four percentage point gains in student success and nine percentage point gains for students working with lower-rated tutors. The cost of the AI coaching layer was approximately \$20 per tutor per year. The Wang et al. design differs from TITL in important respects. Wang et al. placed AI in a coaching role for human tutors who were already present; TITL uses AI as the first-tier responder with human teachers as escalation targets. The common finding, however, is that AI-human collaboration structures in tutoring contexts produce gains larger than either human or AI alone would produce.

The mapping strength is strong. The exact TITL cascade has not been tested, but the general pattern of AI-human collaboration with context handoff has empirical validation.

7.3 Parent Portal with Configurable Engagement

TITL provides parents with a dedicated portal showing their child's daily instructional activity, accommodation delivery verification, and progress toward individualized education program goals. Engagement frequency is parent-controlled rather than platform-pushed. The design is grounded in specific statutory language.

The Individuals with Disabilities Education Improvement Act of 2004 (U.S. Congress, 2004) permits alternative progress reporting methods with parental consent. This provision, frequently underutilized, transforms parent engagement from a convenience feature into a compliance mechanism that simultaneously serves the parent's right to information and the district's obligation to report. The portal therefore functions in two legal registers at once, and the dual function is the platform's contribution on this feature rather than a novel pedagogical insight.

The broader empirical foundation for parent engagement in special education outcomes is extensive and will not be recapitulated here. The mapping strength of the specific feature (a portal grounded in alternative progress reporting provisions) is high on the legal side and sufficient on the pedagogical side.

8 Aspirational Features with Established Literature

The following features appear in TITL's public documentation but have not yet been implemented. They are included here because the pedagogical traditions that would justify their implementation are well-established, and presenting them alongside the built features allows the reader to distinguish the current platform from the planned platform. Each feature in this section is labeled as aspirational and is mapped to the literature that would ground it when built.

8.1 Peer Tutoring Matching (Aspirational)

The platform's documentation describes an AI-optimized peer tutoring matching feature that pairs students based on knowledge complementarity and social compatibility, with the explicit intent of activating the gains that accrue to the student in the tutor role. This feature has not been implemented.

The pedagogical tradition it would draw on is learning-by-teaching research, sometimes termed the protégé effect. Fiorella and Mayer (2013) conducted two experiments demonstrating that students who prepared materials with the expectation of teaching, and then actually taught, outperformed students who prepared only for a test, with the learning advantage strongest on delayed comprehension measures. The effect has been replicated across multiple populations and content domains, with meta-analytic estimates converging on small-to-medium effect sizes for teaching expectancy alone and medium effects for actual teaching (Fiorella & Mayer, 2013). The mechanism most strongly supported in the literature concerns generative processing: students who prepare to teach organize material more coherently and integrate it more deeply with prior knowledge than students who prepare for recognition or recall.

In a special education context, the peer tutoring feature would need to address two concerns not present in general education research. First, accommodation needs differ across paired students, and matching on knowledge complementarity alone may produce pairs in which one student's accommodation needs are inadvertently exposed to the partner. Second, the social cost findings in Aquino and Bittinger (2019) suggest that visibility as a tutor or as a tutee carries different stigma implications for neurodivergent students than for neurotypical populations. A responsible implementation would therefore need to integrate peer matching with the invisible accommodation delivery architecture described in Section 4.1, and would need to provide explicit teacher review of all proposed pairings before activation. The feature is aspirational and not built.

8.2 Bayesian Knowledge Tracing with Forgetting Curves (Aspirational)

The platform's documentation describes a feature that models probabilistic mastery of each skill component and schedules review based on estimated knowledge decay. This feature has not been implemented.

The theoretical foundation is strong and long-established. Bayesian knowledge tracing as a formal mechanism for modeling a student's probability of having learned a skill, given a sequence of observed performances, was introduced by Corbett and Anderson (1995) in the context of the ACT Programming Tutor. Their formulation uses four parameters per skill (initial probability of knowledge, probability of learning on each opportunity, probability of slip on a known skill, probability of guess on an unknown skill) and updates the probability estimate after each student response. The formulation has been extended, refined, and ensembled in the three decades since, but the core Corbett and Anderson model remains the reference point in the intelligent tutoring systems literature.

Forgetting curves, the related construct, trace back to Ebbinghaus (1885) and have been replicated across modern laboratory and classroom conditions. The integration of knowledge tracing with forgetting dynamics produces a schedule in which review is presented not on a fixed calendar but at the point where the estimated retention probability for a given skill approaches a threshold below which reinstatement becomes inefficient.

For TITL, the justification for building this feature concerns the specific profile of students with documented learning differences. Attention-deficit populations exhibit consolidation patterns that differ from neurotypical populations, and autism spectrum populations benefit from explicit review structures that reduce uncertainty about what will come next (Barkley, 1997; Baron-Cohen, 2009). A knowledge tracing system that schedules review based on each student's own decay curve, rather than a one-size-fits-all interval, would address both of these profiles more directly than fixed-schedule spaced repetition. The feature is aspirational and not built. When built, the implementation should be validated in the target population rather than transferred from general education tutoring studies.

8.3 Explainable AI with Pedagogical Reasoning (Aspirational)

The platform's documentation describes a feature in which every AI decision includes an educational rationale, with explanations tailored to the stakeholder (student, teacher, parent) receiving them. This feature has not been implemented.

The theoretical foundation for this feature is the XAI in education literature, most comprehensively surveyed in Khosravi et al. (2022). The authors argue that explainability in educational AI has needs distinct from explainability in other AI application domains, because educational decisions affect not only the learner's immediate experience but the learner's developing sense of competence and agency. A recommendation that a student receive a particular problem next, or that a student's current struggle be routed to the teacher, or that a particular accommodation be activated, is a decision that the student, teacher, and parent may all reasonably want to understand. The XAI-ED framework proposed by Khosravi et al. (2022) identifies six aspects of explainability (stakeholders, benefits, presentation approaches, AI model classes, human-centered designs, and potential pitfalls) and provides a structure for evaluating any particular educational AI system's explanation capability.

For TITL, the justification for building this feature is twofold. First, the Marras and Pasqualotto (2026) findings discussed in Section 4.1 indicate that teachers' discomfort with opaque AI recommendations is a primary barrier to responsible integration. Explanations that allow teachers to evaluate and contest recommendations make the human-in-the-loop commitment operational rather than aspirational. Second, the COPPA and FERPA frameworks governing student data require that parents receive understandable accounts of how automated systems make decisions affecting their children. An explainability layer is therefore not only a pedagogical feature but a compliance feature. The feature is aspirational and not built.

8.4 Collaboration Detection (Aspirational)

The platform's documentation mentions a collaboration detection capability under the academic integrity category. The feature's scope in the public documentation is narrow enough that its pedagogical grounding is uncertain. If the feature is intended to detect unauthorized collaboration during assessment, it is an integrity feature that does not require pedagogical grounding beyond what Section 6 already provides. If the feature is intended to detect and support authorized peer collaboration in learning, it would draw on the same peer-learning literature as the peer tutoring matching feature described above (Fiorella & Mayer, 2013), along with collaborative learning research in the cognitive load tradition. The author recommends that the platform's public documentation be revised to clarify which of these two uses is intended, because the two implementations carry different research justifications and different ethical considerations. The feature is aspirational and not built in either form.

8.5 Student Cognitive Profile Modeling (Aspirational)

The platform's documentation describes a feature that creates a portable cognitive model of each student, used to predict learning crises in advance and to follow the student across platforms and years. This feature has not been implemented.

The research basis for predictive modeling of student learning states is the same knowledge tracing and affective computing literature already cited in Sections 3 and 5 (Corbett & Anderson, 1995; S. D'Mello & Graesser, 2012). Modern extensions use ensemble methods and deep learning to improve prediction accuracy, but the core mechanism (inferring a latent student state from observed performance and behavioral signals) is the same as the mechanism described by Corbett and Anderson thirty years ago. The specific claim that the system can predict learning crises five to fifteen minutes in advance is an empirical claim about prediction horizon rather than a theoretical claim about modeling, and would need to be validated in the target population before it could be responsibly included in public documentation.

The portable profile claim is more ethically fraught than the predictive modeling claim. A cognitive profile that follows a student across platforms and years is, in the author's view, among the most personally identifying data that could be held about a minor, potentially more identifying than their name, address, or biometric records, because it reveals patterns of cognition and affect that the student may not consciously recognize in themselves. The Marras and Pasqualotto (2026) findings regarding amplification of categorization effects in neurodivergent populations apply with particular force to portable profiles. A responsible implementation would require (at minimum) a student's continued affirmative consent at each age of majority transition, parental consent below those thresholds, the ability to delete the profile at any time, and default limitations on profile portability outside the institution that collected the data. The feature is aspirational and not built. When built, the implementation should be governed by privacy architecture that exceeds rather than merely meets current COPPA and FERPA standards.

9 Structural Preconditions

Five additional features do not belong in a pedagogical synthesis but are institutional preconditions under which the pedagogical features can operate. They are noted briefly here to complete the architectural picture.

The hybrid privacy architecture maintains all personally identifiable information on school-controlled local infrastructure, with the cloud-based AI engine receiving only tokenized interaction data. This architecture draws on the privacy-by-design principles articulated by Cavoukian (2009) and addresses the privacy concerns documented in Kumar, Chetty, Clegg, and Vitak (2019) regarding elementary school technology use. A locally deployable model option, using a district-controlled server rather than cloud inference, is available for jurisdictions that require fully offline processing. These design choices make privacy a structural property of the system rather than a policy promise.

The contractual anti-displacement clause commits any implementing institution to not reduce teaching, aide, or special education personnel as a result of efficiency gains from the platform. The clause operates as labor protection rather than pedagogy, preserving the conditions under which the pedagogy can function. A platform that claims to augment teachers while enabling their replacement would undermine its own central premise. The clause makes the premise enforceable.

The audit trail produced by accommodation delivery logging, suggestion tracking, and teacher intervention records constitutes a compliance posture that exceeds current special education documentation standards. Compliance is treated throughout TITL as a byproduct rather than a product category. The platform does not market itself as a compliance tool; it markets itself as a pedagogical tool whose operation happens to generate the evidence districts need to satisfy regulatory requirements. This framing is architectural rather than pedagogical, but it matters for how districts should evaluate the platform against existing compliance-first competitors.

9.1 Default Opt-Out for Non-Essential Sensing and Data Collection

A fifth structural commitment organizes the platform's stance on features that collect information about the student beyond what is strictly necessary for instruction. The platform draws a line between features that are structurally required for the pedagogical architecture to function and features that enrich the architecture but are not required for its core operation. Features in the first category are on by default and operate as part of the platform's baseline behavior. Features in the second category are off by default and require explicit opt-in by the teacher, parent, or student as appropriate to the student's age and the feature's nature.

Structurally required features include accommodation delivery as specified in the individualized education program, the Socratic instructional engine, difficulty and pacing adjustment based on in-task performance signals, three-tier escalation routing from AI to teacher, the teacher dashboard, provenance tracking of AI contributions to student work, and the compliance audit trail. These features cannot be disabled without disabling the platform's core value proposition or violating legal obligations around accommodation delivery.

Features that enrich the architecture but are not required for its core operation include camera-based facial affect analysis, any biometric authentication modality beyond the minimum required to verify student identity for secure session handoff, and any data collection that leaves the local environment for research or product improvement purposes. These features are off by default. The parent portal includes explicit controls that allow parents to review, enable, or disable each optional feature at any time, with plain-language descriptions of what each feature does, what data it collects, where the data is stored, and what it is used for. Teachers can override parent-level settings only for structurally required features; optional features always require parent consent.

This arrangement inverts the default pattern in educational technology, which typically enables all sensing features by default and requires families to navigate opt-out workflows. The inversion is deliberate. It places the burden of configuration on the party that benefits from the additional data rather than on the party whose data is being collected. It also narrows the surface area for the kinds of categorization and surveillance concerns documented in Marras and Pasqualotto (2026) and Kumar et al. (2019), because the optional features, which carry the greatest privacy and equity risk, are the ones that require an affirmative act of consent rather than an affirmative act of refusal.

10 Discussion

Three observations emerge from this synthesis.

First, the pedagogical and disability studies foundations for an AI tutoring platform built specifically for students with documented learning needs are more mature than the field's current product offerings suggest. Every major feature of TITL except the differential content adaptation for heterogeneous neurodivergent populations rests on an established theoretical tradition with recent empirical validation. The gap between what the literatures support and what commercial platforms implement exists in design priorities rather than in the literatures themselves.

Second, the 2026 Marras and Pasqualotto study and related work make clear that human-in-the-loop functions as a concrete architectural commitment requiring enforcement through system design, and deteriorates into marketing language when stated only in documentation. TITL's teacher-first governance, full context handoffs, and contractual anti-displacement clause are mechanisms for keeping the principle operational under economic pressure. Platforms that implement teacher-in-the-loop in documentation but not in architecture should not be treated as equivalent to platforms that implement it in both.

Third, the differential content adaptation feature is the platform's central empirical hypothesis rather than a validated design choice. The platform should be honest about this in public documentation and in any grant applications that invoke the feature. A platform that presents a theoretical commitment as an established finding invites the reviewer backlash that has damaged other AI education initiatives. The cleaner move is to present the feature as a theoretically grounded hypothesis that the platform is designed to test, and to plan the empirical work that would constitute the test.

11 Limitations

The most significant limitation of this article is that TITL is at design stage with no deployment data. Every efficacy claim is therefore indirect, grounded in the empirical validation of closely related mechanisms in other platforms rather than in direct evidence from TITL itself. Future empirical work should prioritize a controlled deployment in a special education context, with accommodation delivery verification, teacher agency measures, and student outcome measures as primary endpoints.

The citation base of this article includes several entries flagged for verification in the accompanying bibliography. These entries are plausible and grounded in the established literature but were not independently reverified across three sources during drafting. A final submission to a peer-reviewed venue should complete this verification, and the bibliography file accompanying this manuscript marks each entry with its verification status.

The evidence base transfer from general education AI tutoring research (Kestin et al., 2025; Wang et al., 2024) to special education contexts should be treated as a hypothesis rather than an established extrapolation. The mechanisms are plausible and the design principles translate, but empirical validation in the target population is not available.

The differential content adaptation feature for autism spectrum and attention-deficit populations remains the paper's weakest mapping and should not be treated as empirically validated on the basis of this synthesis.

12 Conclusion

Adaptive AI tutoring for students with documented learning needs is theoretically tractable. The pedagogical traditions that support the core features are well-established, the recent empirical work in general education settings provides indirect support for the core mechanisms, and the disability studies literature provides clear guidance on what accommodation delivery must accomplish to actually serve the target population. What is missing from the field is a platform that has implemented these insights as architectural commitments rather than feature marketing. The Teacher in the Loop platform, as published at design stage, is an attempt at such a platform. Whether the attempt succeeds is an empirical question that deployment will settle. Whether the attempt is grounded in the literatures that should govern its design is the question this article has addressed. On that question, the mapping presented here suggests that the grounding is defensible across the majority of the feature set, theoretically supported but empirically pending in specific identified cases, and honestly acknowledged where it is weakest.

Acknowledgments

The author uses speech-to-text software and AI-assisted tools (Anthropic Claude) as assistive technology for organizing and formatting written output. These tools serve as accessibility accommodations for the author's neurodevelopmental conditions. All research

questions, analytical arguments, theoretical contributions, methodological decisions, and conclusions are solely the author's own.

Conflict of Interest Statement

The author is the founder and Executive Director of the Real Safety AI Foundation, an Illinois registered nonprofit with federal 501(c)(3) recognition pending. The author is the inventor named on a provisional patent filing (filed October 17, 2025) covering the Teacher in the Loop platform described throughout this article. The patent is held personally by the author and will be licensed without fee to the Foundation for its intended use, consistent with the Foundation's mission. The author currently draws no compensation from the Foundation because the Foundation is pre-revenue and pre-grant. This is a matter of current operational reality rather than a permanent arrangement; if the Foundation secures grant funding or platform-related revenue sufficient to support a staffed executive role, the author anticipates drawing a reasonable nonprofit-scale salary consistent with IRS guidance for similarly situated organizations. Favorable reception of this article is therefore connected, through the Foundation's funding pipeline, to the author's own potential future compensation, and readers should weigh the arguments presented here with that pathway in view rather than under the false impression of a permanently volunteer principal investigator. The author has attempted to mitigate this conflict through three methodological choices: presenting the article explicitly as a grounding synthesis rather than an efficacy study, identifying the weakest feature mappings and the gaps in the evidence base rather than concealing them, and acknowledging the 2026 Marras and Pasqualotto study as a concern that the platform's design must address rather than a critique to be deflected. These choices do not eliminate the conflict. The reader is entitled to weigh the author's arguments with full knowledge of the commercial, funding, and compensation pathway that favorable reception would enable.

13 References

Aquino, K. C., & Bittinger, J. D. (2019). The self-(un)identification of disability in higher education. *Journal of Postsecondary Education and Disability*, *32*(1), 5–19.

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94. doi: 10.1037/0033-2909.121.1.65

Baron-Cohen, S. (2009). Autism: The empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, *1156*, 68–80. doi: 10.1111/j.1749-6632.2009.04467.x

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

Burgstahler, S. E. (2015). *Universal design in higher education: From principles to practice* (2nd ed.). Harvard Education Press.

CAST. (2018). *Universal design for learning guidelines version 2.2.* https://udlguidelines.cast.org.

Cavoukian, A. (2009). *Privacy by design: The 7 foundational principles.* Information and Privacy Commissioner of Ontario.

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

Chung, A. T.-H., Zhang, B., Kung, L.-C., Bastani, H., & Bastani, O. (2026, March). *Effective personalized AI tutors via LLM-guided reinforcement learning.* Working paper, SSRN preprint. (Preprint; not yet peer-reviewed) doi: 10.2139/ssrn.6423358

Clouder, L., Karakus, M., Cinotti, A., Ferreyra, M. V., Amador Fierros, G., & Rojo, P. (2020). Neurodiversity in higher education: A narrative synthesis. *Higher Education*, *80*(4), 757–778. doi: 10.1007/s10734-020-00513-6

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278. doi: 10.1007/BF01099821

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. doi: 10.1146/annurev-psych-113011-143750

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. doi: 10.1016/j.learninstruc.2011.10.001

D'Mello, S. K. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, *105*(4), 1082–1099.

Dutta, S., Roy, S., & Roy, U. (2026). Quantum kernel anomaly detection: A fidelity-based framework for robust behavioral biometric authentication. *The Journal of Supercomputing*, *82*(3). doi: 10.1007/s11227-026-08300-3

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Leipzig: Duncker & Humblot. (English translation by H. A. Ruger and C. E. Bussenius, 1913, Teachers College, Columbia University)

Epp, C., Lippold, M., & Mandryk, R. L. (2011). Identifying emotional states using keystroke dynamics. In *Proceedings of the sigchi conference on human factors in computing systems (CHI '11)* (pp. 715–724). ACM. doi: 10.1145/1978942.1979046

Fajardo-Ramos, D. C., Chiappe, A., & Mella-Norambuena, J. (2025). Human-in-the-loop assessment with AI: Implications for teacher education in ibero-american universities. *Frontiers in Education*, *10*, 1710992.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(4), 281–288. doi: 10.1016/j.cedpsych.2013.06.001

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. doi: 10.58680/ccc198115885

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424. doi: 10.1080/07370000802212669

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299. doi: 10.1080/00461520.2016.1155457

Kapur, M. (2024). *Productive failure: Unlocking deeper learning through the science of failing*. Wiley. doi: 10.1002/9781394308712

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. doi: 10.1080/10508406.2011.591717

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025, June). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. doi: 10.1038/s41598-025-97652-6

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., … Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. doi: 10.1016/j.caeai.2022.100074

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264.

Kumar, P. C., Chetty, M., Clegg, T. L., & Vitak, J. (2019). Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI conference on human factors in computing systems (CHI '19)*. ACM.

Le Cunff, A.-L., Giampietro, V., & Dommett, E. (2024). Neurophysiological measures and correlates of cognitive load in ADHD, ASD and dyslexia: A scoping review and research recommendations. *European Journal of Neuroscience*, *59*(2), 256–282. doi: 10.1111/ejn.16201

Marras, A., & Pasqualotto, A. (2026). Negotiating assistive technologies and AI in inclusive education: Professional agency in neurodivergent contexts. *Frontiers in Child and Adolescent Psychiatry*, *5*, 1820276. doi: 10.3389/frcha.2026.1820276

Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal design for learning: Theory and practice*.

Wakefield, MA: CAST Professional Publishing.

Paas, F., & van Merrienboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, *29*(4). doi: 10.1177/0963721420922183

Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Toutain, C. (2019). Barriers to accommodations for students with disabilities in higher education: A literature review. *Journal of Postsecondary Education and Disability*, *32*(3), 297–310.

U.S. Congress. (2004). *Individuals with disabilities education improvement act of 2004.* 20 U.S.C. §1400 et seq.

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations.* Washington, DC.

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

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2024). *Tutor CoPilot: A human-AI approach for scaling real-time expertise* (EdWorkingPaper No. No. 24-1054). Annenberg Institute at Brown University. Retrieved from `https://edworkingpapers.com/ai24-1054` doi: 10.26300/81nh-8262

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x