

Development of an Adaptive AI Physics Tutor Based on a Large Language Model to Improve Scientific Literacy and Remediate High School Students' Misconceptions

Ramadani Khotibul Umam

Physics Education Study Program, Faculty of Education, Universitas KH Abdul Wahab Hasbullah, Jombang, Indonesia

Email: akunyumanto@gmail.com

Received : 15 August - 2026

Accepted : 15 September - 2026

Published online : 16 September - 2026

Abstract

The shift toward Society 5.0 and generative AI (Artificial Intelligence) has opened new opportunities for personalizing physics learning, yet Indonesian students' scientific literacy in PISA (Programme for International Student Assessment) 2022 remains far below the OECD (Organization for Economic Co-operation and Development) average, and misconceptions involving force, motion, energy, electricity, and waves remain resistant to conventional instruction. LLM (Large Language Model) chatbots such as ChatGPT, Gemini, and Claude are used ad hoc by students, without being designed to diagnose misconceptions, grounded in conceptual change theory, or equipped with PISA-aligned scientific literacy assessment. This article formulates a theoretical framework for an Adaptive AI Physics Tutor, a new-generation Intelligent Tutoring System integrating LLM architecture, Retrieval-Augmented Generation, a Misconception Detector, and a Student Model into one personalized, adaptive learning ecosystem. Using narrative review, critical review, concept analysis, and theory synthesis on 2021-2026 literature, it combines Conceptual Change theory, Interactive Engagement, Modeling Instruction, Vygotsky-Piaget constructivism, and TPACK (Technological Pedagogical Content Knowledge) into a conceptual model mapping student input, the AI engine, and learning output. The resulting model positions the AI Physics Tutor as a mediator of misconception remediation explicitly aligned with three PISA competencies, namely explaining phenomena scientifically, evaluating and designing scientific inquiry, and interpreting data and evidence, and offers a staged 2026-2045 development roadmap with a Design-Based Research agenda for prototype validation, along with theoretical and practical implications for curriculum developers, physics teachers, universities, and policymakers in Indonesia.

Keywords: Adaptive Learning, AI Physics Tutor, Intelligent Tutoring System, Large Language Model, Physics Misconceptions.

1. Introduction

Education in the third decade of the 21st century stands at a crossroads of transformation marked by the convergence of a super-smart society (Society 5.0), artificial intelligence (AI) based automation, and the acceleration of generative Large Language Model (LLM) capabilities. The concept of Society 5.0, introduced by Japan, places humans at the center of digital technology utilization to solve social problems, including gaps in education quality, and requires students to possess adequate digital literacy, data literacy, and scientific literacy in order to participate productively in an increasingly automated socio-technical ecosystem. At the same time, the pace of LLM development since the introduction of the Transformer architecture (Vaswani et al., 2025) and reinforced by Reinforcement Learning



from Human Feedback (RLHF) techniques (Ouyang et al., 2022) has given rise to generative models such as GPT, Gemini, Claude, Llama, DeepSeek, and Mistral that are capable of multi-step reasoning, explaining scientific concepts across various language registers, and interacting conversationally with users in a natural manner. This capability opens great opportunities for science education, including physics, to transition from a one-size-fits-all learning model toward personalized learning at scale (Kasneci et al., 2023; Yan et al., 2023).

However, this technological optimism confronts an empirical reality that is far bleaker at the level of Indonesian students' scientific literacy achievement. The results of the Programme for International Student Assessment (PISA) 2022, released by the Organization for Economic Co-operation and Development (OECD) on December 5, 2023, show Indonesia's scientific literacy score at 383, down 13 points from the 2018 PISA score of 396, and roughly 102 points behind the OECD country average of 485 (Kemendikdasmen, 2023; OECD, 2023). Indonesia ranked around 70th out of 81 participating countries in the science domain. Although Indonesia's relative ranking for scientific literacy is reported to have risen by about six positions compared to PISA 2018, an achievement interpreted by Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) in 2023 as evidence of the national education system's resilience in facing learning loss caused by the COVID-19 pandemic, this rise in ranking occurred alongside a decline in the absolute scores of nearly all participating countries due to the pandemic's global impact, so it cannot be interpreted as substantive improvement in scientific literacy quality in absolute terms. This fact confirms that Indonesia's scientific literacy problem is structural and chronic in nature, not merely a momentary anomaly caused by the pandemic.

Among the science subdomains tested, physics is one of the subjects most widely reported to experience conceptual resistance among students at various levels. Physics misconception studies that have been ongoing for a long time consistently identify alternative conceptions regarding force and motion (Clement, 1982; McDermott, 1984), energy and work, momentum and collisions, electricity and circuits, magnetic fields, fluids, waves, and thermodynamics. These misconceptions have proven highly resistant to conventional lecture-based instruction because students tend to maintain intuitive cognitive schemas (naïve physics) that were formed before formal instruction took place (diSessa, 1993; Posner et al., 1982). Hake (1998), through a large-scale interactive engagement study involving more than six thousand students, showed that interactive learning methods providing immediate feedback produce far higher conceptual gains compared to traditional methods, while Hestenes (1987) and Hestenes et al. (1992), through the Modeling Instruction approach, emphasized the importance of students building and testing their own mental models through a modeling-experiment-revision cycle. What all these research traditions share is the recognition that misconception remediation requires intervention that is personal, repeated, diagnosis-based, and provides targeted cognitive conflict, a need that is structurally difficult for teachers to fulfill in large classrooms with an unfavorable teacher-to-student ratio.

At this point, Intelligent Tutoring System (ITS) technology has long been projected as a solution (J. R. Anderson et al., 1985; Vanlehn, 2011), but earlier generations of ITS were built on rule-based knowledge bases that were expensive to develop, rigid in responding to variations in students' natural language, and difficult to scale across topics (Jiang & Jiang, 2024; Nwana, 1990). The emergence of generative LLMs fundamentally changes this landscape because LLMs are able to understand students' questions in diverse natural language, generate contextual explanations, and adapt across subjects without manual rule engineering (Jiang & Jiang, 2024; Nye et al., 2023). Recent research has begun exploring this potential in the physics domain, such as LLM-powered tutoring systems designed to guide

deep learning and precise understanding in students (Jiang & Jiang, 2024), aiPlato, which provides step-wise feedback on physics problem solving (2026), LLM-based automatic scoring systems for physics problems (Educational Sciences, 2025), as well as critical studies on the need to integrate LLMs sensitive to physics pedagogy rather than the careless use of generic chatbots (Küchemann et al., 2024).

Nevertheless, most students and teachers in Indonesia currently use generic AI chatbots, ChatGPT, Gemini, and Copilot, on an ad hoc basis to help with physics assignments, without explicit pedagogical design. This practice carries a number of fundamental weaknesses. First, generic chatbots lack a misconception diagnosis mechanism based on conceptual change theory, so the answers given tend to be direct answer-giving rather than scaffolding that triggers cognitive conflict (Sonkar, Chen, et al., 2024). Second, generic chatbots are prone to conceptual hallucination, that is, generating plausible-sounding but factually incorrect physics content, and do not guarantee alignment of answers with the curriculum or national learning outcomes because they are not equipped with a verified physics knowledge base or a Retrieval-Augmented Generation mechanism anchored to credible sources (Gao et al., 2024; Lewis et al., 2021). Third, generic chatbots are not designed to conduct scientific literacy assessment aligned with the OECD-PISA competency framework, namely explaining phenomena scientifically, evaluating and designing scientific inquiry, and interpreting data and evidence scientifically (OECD, 2019). Fourth, the use of generic chatbots without adequate governance raises issues of ethics, student data privacy, algorithmic bias, and the potential for cognitive dependency (cognitive offloading) that could actually weaken students' independent reasoning ability (UNESCO, 2023; Zhai, 2023).

A review of the literature reveals a clear research gap at three levels. At the theoretical level, AI in Education (AIED) research has developed rapidly in the domains of mathematics, language, and programming, but specific research integrating LLM architecture with conceptual change theory and Modeling Instruction in the physics domain remains limited in number and has not been integrated into a coherent conceptual framework (Jiang & Jiang, 2024; Yan et al., 2024). At the methodological level, most LLM-based ITS studies still focus on automatic scoring and automatic question generation without including an explicit and validated misconception detection mechanism (Sonkar et al., 2024). At the policy-practice level in Indonesia, Ministry of Primary and Secondary Education (Kemendikdasmen) policy documents and the National Research and Innovation Agency (BRIN) research roadmap on the digital transformation of education have not yet contained a conceptual framework that specifically aligns AI Tutor development with the target of improving PISA scientific literacy and integrated physics misconception remediation. This theory-practice gap becomes even more apparent given the fact that physics teacher training on the pedagogical use of generative AI remains very limited, while the TPACK (Technological Pedagogical Content Knowledge) of physics teachers in the field generally still remains at a shallow level of technology integration (Koh et al., 2014; Mishra & Koehler, 2006).

Based on the above description, the urgency of developing an Adaptive AI Physics Tutor becomes unavoidable. This conceptual article is intended to fill this theoretical framework gap by formulating a conceptual model of an LLM-based Adaptive AI Physics Tutor that explicitly integrates: (1) modern LLM technical architecture (Transformer, RLHF, RAG, fine-tuning, and multimodal capabilities); (2) conceptual change, constructivism, and Modeling Instruction learning theories as the pedagogical foundation for misconception remediation; and (3) the OECD-PISA scientific literacy framework as an instrument for assessing learning outcomes. The conceptual novelty of this article lies in three things. First, the explicit alignment between the technical components of the LLM (Misconception Detector, Student

Model, RAG Physics Knowledge Base) and the three dimensions of PISA scientific literacy competency, which has rarely been discussed in an integrated manner within a single framework. Second, positioning the AI Physics Tutor not as a replacement for the teacher, but as a cognitive mediator in the conceptual change cycle that works alongside the teacher's epistemic role as a learning designer (teacher-in-the-loop). Third, the formulation of a staged development roadmap for 2026-2045 that links the development of AI Physics Tutor technology with the achievement of the 8 Dimensions of Graduate Profile, a framework of graduate character and competencies that replaces the Pancasila Student Profile since Regulation of the Minister of Primary and Secondary Education (Permendikdasmen) Number 10 of 2025 concerning Graduate Competency Standards, and the vision of a Golden Indonesia 2045. The specific objectives of this article are: (a) to synthesize LLM, ITS, and physics education literature to formulate a conceptual model of the Adaptive AI Physics Tutor; (b) to map the relationship between AI Tutor components and conceptual change theory and the PISA scientific literacy framework; and (c) to formulate a Design-Based Research/R&D agenda as a continuation of this conceptual article toward empirical prototype validation.

2. Literature Review

2.1. Large Language Models and Intelligent Tutoring Systems

LLMs use the Transformer's self-attention architecture (Vaswani et al., 2025) and are refined through RLHF to better align with human preferences (Ouyang et al., 2022; Bai et al., 2022). Retrieval-Augmented Generation grounds outputs in verified sources, reducing hallucination (Lewis et al., 2021; Gao et al., 2024), while fine-tuning adapts models to domain-specific data such as student misconception patterns (Sonkar, Ni, et al., 2024). ITS research predates LLMs (J. R. Anderson et al., 1985; VanLehn, 2011), but earlier rule-based systems were rigid and costly to scale (Nwana, 1990). LLMs promise more flexible, natural-language tutoring, with early physics-specific applications emerging (Jiang & Jiang, 2024; Wei et al., 2025), though scholars caution these tools still struggle with multi-step physics reasoning and require deliberate pedagogical design rather than ad hoc use (Küchemann et al., 2024; Kasneci et al., 2023; Yan et al., 2023).

2.2. Physics Misconceptions and Conceptual Change Theory

Students consistently hold intuitive misconceptions that resist instruction, including an impetus-based view of motion (Clement, 1982; Hestenes et al., 1992; McDermott, 1984) and a "current-consumed" model of electricity (McDermott & Shaffer, 1992), reflecting stable pre-instructional knowledge structures (diSessa, 1993). Posner et al.'s (1982) conceptual change theory holds that revising these requires dissatisfaction with the old conception and a new one that is intelligible, plausible, and fruitful, building on Piaget's (1970) and Vygotsky's (1980) constructivism. Hake (1998) showed interactive, feedback-rich instruction produces far larger conceptual gains than passive lecture, and Hestenes's (1987) Modeling Instruction frames remediation as an iterative cycle of model building, testing, and revision.

2.3. Scientific Literacy and the OECD-PISA Framework

PISA defines scientific literacy through three competencies: explaining phenomena scientifically, evaluating and designing scientific inquiry, and interpreting data and evidence (OECD, 2019). Indonesia's PISA 2022 science score fell in absolute terms from 2018, even as its relative ranking improved (OECD, 2023; Kemendikdasmen, 2023). Because unresolved misconceptions limit a student's ability to explain phenomena or interpret related data, misconception remediation and PISA-defined literacy are conceptually linked, though the two

literatures have developed separately. Uneven teacher TPACK (Mishra & Koehler, 2006; Koh et al., 2014) further constrains how well any AI-based intervention can be implemented as intended.

2.4. Linking the Constructs: A Conceptual Chain

The constructs reviewed above suggest a multi-level causal chain linking technology to long-term educational achievement, rather than isolated relationships. LLM architecture, through Transformer, RLHF, and RAG components (Vaswani et al., 2025; Ouyang et al., 2022; Lewis et al., 2021), provides the technical foundation for adaptive tutoring systems capable of personalized, real-time instructional adjustment (Bulger, 2016). When personalization is explicitly grounded in conceptual change theory and cognitive conflict (Posner et al., 1982), it becomes a plausible mechanism for remediating misconceptions that resist conventional instruction (Clement, 1982; Hestenes et al., 1992; McDermott & Shaffer, 1992). Because unresolved misconceptions constrain a student's capacity to explain phenomena or interpret data validly, successful remediation is theoretically a prerequisite for strengthening scientific literacy as defined by PISA (OECD, 2019), which in turn underlies the Higher Order Thinking Skills, critical thinking, scientific reasoning, and argumentation, associated with 21st-century competency frameworks. This chain is not assumed to be strictly linear: interaction data at the literacy and HOTS stage would need to feed back into the student model, allowing the system's personalization strategy to be continuously refined.

3. Methods

This article was prepared using the conceptual paper method (Gilson & Goldberg, 2015; Jaakkola, 2020), a scholarly writing genre that aims to produce novelty through the synthesis and reorganization of ideas from existing literature, rather than through new empirical data collection. The writing follows five complementary steps.

First, a narrative review was conducted on literature from 2021-2026 in the domains of Large Language Model, Intelligent Tutoring System, Artificial Intelligence in Education, and physics education, sourced from the Scopus, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect (Elsevier), Nature Portfolio, and arXiv databases, as well as policy documents from OECD, UNESCO, Kemendikdasmen, and BRIN. Second, a critical review was applied to identify the strengths, limitations, and gaps in existing LLM-based ITS research, particularly regarding the extent to which such systems accommodate theory-based physics misconception remediation. Third, concept analysis was used to dissect and redefine key constructs, namely AI Physics Tutor, Adaptive Learning, Misconception Detector, and Scientific Literacy, in order to obtain precise and non-overlapping conceptual boundaries.

Fourth, theory synthesis was conducted by combining classical learning theories (Piaget (1970); Vygotsky (1980) constructivism, Bruner (1966) discovery learning, Ausubel (1968) meaningful learning, L. W. Anderson and Krathwohl (2001) revised Bloom's taxonomy, Posner et al. (1982) conceptual change, Hake (1998) interactive engagement, Hestenes's Modeling Instruction, Mishra & Koehler's TPACK) with contemporary artificial intelligence theory and technical architecture (Transformer, RLHF, RAG) into a single coherent conceptual model. Fifth, a future research framework was formulated in the final section of the article in the form of a staged development roadmap and a Design-Based Research agenda to direct further research toward the development and validation of an empirical prototype.

The literature inclusion criteria include: (a) published between 2021-2026; (b) discussing LLM, ITS, AIED, scientific literacy, or physics misconceptions; (c) originating from Scopus/WoS indexed journals, reputable proceedings, or official documents of authoritative

institutions (OECD, UNESCO, Kemendikdasmen, BRIN). Classic science education literature (pre-2021) is still included as long as it constitutes a foundational (seminal) reference that remains a reference point for contemporary research. Conversely, the exclusion criteria include: (a) sources published before 2021 that do not meet the seminal-reference threshold described above; (b) publications with no substantive discussion of LLM, ITS, AIED, scientific literacy, or physics misconceptions; (c) non-indexed sources such as personal blogs, unreviewed preprints outside arXiv, opinion pieces, or promotional/vendor material without peer review or institutional authority; and (d) non-Scopus/WoS-indexed journal articles and proceedings that are not official documents of an authoritative institution (OECD, UNESCO, Kemendikdasmen, BRIN).

It should be emphasized that this article uses a conceptual paper approach as formulated by Jaakkola (2020) as well as Gilson and Goldberg (2015), not a systematic literature review (SLR); therefore, this article does not present a PRISMA-style search protocol (standard search strings, a single search date, a quantitative selection flow diagram) because the purpose of the writing is not to exhaustively map the entire population of studies, but rather to purposively and argumentatively synthesize constructs across disciplines to produce theoretical novelty. This limitation is explicitly acknowledged as a methodological delimitation: the selection of literature is purposive based on conceptual relevance, so independent replication of the literature selection process cannot be carried out as in an SLR, and the results of the synthesis need to be read as theoretical propositions open to empirical testing in further research, not as conclusions that have been empirically validated.

4. Results and Discussion

4.1. Large Language Models in Education

A Large Language Model is a large-scale language model trained using the Transformer architecture (Vaswani et al., 2025), a neural network architecture that relies on a self-attention mechanism to model contextual relationships between tokens in a sequence of text without depending on the sequential processing used in earlier recurrent architectures. The attention mechanism allows the model to learn relevance weights between words in parallel so that the LLM is able to process long contexts and produce semantically coherent output.

Contemporary generations of LLMs, GPT (OpenAI), Gemini (Google DeepMind), Claude (Anthropic), Llama (Meta), DeepSeek, Mistral, and Copilot (Microsoft), essentially share the same architectural foundation but differ in parameter scale, training data, and alignment strategy. Reinforcement Learning from Human Feedback (RLHF) is an alignment technique that trains a model to produce responses that are more helpful, honest, and harmless based on human evaluator preferences (Bai et al., 2022; Ouyang et al., 2022), so that model output becomes more aligned with pedagogical needs compared to a raw language model.

Two other technical techniques are highly relevant to the development of the AI Physics Tutor. First, Retrieval-Augmented Generation (RAG) is a technique that anchors the text generation process to external documents retrieved dynamically from a verified knowledge base, thereby reducing the risk of conceptual hallucination and ensuring alignment of answers with valid physics curricula and textbooks (Gao et al., 2024; Lewis et al., 2021). Second, fine-tuning is a further training process applied to a foundation model using a domain-specific data corpus, for example a collection of high school physics problems, student misconception transcripts, and HOTS assessment rubrics, so that the model becomes more sensitive to the terms, notation, and typical patterns of conceptual error characteristic of physics (Sonkar, Ni, et al., 2024).

In addition, the development of multimodal AI capability allows LLMs to process not only text but also images of force diagrams, motion graphs, electrical circuits, and experimental video recordings, so that it has the potential to support explanations of physics phenomena that are inherently visual-spatial in nature (Yang et al., 2023). This combination of Transformer, RLHF, RAG, fine-tuning, and multimodality is the technical foundation for the Adaptive AI Physics Tutor proposed in this article.

4.2. AI in Physics Education

An Intelligent Tutoring System (ITS) is a computer system designed to provide instruction and feedback tailored to individual student needs without the direct presence of a human instructor (J. L. Anderson et al., 1985; VanLehn, 2011). The concepts of Adaptive Learning and Personalized Learning refer to the mechanism of adjusting the path, pace, and difficulty level of material based on a student's profile and ongoing performance (Bulger, 2016). Learning Analytics provides a mechanism for collecting and analyzing traces of student interaction to generate insight into learning patterns, while AI Assessment and AI Feedback enable automatic scoring of open-ended answers and instant formative feedback (Zawacki-Richter et al., 2019).

Automatic Question Generation uses LLMs to compose new questions aligned with a particular cognitive taxonomy, while Automatic Scoring, including innovative approaches such as Code Generation Based Grading that transforms narrative answers into executable code for verification, has proven capable of assessing complex physics answers with accuracy approaching human assessment (Wei et al., 2025). Conversational AI has become the main interface enabling students to engage in naturalistic dialogue with the tutor system, as demonstrated by research on an LLM-powered tutoring system specifically designed for deep learning of physics (Jiang & Jiang, 2024) and aiPlato, which provides step-wise feedback on physics problem solving.

The opportunities offered by this technology for physics education are extensive: the availability of 24-hour guidance, unlimited scalability regardless of the number of students, personalized remediation pathways, and the ability to generate analogies and multiple representations suited to the student's level of understanding (Küchemann et al., 2024). However, the accompanying challenges are no less significant. Pedagogically, research shows that generic LLMs still experience considerable difficulty when confronted with complex physics problems that demand multi-step reasoning and rigorous mathematical-physical representation (Wei et al., 2025).

Ethically, the use of AI in assessment raises the risk of algorithmic bias due to unrepresentative training data, threats to student data privacy related to learning history and cognitive profile, and the risk of data security breaches when sensitive information is processed by third-party service providers (Holmes et al., 2023; UNESCO, 2023). Therefore, the governance of the AI Physics Tutor must include the principles of algorithmic transparency, data minimization, informed consent from students and parents/guardians, and a periodic bias audit mechanism as recommended by the UNESCO Recommendation on the Ethics of Artificial Intelligence and AI in Education guidelines (UNESCO, 2022, 2023).

4.3. Scientific Literacy in Physics Education

The OECD-PISA scientific literacy framework defines three main scientific competencies that serve as assessment indicators: (1) explain phenomena scientifically, namely the ability to recognize, offer, and evaluate explanations for various natural and technological phenomena; (2) evaluate and design scientific inquiry, namely the ability to describe and assess scientific investigations and propose methodologically appropriate ways

to answer scientific questions; and (3) interpret data and evidence scientifically, namely the ability to analyze and evaluate scientific data, claims, and arguments in various representations and draw valid conclusions (OECD, 2019, 2023). These three competencies demand more than mere mastery of physics facts and formulas; students are required to be able to reason scientifically in authentic contexts that are often unfamiliar.

The use of an AI Physics Tutor has the potential to simultaneously strengthen these three dimensions when deliberately designed to do so (by design). For the competency of explaining phenomena scientifically, the AI Tutor can present various analogies and multiple representations tailored to the student's specific misconceptions, thereby facilitating a shift from naive explanation toward coherent scientific explanation. For the competency of evaluating and designing scientific inquiry, virtual simulations and virtual laboratory practicals generated by the AI Tutor can give students room to design experiments, manipulate variables, and repeatedly evaluate the validity of their own designs without the constraints of physical laboratory equipment. For the competency of interpreting data and scientific evidence, adaptive analytics dashboards and adaptive Higher Order Thinking Skills (HOTS) questions generated by the AI Tutor can train students to critically read graphs, tables, and experimental data. Thus, the relationship between AI use and the strengthening of Higher Order Thinking Skills, critical thinking, scientific reasoning, and scientific argumentation is mediated, that is, entirely dependent on the extent to which the AI Tutor's pedagogical design explicitly targets higher-order cognitive processes, rather than merely accelerating the delivery of answers (Yan et al., 2023; Zhai, 2023).

4.4. Misconceptions in Physics Learning

Studies of physics misconceptions have identified relatively consistent patterns of conceptual error across student populations and across countries. On the concept of force and motion, students tend to believe that force is required to maintain constant motion (impetus theory), contrary to Newton's First Law (Clement, 1982; Hestenes et al., 1992; McDermott, 1984), using the Force Concept Inventory instrument). On the concept of energy, students often conflate energy with force or work, and fail to distinguish between forms of energy and the principle of its conservation.

On the concept of momentum, common misunderstandings occur regarding elastic-inelastic collisions and the law of conservation of momentum. On the concept of electricity, the "current-consumed" mental model is found to be highly resistant (McDermott & Shaffer, 1992). On the concepts of magnetic fields, fluids, waves, and thermodynamics, similar patterns are found in the form of applying linear-sequential causal schemas that fail to accommodate the systemic and proportional relationships characteristic of these concepts.

Conceptual Change Theory, formulated by Posner et al. (1982), asserts that conceptual change only occurs when four conditions are met: the student feels dissatisfaction with the old conception; the new conception must be intelligible; it must be plausible; and it must be fruitful for explaining new phenomena. This theory is rooted in Vygotsky (1980) constructivism, which emphasizes the role of social mediation and the zone of proximal development, as well as Piaget (1970) constructivism, which explains the process of assimilation and accommodation of cognitive schemas.

Cognitive Conflict, namely the creation of a situation that confronts a student's intuitive predictions with actual observed results, becomes the main trigger mechanism for dissatisfaction in the conceptual change cycle. Interactive Engagement (Hake, 1998) provides large-scale quantitative evidence that methods involving students in predict-observe-explain activities produce a normalized gain far higher than passive learning, while Modeling Instruction (Hestenes, 1987) emphasizes the cycle of model development, deployment of the

model to new phenomena, and revision of the model based on empirical evidence as the core of meaningful physics learning. These four theoretical frameworks, conceptual change, constructivism, cognitive conflict, and Modeling Instruction, become the pedagogical foundation that must be internalized into the AI Physics Tutor's algorithm, not merely used as rhetorical justification.

4.5. The Concept of the Adaptive AI Physics Tutor

The conceptual model of the Adaptive AI Physics Tutor proposed in this article is composed of three interconnected layers arranged cyclically: input, AI engine, and output. At the input layer, the system collects four types of student data: (1) student profile, covering basic demographic-pedagogical data and initial learning conditions; (2) learning style, which is represented not as a fixed category but as a representation preference (visual, verbal, mathematical, or virtual-kinesthetic) that can change over time; (3) learning history, namely the trace of the student's interaction with previous material and problems; and (4) misconception diagnostic results and initial scientific literacy scores obtained from diagnostic assessment based on validated instruments such as the Force Concept Inventory or PISA-equivalent instruments.

At the AI Engine layer, five components work in an integrated manner. The Physics Knowledge Base is a curated physics knowledge base containing curriculum, textbooks, common misconceptions along with their clarifications, and verified problems. Retrieval-Augmented Generation anchors the language model's answer generation process to this Physics Knowledge Base so that output remains accurate, contextual, and curriculum-aligned, while minimizing hallucination. The Student Model dynamically represents the student's knowledge status and misconceptions and is updated with every interaction, adapting the principle of knowledge tracing (Sonkar et al., 2023) to the physics domain.

The Misconception Detector is a classification module based on an LLM that has been fine-tuned to recognize patterns of language and reasoning that indicate specific misconceptions from students' open-ended answers, in line with research on LLM-based cognitive modeling of students with misconceptions (Sonkar, Ni, et al., 2024). The Adaptive Learning Engine integrates the output of the Student Model and Misconception Detector to determine the next pedagogical intervention, whether in the form of verbal scaffolding, a new analogy, a virtual simulation, or a specially designed cognitive-conflict problem.

At the output layer, the system produces eight forms of output accessible to both students and teachers: a personal tutor in the form of adaptive conversational dialogue; misconception diagnostics that explain the type and source of the student's conceptual errors; tiered conceptual explanations adjusted to the level of understanding; virtual simulations of physics phenomena that can be manipulated by the student; adaptive HOTS problems tailored to difficulty level and targeted misconception; virtual practicals that allow students to design and run simulated experiments; automatic feedback that is formative and immediate; and a teacher dashboard that presents a map of misconceptions and scientific literacy achievement for the entire class to support the teacher's pedagogical decision-making (teacher-in-the-loop). These three layers work cyclically: the output of student interaction at the output layer is fed back as an update to the learning history at the input layer, so that the system continues to adapt throughout the learning process.

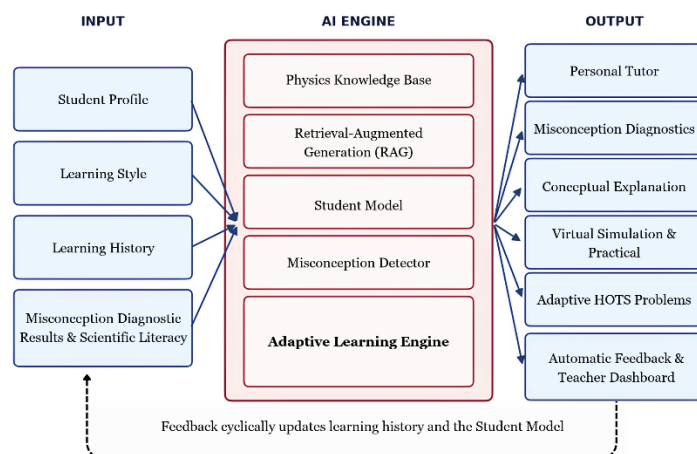


Figure 1. Conceptual Architecture of the LLM-Based Adaptive AI Physics Tutor

4.6. Novelty of the AI Physics Tutor

The novelty of the proposed Adaptive AI Physics Tutor lies in its eight core capabilities: understanding student answers with deep semantic comprehension (not merely keyword matching); detecting misconceptions specifically based on reasoning patterns; providing tiered adaptive scaffolding in accordance with the principle of the Zone of Proximal Development; generating analogies tailored to the student's level and cultural context; creating manipulable virtual experiments; automatically generating problems with calibrated difficulty levels; explicitly assessing scientific literacy achievement aligned with the PISA framework; and recommending subsequent material based on an individual misconception map.

Compared to generic chatbots such as ChatGPT, Gemini, or Claude without pedagogical customization, the Adaptive AI Physics Tutor has advantages in three main aspects: (1) anchoring of answers to a curated Physics Knowledge Base through RAG, which significantly reduces the risk of conceptual hallucination and ensures curriculum alignment, something not guaranteed by generic chatbots that answer based on general knowledge from training data without domain-specific verification; (2) the presence of an explicit Misconception Detector and Student Model based on conceptual change theory, whereas generic chatbots are essentially reactive to prompts without a persistent representation of the student's cognitive status across sessions; and (3) the integration of scientific literacy assessment aligned with the PISA framework as an inherent part of the system, not merely an additional feature.

Compared to conventional Learning Management Systems (LMS), the Adaptive AI Physics Tutor offers two-way dialogic interaction and real-time adaptation, rather than the delivery of uniform static content to all students. Compared to earlier generations of rule-based ITS, the Adaptive AI Physics Tutor offers natural language flexibility and far more efficient content development scalability because it does not require coding logical rules for every variation of misconception (Jiang & Jiang, 2024; Nwana, 1990).

4.7. Theoretical Propositions and Boundary Conditions of the Model

So that the conceptual contribution of this article does not stop at describing the system architecture, the Adaptive AI Physics Tutor model is reformulated into five theoretical propositions that can be empirically tested in further research. Proposition 1 holds that the effectiveness of misconception remediation by the AI Physics Tutor is mediated by the accuracy of the Misconception Detector in classifying the type of student misconception; the

higher the classification accuracy, the greater the likelihood that the cognitive-conflict intervention given is well targeted.

Building on this, Proposition 2 states that the relationship between AI Physics Tutor use and improvement in scientific literacy is moderated by the extent to which the system's scaffolding design explicitly targets the three PISA competencies, rather than merely accelerating problem completion; without this explicit alignment, AI Tutor use is predicted to show no significant correlation with scientific literacy achievement even with high frequency of use. Relatedly, Proposition 3 proposes that the success of misconception remediation depends on sequentially fulfilling the four conditions of conceptual change, namely dissatisfaction, intelligible, plausible, and fruitful; AI interventions that skip the dissatisfaction stage are predicted to produce misconceptions that appear superficially resolved but reemerge in different transfer contexts.

Beyond the system itself, Proposition 4 addresses the human factor, holding that the teacher's role as teacher-in-the-loop moderates the impact of the AI Physics Tutor on Higher Order Thinking Skills; without teacher involvement in following up on the diagnostic dashboard, the system's impact on higher-order reasoning is predicted to be weaker than when the teacher actively designs pedagogical follow-up. Finally, at the structural level, Proposition 5 asserts that the digital infrastructure gap between regions moderates the scalability of the AI Physics Tutor's impact nationally, so that without equitable access, the system's benefits are predicted to be concentrated in schools with adequate digital resources and have the potential to widen, rather than narrow, the education quality gap.

These five propositions apply within certain boundary conditions that need to be emphasized so that the model is not claimed to apply universally. First, this model is designed for the context of high school students who already possess basic digital literacy and adequate device access; application at the primary education level or in contexts with severely limited digital access requires substantial modification. Second, the model assumes the availability of a Physics Knowledge Base that has been curated and validated by experts; without a valid knowledge base, the RAG mechanism cannot guarantee curriculum alignment as claimed, and the risk of conceptual hallucination remains, albeit reduced.

Third, the model assumes the active involvement of the teacher as a learning designer; in contexts where the AI Tutor is used entirely without pedagogical supervision, the risk of cognitive dependency and the decline in students' independent reasoning ability may actually increase rather than decrease. Fourth, the claims of LLM technical superiority in this article are probabilistic, not absolute: retrieval errors, data corpus obsolescence, sensitivity to variations in student prompts, response latency, computational cost, and the potential for model failure remain inherent risks that require ongoing evaluation, human oversight mechanisms, and an escalation path to the teacher when the system is unable to answer adequately.

4.8. AI Physics Tutor Development Roadmap 2026-2045

The development of the Adaptive AI Physics Tutor is projected to proceed through four phases toward 2045, in line with the vision of a Golden Indonesia 2045. Phase one (2026-2030) focuses on prototype development: formulating the Physics Knowledge Base for the high school physics curriculum, initial training of the Misconception Detector in the domain of force-motion and energy, limited trials at several partner schools, and validation by subject matter experts, media experts, and scientific literacy assessment experts. Phase two (2031-2035) focuses on integration with school Learning Management Systems and Ruang GTK, a teacher and education personnel competency development platform resulting from Kemendikdasmen's digital transformation under the Rumah Pendidikan ecosystem

(Kemendikdasmen, 2025), as well as expanding material coverage to all high school physics subdomains and developing a provincial-scale teacher analytics dashboard.

Phase three (2036-2040) focuses on national implementation, covering interoperability with Kemendikdasmen's education data system, provision of national education cloud infrastructure, and mass TPACK training for physics teachers throughout Indonesia. Phase four (2041-2045) moves toward the formation of an AI Physics Ecosystem, namely an integrated physics learning ecosystem connecting the AI Physics Tutor with a national virtual laboratory, a national-scale adaptive question bank, and a STEM career recommendation system, as a concrete contribution to strengthening the scientific literacy of Indonesia's golden generation in 2045.

Building on the theoretical chain established above, the proposed model instantiates this relationship structurally: the Adaptive AI Tutor, equipped with a Physics Knowledge Base, Student Model, and Misconception Detector, delivers personalized learning intended to remediate misconceptions and strengthen scientific literacy. Strengthened scientific literacy is positioned as a contributor to Higher Order Thinking Skills, one of the pillars of the 8 Dimensions of Graduate Profile framework (Kemendikdasmen, 2025), and by extension a foundation for the Golden Indonesia 2045 human resource agenda. As in the underlying theoretical chain, this model is cyclical: output from the scientific literacy and HOTS stages feeds back to update the Student Model, allowing the Tutor to refine its personalization over time.

5. Conclusion

This conceptual article formulates an Adaptive AI Physics Tutor based on a Large Language Model, integrating generative AI architecture (Transformer, RLHF, RAG, fine-tuning) with established physics learning theories (conceptual change, constructivism, cognitive conflict, Modeling Instruction) and the OECD-PISA scientific literacy framework. Unlike generic chatbots used ad hoc, it is deliberately designed to diagnose misconceptions, provide tiered scaffolding, and measurably assess scientific literacy. Its novelty lies in aligning AI technical components with the three PISA competencies and in a staged 2026-2045 development roadmap linked to the 8 Dimensions of Graduate Profile and the Golden Indonesia 2045 vision.

Theoretically, the article extends conceptual change theory (Posner et al., 1982) by adding automatic misconception detection as a mechanism for operationalizing dissatisfaction and cognitive conflict, previously dependent entirely on teacher sensitivity. It also extends TPACK (Mishra & Koehler, 2006) with a new dimension, AI-Pedagogical Content Knowledge, the competency to critically design and evaluate AI Tutor use pedagogically, not just operate it technically. Practically, it implies strengthened teacher TPACK, infrastructure and data governance investment by schools, curriculum renewal in teacher education, ethical guidelines for policymakers (Kemendikdasmen), cross-disciplinary research (BRIN), and functional specifications for EdTech developers building misconception-oriented, evidence-based tools.

Realizing this potential requires addressing real constraints: uneven digital infrastructure risks widening rather than narrowing the digital divide; teacher TPACK and AI literacy need systematic strengthening so teachers remain critical designers rather than passive users; and student data governance and AI ethics, including cognitive dependency, equity of access, and algorithmic accountability, must be central at every development stage. Future work should follow a five-stage Design-Based Research/R&D agenda: (1) prototype

development in a limited domain (e.g., force and motion); (2) expert validation of content and construct validity; (3) limited usability and misconception-detection trials; (4) quasi-experimental effectiveness testing against a control group; and (5) national-scale implementation with interoperability and long-term impact evaluation. This agenda positions the article as a foundation for empirical development of the Adaptive AI Physics Tutor in Indonesia.

6. References

- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent Tutoring Systems. *Science*, 228(4698), 456–462. <https://doi.org/10.1126/science.228.4698.456>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.
- Ausubel, D. P. (1968). *Educational Psychology: A Cognitive View*. Holt, Rinehart and Winston.
- Bai, Y., Kadavath, S., Kundu, S., Askill, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., ... Kaplan, J. (2022). Constitutional AI: Harmlessness from AI Feedback. *ArXiv Preprint*. <http://arxiv.org/abs/2212.08073>
- Bruner, J. (1966). *Toward a Theory of Instruction*. Harvard University Press.
- Bulger, M. (2016). *Personalized Learning: The Conversations We're Not Having*. Data & Society Research Institute. <https://doi.org/10.69985/obue9001>
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50(1), 66–71. <https://doi.org/10.1119/1.12989>
- diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10(2–3), 105–225. <https://doi.org/10.1080/07370008.1985.9649008>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). Retrieval-Augmented Generation for Large Language Models: A Survey. *ArXiv Preprint*. <http://arxiv.org/abs/2312.10997>
- Gilson, L. L., & Goldberg, C. B. (2015). Editors' Comment: So, What is A Conceptual paper? *Group & Organization Management*, 40(2), 127–130. <https://doi.org/10.1177/1059601115576425>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, 55(5), 440–454. <https://doi.org/10.1119/1.15129>
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141–158. <https://doi.org/10.1119/1.2343497>
- Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In *Data ethics : building trust : how digital technologies can serve humanity* (pp. 621–653). Globethics Publications. <https://doi.org/10.58863/20.500.12424/4276068>
- Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS Review*, 10(1–2), 18–26. <https://doi.org/10.1007/s13162-020-00161-0>
- Jiang, Z., & Jiang, M. (2024). Beyond Answers: Large Language Model-Powered Tutoring System in Physics Education for Deep Learning and Precise Understanding. *ArXiv Preprint*. <http://arxiv.org/abs/2406.10934>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T.,

- Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*. <https://doi.org/10.35542/osf.io/5er8f>
- Kemendikdasmen. (2023). *Peringkat Indonesia pada PISA 2022 Naik 5-6 Posisi Dibanding 2018*. Kementerian Pendidikan Dasar dan Menengah Republik Indonesia. <https://www.kemendikdasmen.go.id/siaran-pers/8990-peringkat-indonesia-pada-pisa-2022-naik-5-6-posisi-dibanding-2018>
- Koh, J. H. L., Chai, C. S., & Tsai, C.-C. (2014). Demographic factors, TPACK constructs, and teachers' perceptions of constructivist-oriented TPACK. *Educational Technology & Society*, 17(1), 185–196. <https://www.jstor.org/stable/jeductechsoci.17.issue-1>
- Küchemann, S., Steinert, S., Kuhn, J., Avila, K., & Ruzika, S. (2024). Large language models—Valuable tools that require a sensitive integration into teaching and learning physics. *The Physics Teacher*, 62(5), 400–402. <https://doi.org/10.1119/5.0212374>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *ArXiv Preprint*. <http://arxiv.org/abs/2005.11401>
- McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today*, 37(7), 24–32. <https://doi.org/10.1063/1.2916318>
- McDermott, L. C., & Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding. *American Journal of Physics*, 60(11), 994–1003. <https://doi.org/10.1119/1.17003>
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1177/016146810610800610>
- Nwana, H. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4). <https://doi.org/10.1007/bf00168958>
- Nye, B. D., Mee, D., & Core, M. G. (2023). Generative Large Language Models for Dialog-Based Tutoring: An Early Consideration of Opportunities and Concerns. *CEUR Workshop Proceedings*, 1–11. <https://ceur-ws.org/Vol-3487/paper4.pdf>
- OECD. (2019). PISA 2018 Assessment and Analytical Framework. In *PISA*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- OECD. (2023). PISA 2022 Results (Volume I). In *PISA*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training Language Models to Follow Instructions with Human Feedback. In *Advances in Neural Information Processing Systems 35* (pp. 27730–27744). Neural Information Processing Systems Foundation, Inc. (NeurIPS). <https://doi.org/10.52202/068431-2011>
- Piaget, J. (1970). *Genetic Epistemology*. Columbia University Press. <https://doi.org/10.7312/piag91272>
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227. <https://doi.org/10.1002/sce.3730660207>
- Sonkar, S., Chen, X., Liu, N., Baraniuk, R. G., & Sachan, M. (2024). LLM-based Cognitive Models of Students with Misconceptions. *ArXiv Preprint*. <http://arxiv.org/abs/2410.12294>
- Sonkar, S., Liu, N., Mallick, D., & Baraniuk, R. (2023). CLASS: A Design Framework for Building Intelligent Tutoring Systems Based on Learning Science principles. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 1941–1961). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings->

emnlp.130

- Sonkar, S., Ni, K., Chaudhary, S., & Baraniuk, R. (2024). Pedagogical Alignment of Large Language Models. *Findings of the Association for Computational Linguistics: EMNLP 2024*, 13641–13650. <https://doi.org/10.18653/v1/2024.findings-emnlp.797>
- UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Vanlehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., N.Gomez, A., Kaiser, L., & Polosukhin, I. (2025). *Attention Is All You Need*. Shenzhen Medical Academy of Research and Translation. <https://doi.org/10.65215/ctdc8e75>
- Vygotsky, L. S. (1980). *Mind in Society*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Wei, Y., Zhang, R., Zhang, J., Qi, D., & Cui, W. (2025). Research on intelligent grading of physics problems based on large language models. *Education Sciences*, 15(2), 1–15. <https://doi.org/10.3390/educsci15020116>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- Yang, Z., Li, L., Lin, K., Wang, J., Lin, C.-C., Liu, Z., & Wang, L. (2023). The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). *ArXiv Preprint*. <http://arxiv.org/abs/2309.17421>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (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), 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhai, X. (2023). ChatGPT and AI: The Game Changer for Education. *Tellus*, 1–4. <https://doi.org/10.13140/RG.2.2.31107.37923>