SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Intelligent Tutoring Systems Using Large Language Models for Personalized and Context-Aware Learning

G.M. Prem Kumar[1], Dr. N. HEMA[2]

[1]Research Scholar, Department of Language, Culture and Society, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu (603203), Tamil Nadu, India. E-mail: iamprem009@gmail.com, ORCID: 0009-0006-9723-9983
[2]Professor, Directorate of Learning and Development, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India. E-mail: heman@srmist.edu.in, ORCID:0000-0003-3536-8676

Abstract
Intelligent Tutoring Systems (ITS) have evolved from rule-based instructional environments toward adaptive learning platforms capable of modelling learner characteristics, providing individualized feedback, and dynamically adjusting instructional content. The emergence of Large Language Models (LLMs) introduces a significant opportunity to enhance ITS through natural-language interaction, contextual reasoning, formative assessment, personalized explanations, and learner-aware instructional adaptation. This paper examines the integration of LLMs into intelligent tutoring environments for delivering personalized and context-aware learning experiences. It conceptualizes an LLM-driven ITS architecture comprising learner modelling, contextual representation, pedagogical orchestration, content generation, feedback mechanisms, and continuous learner assessment. The study further examines the potential of LLMs to adapt explanations according to learner proficiency, learning history, misconceptions, interaction patterns, and task context. Particular attention is given to challenges involving hallucination, pedagogical reliability, bias, privacy, explainability, assessment validity, and over-reliance on generated responses. Existing research on intelligent tutoring, adaptive learning, conversational agents, and LLM-based education is synthesized to identify emerging research directions. The analysis indicates that LLM-enhanced ITS can substantially improve instructional flexibility and conversational personalization when combined with structured learner models, retrieval mechanisms, pedagogical constraints, and human oversight. The paper proposes a research-oriented framework for developing trustworthy, context-aware, and pedagogically aligned LLM-based tutoring systems capable of supporting diverse learners across educational settings.

Keywords: Intelligent Tutoring Systems, Large Language Models, Personalized Learning, Context-Aware Learning, Adaptive Learning, Artificial Intelligence in Education, Learner Modelling, Conversational AI, Generative AI, Educational Technology

1. Introduction

The rapid advancement of artificial intelligence (AI) has transformed digital education from static content-delivery environments toward adaptive, interactive, and learner-centred systems. Intelligent Tutoring Systems (ITS) represent one of the most established applications of AI in education, designed to emulate important functions of human tutors by combining domain knowledge, learner modelling, pedagogical strategies, and interactive interfaces [1], [2]. Conventional ITS have demonstrated their ability to provide individualized instruction, diagnose learner performance, and deliver immediate feedback, particularly in structured domains such as mathematics, programming, and science [3], [4]. However, many traditional systems depend heavily on predefined rules, handcrafted instructional paths, and domain-specific knowledge representations, which can constrain their flexibility when learners formulate unexpected questions or require explanations that differ from predefined instructional patterns.

Personalization is a central characteristic of effective intelligent tutoring. Learners differ considerably in prior knowledge, cognitive abilities, learning pace, interaction behaviour, misconceptions, and preferred forms of instructional support. Consequently, an effective tutoring system should not merely present the same instructional material to every learner but should continuously estimate the learner's current state and modify instructional content accordingly [5]. Student modelling research has emphasized the importance of representing learner characteristics, knowledge states, goals, and interaction histories to support meaningful adaptation [6]. Recent research on adaptive e-learning further indicates that learner knowledge and individual characteristics are frequently used to determine the selection, sequencing, and presentation of instructional resources. Nevertheless, the development and empirical justification of comprehensive learner models remain important challenges [7].

The emergence of Large Language Models (LLMs) has created a new technological foundation for addressing several limitations of conventional ITS. Models based on large-scale transformer architectures can understand and generate natural language, follow instructional prompts, summarize information, construct examples, explain concepts at different levels of complexity, and sustain multi-turn dialogue [8], [9]. Unlike traditional rule-based tutoring mechanisms, LLMs can generate responses dynamically rather than selecting exclusively from a predefined collection of instructional statements. This capability creates opportunities for tutoring systems to respond to learners' questions in a conversational manner while adapting explanations to the immediate educational context.

The educational potential of LLMs became particularly prominent following the widespread availability of generative AI systems. Research has identified applications including personalized tutoring, automated feedback, question generation, writing assistance, programming support, assessment, and learning-content generation [10], [11]. LLMs can potentially act as conversational instructional companions capable of providing explanations, asking follow-up questions, identifying apparent misconceptions, generating examples, and adjusting the complexity of responses according to learner needs. Their flexibility also provides an opportunity to extend tutoring beyond predetermined learning paths and support more natural interactions between learners and educational technologies.

However, the integration of LLMs into ITS should not be interpreted simply as replacing a conventional tutoring engine with a language model. Effective tutoring requires pedagogical reasoning in addition to linguistic competence. A system may generate a fluent and apparently convincing explanation while still providing incorrect information, inappropriate difficulty, insufficient scaffolding, or an instructional strategy that does not correspond to the learner's current knowledge state. Foundation models also present challenges involving hallucination, bias, reliability, transparency, and responsible deployment [12]. Therefore, LLM-based tutoring requires additional mechanisms for learner modelling, knowledge grounding, contextual memory, response verification, pedagogical control, and human oversight.

Context awareness is particularly important for developing effective LLM-based ITS. In ordinary conversational systems, an isolated question can often be answered adequately without extensive knowledge of the user's previous interactions. In tutoring, however, the same question may require substantially different responses depending on what the learner has previously studied, which concepts have already been mastered, what errors have occurred, and what learning objective is currently being pursued. Context-aware tutoring therefore requires integration of short-term conversational context with longer-term learner information. Such integration can allow the system to determine whether a learner requires a definition, an example, a hint, a simpler explanation, additional practice, or a more advanced challenge.

The combination of LLMs with learner modelling consequently provides an opportunity to develop a new generation of ITS in which personalization is dynamic, conversational, and context-sensitive. Instead of treating learner modelling and natural-language interaction as separate components, an LLM-enhanced ITS can integrate learner information into the generation and selection of instructional responses. This approach is consistent with the broader evolution of AI-enabled adaptive learning systems, where personalization increasingly depends on the continuous interpretation of learner data and adaptation of instructional strategies [13]. At the same time, recent evidence indicates that although AI-driven ITS generally demonstrate positive educational effects, the magnitude and consistency of these benefits require further empirical investigation, particularly through larger and longer-duration studies [14].

The central research problem, therefore, concerns how LLM capabilities can be integrated with established ITS principles without compromising pedagogical reliability and learner safety. Recent research on LLM applications in education suggests substantial potential for improving academic performance, engagement, accessibility, and individualized learning support, while simultaneously identifying concerns related to over-reliance, privacy, fairness, and technical reliability [15]. These competing opportunities and risks demonstrate the need for a structured framework that considers both the technological and pedagogical dimensions of LLM-based tutoring.

Accordingly, this paper investigates **Intelligent Tutoring Systems Using Large Language Models for Personalized and Context-Aware Learning**. The study focuses on the architecture and functional integration of learner modelling, contextual representation, LLM-based interaction, pedagogical orchestration, adaptive assessment, and feedback. It further examines how LLMs can support personalized explanations and learning pathways while addressing reliability, privacy, bias, explainability, and assessment concerns. The paper ultimately seeks to establish a conceptual foundation for trustworthy LLM-enhanced ITS that can complement human educators and provide scalable, adaptive, and context-sensitive learning support.

2. Literature Review

2.1 Intelligent Tutoring Systems and Adaptive Learning

Intelligent Tutoring Systems have evolved as interdisciplinary educational technologies incorporating artificial intelligence, cognitive science, learning sciences, natural-language processing, and computational modelling. Their fundamental objective is to provide individualized instructional support that approximates selected functions of human tutoring [1]. Unlike conventional computer-assisted instruction, ITS attempt to infer learner knowledge and adapt instructional decisions according to the learner's evolving state. Early cognitive tutors demonstrated how explicit representations of domain knowledge and student cognition could be used to provide targeted assistance during problem solving [3]. Such systems established an important foundation for adaptive instruction by demonstrating that instructional feedback could be generated according to the learner's demonstrated knowledge rather than solely according to task completion.

The effectiveness of ITS has been investigated across multiple educational domains. Comparative research has indicated that intelligent tutoring can achieve learning outcomes approaching those associated with human tutoring under appropriate conditions [2]. However, ITS effectiveness depends considerably on the quality of their domain models, learner models, pedagogical strategies, and interaction mechanisms. Traditional architectures often require extensive expert knowledge engineering, making them expensive and time-consuming to develop and maintain [4]. This limitation becomes particularly important when systems must support open-ended questions and diverse instructional contexts.

Adaptive learning extends the principles of ITS by continuously modifying instructional content, task sequencing, difficulty, feedback, and learning resources according to individual learner characteristics. Recent research indicates that adaptive learning systems commonly incorporate learner models, domain models, adaptation mechanisms, interfaces, and instructional resources to support individualized learning experiences [7]. AI-enabled adaptive systems have further expanded these capabilities through machine learning, educational data mining, and learning analytics [13]. Nevertheless, existing systems frequently remain constrained by the learner characteristics that they explicitly model and the instructional alternatives encoded within their adaptation mechanisms.

2.2 Learner Modelling and Personalization

Learner modelling is a fundamental component of personalized tutoring because adaptation requires some representation of the learner's current state. A learner model may contain information concerning knowledge acquisition, skills, misconceptions, learning progress, preferences, goals, interaction patterns, and other characteristics relevant to instructional decision-making [5]. Knowledge tracing approaches provide an important methodological foundation by estimating the probability that a learner has acquired specific knowledge components based on their sequence of responses and interactions [6]. Such models allow ITS to determine whether a learner should proceed to a new concept, receive additional practice, or revisit previously introduced material.

The literature demonstrates that learner modelling is not limited to academic knowledge. Adaptive systems may incorporate learner characteristics such as prior experience, learning preferences, behavioural patterns, and performance history. A recent systematic analysis of adaptive e-learning systems found that learning styles and knowledge levels are among the commonly used characteristics for personalization, while also highlighting insufficient reporting and theoretical justification concerning learner-model development [7]. This finding is significant for LLM-based ITS because the effectiveness of generated personalization depends directly on the quality, relevance, and temporal accuracy of the learner information supplied to the model.

LLMs create an opportunity to extend learner modelling from structured variables toward richer representations of learner interaction. For example, a tutoring system could maintain a structured learner profile containing mastered concepts, recurring errors, unanswered questions, recent performance, instructional preferences, and current objectives. The LLM could then use this information to adapt explanations and dialogue strategies. Such an approach combines the interpretability and persistence of explicit learner models with the linguistic flexibility of generative models.

2.3 Conversational Agents and Natural-Language Tutoring

Natural-language interaction has long been recognized as an important mechanism for making tutoring systems more flexible and accessible. Traditional conversational tutors, however, generally rely on restricted dialogue structures, predefined intents, or domain-specific linguistic rules. These approaches can provide reliable responses within narrow domains but may struggle with unexpected learner formulations and complex multi-turn interactions.

LLMs substantially alter this paradigm because they can process natural-language questions and generate contextually appropriate responses across a broad range of topics. Their ability to engage in multi-turn dialogue makes them particularly suitable for tutoring scenarios in which learners ask follow-up questions, request clarification, challenge an explanation, or move between related concepts [10], [11]. An LLM-based tutor can potentially transform a single instructional resource into a conversational learning experience in which the learner actively interacts with the instructional system.

However, conversational fluency should not be equated with pedagogical effectiveness. A tutoring dialogue should ideally encourage reasoning rather than simply provide answers. Effective tutoring may require Socratic questioning, graduated hints, error diagnosis, retrieval of prerequisite concepts, or deliberate withholding of complete solutions. Therefore, LLM-based conversational agents require pedagogical orchestration mechanisms that determine not only **what** information should be generated but also **when**, **how**, and **at what level of detail** it should be delivered.

2.4 Large Language Models in Education

The emergence of LLMs has expanded the scope of generative AI applications in education. Educational applications include personalized explanation, content generation, automated feedback, question generation, assessment assistance, language learning, programming support, and conversational tutoring [10], [11]. Their broad linguistic capabilities enable the same model to support different instructional functions without requiring a completely separate rule set for each task.

The literature also indicates that LLMs can support personalization through dynamic generation. A learner asking for a simple explanation can receive an introductory response, whereas another learner can request a technically detailed explanation of the same concept. Similarly, examples can potentially be generated according to the learner's academic discipline, interests, proficiency level, or immediate learning objective. This flexibility distinguishes LLM-based tutoring from systems in which personalization depends primarily on selecting among pre-authored instructional materials.

Nevertheless, the educational deployment of LLMs presents significant reliability concerns. LLMs can generate factually incorrect statements with high linguistic confidence, making hallucination particularly problematic in educational contexts. Incorrect explanations may directly contribute to misconception formation rather than merely failing to answer a question. Concerns have also been raised regarding bias, academic integrity, student over-reliance, privacy, and the appropriate role of generative AI in education [12], [15]. Consequently, LLMs should be embedded within architectures that provide access to validated knowledge sources and mechanisms for controlling and evaluating generated instructional content.

2.5 Context-Aware and Personalized LLM Tutoring

Context awareness represents an important distinction between generic educational chatbots and advanced LLM-based ITS. A context-aware tutor should consider the learner's immediate question together with relevant historical and instructional information. This may include the learner's prior performance, concepts already mastered, recent mistakes, current course objectives, difficulty level, and previous dialogue. Such contextual information enables the system to produce responses that are relevant not merely to the question but to the learner's position within the learning process.

The integration of learner modelling with LLMs can therefore create a feedback loop: learner interactions generate behavioural and performance information; the learner model interprets this information; the pedagogical engine determines an appropriate instructional action; and the LLM generates the corresponding natural-language response. Subsequent learner responses can then update the model and influence future tutoring decisions. This architecture can support continuous personalization rather than one-time adaptation.

Recent research on learner models emphasizes the importance of understanding what learner information should be represented, how it should be updated, and how it should subsequently be used by adaptive systems [7]. For LLM-based ITS, these questions become even more important because excessive contextual information may introduce irrelevant or outdated information into generation, whereas insufficient context can cause the tutor to behave as though every interaction were independent. Effective context management therefore requires mechanisms for selecting, updating, prioritizing, and validating learner information.

2.6 Research Gap

Although substantial literature exists on ITS, adaptive learning, learner modelling, and LLM-based education, these research streams have not yet been fully integrated into a unified tutoring architecture. Conventional ITS offer strong pedagogical structure but can be limited in natural-language flexibility and content generation. LLMs offer exceptional conversational and generative capabilities but do not inherently provide reliable learner modelling, pedagogical sequencing, or guaranteed factual correctness [8], [12].

Recent empirical reviews demonstrate that LLM applications in education are expanding rapidly, with ITS emerging as one of the most prominent application areas [15]. At the same time, evidence concerning long-term learning outcomes, standardized evaluation, pedagogical alignment, and reliable learner adaptation remains comparatively limited. Existing ITS research similarly indicates that educational effects are generally positive but that stronger experimental designs, longer interventions, and more diverse learner populations are needed [14].

The major research gap consequently lies in designing **LLM-enhanced ITS that combine generative flexibility with explicit learner modelling, contextual memory, pedagogical control, grounded knowledge, and rigorous evaluation.** A successful system should not merely generate personalized text; it should determine the learner's instructional needs, select an appropriate pedagogical intervention, generate contextually suitable content, evaluate the learner's response, and continuously update the learner model. Addressing this gap provides the foundation for the proposed conceptual framework and subsequent analysis of LLM-driven personalized and context-aware tutoring systems.

3. Methodology and Conceptual Framework

3.1 Research Design

This study adopts a **conceptual and analytical research design** to develop a comprehensive framework for integrating Large Language Models (LLMs) into Intelligent Tutoring Systems (ITS) for personalized and context-aware learning. The framework is derived through a structured synthesis of established ITS architectures, learner-modelling approaches, adaptive learning principles, conversational artificial intelligence, and recent research on generative AI in education [1], [2], [7], [10], [15]. Rather than evaluating a particular software implementation, the study examines how the principal functional components of an ITS can be reorganized around an LLM while preserving pedagogical structure and learner-centred adaptation.

The methodological approach consists of four interconnected stages. First, the established theoretical foundations of ITS and learner modelling are examined to identify the essential functions required for adaptive tutoring [1], [3], [5]. Second, the capabilities of contemporary LLMs are mapped against these functions, particularly natural-language interaction, explanation generation, question generation, feedback, contextual dialogue, and content adaptation [8], [9]. Third, potential integration points and associated limitations are identified, including hallucination, bias, learner-data privacy, pedagogical inconsistency, and over-reliance on automated responses [10]–[12], [15]. Finally, these findings are synthesized into a conceptual architecture for an LLM-driven ITS.

The proposed framework considers personalization as a continuous instructional process rather than a single adaptation event. Learner interactions are used to update a learner representation, contextual information is incorporated into the tutoring decision, and an appropriate pedagogical action is subsequently translated into an individualized natural-language response. This approach aligns the generative capabilities of LLMs with the structured instructional principles established in ITS research.

Table 1. Methodological Components of the Proposed Study

Component	Primary Purpose	Principal Inputs	Expected Output
Literature synthesis	Establish theoretical foundation	ITS, adaptive learning and LLM research	Consolidated knowledge base
Functional analysis	Identify tutoring requirements	ITS capabilities and limitations	Core system functions
LLM capability mapping	Determine appropriate LLM applications	Natural-language and generative capabilities	LLM-enabled functions
Learner-context analysis	Define personalization variables	Knowledge, performance, history and goals	Learner-context representation
Architectural synthesis	Integrate technological and pedagogical components	Findings from preceding stages	Proposed ITS architecture
Risk analysis	Identify deployment constraints	Reliability, privacy, bias and pedagogical risks	Risk-control requirements

3.2 Theoretical Foundation of the Framework

The proposed framework is grounded in the established architecture of intelligent tutoring systems, generally consisting of a **domain model, learner model, pedagogical model, and interaction interface** [1], [4]. The domain model represents the subject matter and relationships among concepts. The learner model estimates the learner's knowledge and performance. The pedagogical model determines appropriate instructional strategies, while the interface facilitates interaction between learner and system.

LLMs introduce an additional generative and conversational capability into this architecture. Rather than replacing the four established components, the LLM is positioned as an intelligent interaction and generation layer operating under pedagogical and knowledge constraints. This distinction is important because an LLM by itself does not constitute a complete ITS. It requires access to learner information, educational knowledge, instructional objectives, and appropriate pedagogical policies to provide genuinely adaptive tutoring.

The framework therefore combines **structured components** with **generative components**. Structured learner and domain representations provide consistency and control, while the LLM provides flexibility in natural-language generation and conversational interaction. This hybrid structure addresses the principal limitation of relying exclusively on either conventional rule-based ITS or unrestricted generative AI.

Table 2. Relationship Between Conventional ITS Components and LLM-Enhanced Functions

Conventional ITS Component	Traditional Function	LLM-Enhanced Function	Expected Advantage
Domain model	Stores structured subject knowledge	Knowledge-grounded generation and explanation	More flexible instructional presentation
Learner model	Represents learner knowledge and progress	Contextualized learner profile for generation	Deeper personalization
Pedagogical model	Selects instructional strategies	Controls prompting, response type and scaffolding	Adaptive instructional behaviour
User interface	Enables learner-system interaction	Conversational natural-language interface	More natural interaction
Assessment module	Evaluates learner responses	Generates and evaluates context-specific tasks	Dynamic formative assessment
Feedback mechanism	Provides predefined or rule-based feedback	Generates individualized explanatory feedback	Greater feedback flexibility

3.3 Proposed LLM-Driven ITS Architecture

The proposed architecture consists of six principal layers: **learner interaction, learner modelling, contextual memory, knowledge and retrieval, pedagogical orchestration, and LLM generation and evaluation**. These layers operate as an integrated instructional cycle.

The learner interaction layer captures questions, answers, requests for clarification, problem-solving attempts, and other learning behaviours. These interactions are passed to the learner-modelling layer, where relevant information concerning proficiency, mastery, misconceptions, previous performance, and learning objectives is maintained. Learner modelling is particularly important because adaptive tutoring depends on an accurate estimation of the learner's current state [5], [6].

The contextual memory layer maintains both immediate conversational context and relevant long-term instructional history. Short-term context enables the system to understand references made during a current dialogue, while long-term context allows the tutor to recognize previously demonstrated strengths, weaknesses, and learning progress. Context selection should be controlled so that only relevant learner information is incorporated into a particular tutoring decision.

The knowledge and retrieval layer provides access to validated educational content. This component is particularly important for reducing the risk of unsupported or inaccurate LLM responses. Instead of relying exclusively on information encoded within model parameters, the tutoring system can retrieve relevant material from curated textbooks, institutional resources, course notes, knowledge bases, or other validated educational sources.

The pedagogical orchestration layer determines the instructional strategy. For example, when a learner repeatedly demonstrates the same misconception, the system may select remediation rather than simply providing another direct answer. When mastery is demonstrated, the system may increase task difficulty. This layer provides the pedagogical constraints necessary to ensure that language generation remains aligned with instructional objectives.

Finally, the LLM generation and evaluation layer transforms the selected instructional action into natural-language content and evaluates the resulting response for relevance, factual consistency, contextual appropriateness, and pedagogical suitability. This creates a controlled generative architecture rather than an unrestricted chatbot.

Figure-Level Architectural Flow

Learner → Interaction Interface → Learner Model → Context Manager → Knowledge Retrieval → Pedagogical Orchestrator → LLM Tutor → Response Validation → Personalized Feedback → Learner

The resulting interaction forms a continuous cycle because the learner's subsequent response becomes a new input to the learner model. This enables progressive adaptation across multiple interactions and potentially across multiple learning sessions.

3.4 Learner Context Representation

Personalization requires more than identifying whether an answer is correct or incorrect. The proposed framework represents learner context through multiple dimensions, including prior knowledge, current competency, learning objectives, historical performance, misconceptions, interaction behaviour, task difficulty, and recent dialogue.

Prior knowledge determines the level at which a concept should initially be introduced. Current performance indicates whether the learner is progressing, stagnating, or requiring intervention. Misconception information allows the system to provide corrective explanations rather than generic feedback. Learning objectives determine the relevance of instructional content, while interaction history supports continuity across a learning session.

The framework also distinguishes between **stable learner characteristics** and **dynamic learner states**. Stable characteristics may include relatively persistent educational information, whereas dynamic states change as the learner interacts with the system. This distinction prevents temporary difficulties from being interpreted as permanent learner characteristics.

Table 3. Learner Context Variables for Personalized Tutoring

Context Dimension	Examples	Adaptation Supported
Prior knowledge	Prerequisite mastery, previous coursework	Starting instructional level
Current proficiency	Beginner, intermediate, advanced	Explanation complexity
Knowledge gaps	Unmastered concepts	Remedial instruction
Misconceptions	Repeated conceptual errors	Corrective explanation
Learning objective	Examination, concept mastery, skill development	Content prioritization
Interaction history	Previous questions and responses	Conversational continuity
Performance history	Accuracy, completion, attempts	Difficulty adjustment
Current task	Problem type and complexity	Context-specific assistance
Learning pace	Fast, moderate, slow progression	Instructional pacing
Feedback history	Previous hints and corrections	Feedback adaptation

3.5 Personalization and Pedagogical Adaptation

The proposed system personalizes learning at several levels. **Content personalization** modifies the subject matter selected for presentation. **Difficulty personalization** adjusts the complexity of questions and learning tasks. **Explanation personalization** changes the depth, terminology, examples, and presentation style. **Feedback personalization** adapts the degree of assistance according to the learner's demonstrated needs.

A critical feature is progressive scaffolding. Instead of immediately providing the complete solution, the system can provide a conceptual hint, identify the relevant principle, ask a guiding question, or provide a partially worked example. If the learner continues to experience difficulty, progressively stronger assistance can be provided. This approach is consistent with the central objective of tutoring systems: supporting learning rather than merely maximizing the number of questions answered correctly [1], [3].

LLMs are particularly useful in this context because the same instructional concept can be expressed in multiple ways. A complex technical explanation can be reformulated using simpler language, an analogy, an example, a step-by-step explanation, or a domain-specific application. This flexibility supports individualized instruction while retaining the same underlying learning objective.

3.6 Evaluation Dimensions of the Proposed Framework

The effectiveness of an LLM-driven ITS should be evaluated through both **learning outcomes** and **system quality**. Conventional accuracy metrics alone are insufficient because a response can be factually correct but pedagogically inappropriate. Accordingly, the framework proposes six principal evaluation dimensions: learning effectiveness, personalization, contextual consistency, pedagogical quality, reliability, and learner experience.

Learning effectiveness concerns measurable knowledge or skill improvement. Personalization measures whether system behaviour reflects differences among learners. Contextual consistency examines whether the system appropriately uses current and historical context. Pedagogical quality evaluates whether feedback and instructional strategies promote learning. Reliability concerns factual correctness and consistency, while learner experience considers engagement, usability, perceived usefulness, and trust.

Table 4. Evaluation Framework for LLM-Enhanced ITS

Evaluation Dimension	Key Indicators	Evaluation Objective

Learning effectiveness	Pre/post performance, knowledge gain, retention	Determine educational impact
Personalization	Adaptation to proficiency and learner needs	Measure individualized instruction
Context awareness	Context retention, relevant references, continuity	Assess contextual tutoring
Pedagogical quality	Scaffolding, feedback quality, instructional alignment	Assess teaching effectiveness
Reliability	Factual accuracy, consistency, hallucination rate	Assess response trustworthiness
Learner experience	Engagement, satisfaction, usability, confidence	Assess learner acceptance
Teacher support	Intervention quality, transparency, control	Assess human-AI collaboration

4. Proposed LLM-Based Intelligent Tutoring System

4.1 Functional Architecture

The proposed LLM-based ITS is designed as a **hybrid adaptive tutoring environment** in which deterministic educational components and generative AI operate together. The system begins with learner interaction and progressively transforms the interaction into a personalized instructional response.

The first stage captures the learner's input. The system then identifies the subject, task, intent, and relevant conversational context. The learner model provides information regarding the learner's previous performance and current competency. The contextual manager selects relevant historical information, while the knowledge-retrieval component identifies authoritative material related to the current question. The pedagogical engine subsequently determines the most appropriate instructional intervention. Possible interventions include direct explanation, Socratic questioning, worked example, conceptual hint, corrective feedback, remedial instruction, additional practice, or advanced challenge. The LLM then generates the response under these constraints. A validation layer evaluates the generated content before delivery.

Table 5. Major Functional Modules of the Proposed ITS

Module	Core Function	Main Output
Learner Interface	Captures learner interactions	Questions, answers, requests
Intent and Context Analyzer	Interprets current task	Intent and relevant context
Learner Model	Estimates learner state	Competency and learner profile
Context Manager	Maintains relevant history	Context package
Knowledge Repository	Stores validated educational information	Grounded content
Retrieval Module	Retrieves relevant knowledge	Evidence/context
Pedagogical Engine	Selects instructional strategy	Tutoring action
LLM Engine	Generates natural-language response	Personalized response
Validation Layer	Checks generated output	Approved/corrected response
Assessment Module	Evaluates learner performance	Updated learner state
Analytics Module	Tracks progress	Learning analytics

4.2 Adaptive Tutoring Cycle

The proposed system follows a continuous **observe–interpret–adapt–generate–evaluate** cycle. During observation, the system collects learner interaction data. Interpretation determines the learner's apparent intent and instructional state. Adaptation selects the appropriate pedagogical strategy. Generation produces a natural-language response. Evaluation assesses both the learner's subsequent performance and the quality of the system's response.

This cycle enables the tutor to modify its behaviour during a learning session. For example, if a learner correctly answers several increasingly difficult questions, the system can introduce advanced tasks. Conversely, repeated errors can trigger prerequisite revision, simpler examples, additional hints, or alternative explanations.

Table 6. Adaptive Tutoring Cycle

Stage	System Activity	Personalization Decision
Observe	Capture learner interaction	What is the learner doing?

Interpret	Analyse intent and performance	What does the learner currently need?
Retrieve	Obtain rele vant knowledge	Which validated information is required?
Adapt	Select pedagogical strategy	What type of intervention is appropriate?
Generate	Produce individualized response	How should the intervention be communicated?
Validate	Check response quality	Is the response accurate and pedagogically appropriate?
Evaluate	Analyse learner reaction	Did the intervention improve understanding?
Update	Modify learner model	What has been learned about the learner?

4.3 Context-Aware Dialogue Management

Context-aware dialogue is one of the most important differentiating characteristics of the proposed architecture. A learner may ask, for example, "Why does this happen?" without explicitly identifying the concept under discussion. A context-aware system can infer the intended referent from the preceding interaction. More importantly, it can combine the immediate dialogue with the learner's broader instructional state.

The proposed system uses three levels of context: **session context, learner context, and instructional context**. Session context represents the current conversation. Learner context represents relevant historical information concerning performance and knowledge. Instructional context represents the current course, topic, learning objective, task, and expected competency.

Table 7. Context Layers in the Proposed System

Context Layer	Information	Primary Role
Immediate dialogue context	Current question, previous turns	Interpret current interaction
Session context	Current topic and task sequence	Maintain continuity
Learner context	Knowledge, errors, progress	Personalize response
Instructional context	Curriculum and learning objectives	Maintain pedagogical alignment
Historical context	Previous sessions and interventions	Support longitudinal adaptation
Environmental context	Course, assessment or learning setting	Adapt instructional purpose

The use of context should remain selective rather than indiscriminate. Irrelevant historical information may negatively influence generation, while excessive retention can introduce privacy risks. Consequently, contextual memory should prioritize information that has direct instructional relevance.

4.4 LLM-Based Personalized Content Generation

The LLM functions as the primary natural-language generation component of the tutoring system. However, its output is constrained by the learner model, pedagogical strategy, learning objective, and retrieved knowledge. This arrangement allows the system to generate individualized instructional material without treating the LLM as the sole source of educational authority.

Personalized content can include concept explanations, examples, analogies, worked solutions, hints, summaries, practice questions, revision exercises, and formative assessments. The system can also transform the same underlying concept into different representations according to learner proficiency.

For a beginner, the system may use foundational terminology and concrete examples. For an intermediate learner, it may introduce relationships among concepts and progressively complex applications. For an advanced learner, it may emphasize abstraction, critical reasoning, exceptions, and problem-solving. This adaptive generation provides a substantial advantage over static instructional repositories.

Table 8. Personalized LLM Tutoring Functions

Function	Conventional Approach	Proposed LLM-Enhanced Approach
Explanation	Pre-authored explanation	Dynamically adapted explanation
Example generation	Fixed examples	Context-specific examples
Hint generation	Predefined hints	Progressive personalized hints
Question generation	Question bank	Dynamically generated questions
Feedback	Rule-based feedback	Explanatory learner-specific feedback
Remediation	Fixed remedial pathway	Context-sensitive remediation
Summarization	Static summaries	Learner-specific summaries
Revision	Predetermined exercises	Adaptive revision activities

4.5 Knowledge Grounding and Response Reliability

A major architectural requirement is to prevent the tutoring system from relying exclusively on unconstrained LLM generation. Foundation models can generate fluent but incorrect information, which is particularly problematic when learners treat the system as an authoritative educational source [12]. The proposed architecture therefore incorporates a knowledge-grounding mechanism.

The knowledge repository can contain validated textbooks, institutional course material, academic references, instructor-approved resources, structured domain knowledge, and curriculum-aligned learning objects. When a learner asks a question, the retrieval module identifies relevant information and supplies it to the LLM as contextual evidence. The generated response is subsequently evaluated for consistency with the retrieved knowledge.

This architecture does not completely eliminate hallucination, but it provides an additional layer of factual control. It also permits institutions and instructors to determine which educational resources should serve as authoritative sources. Knowledge grounding is therefore not only a technical mechanism but also a governance mechanism for educational AI.

Table 9. Reliability Risks and Proposed Controls

Risk	Potential Educational Consequence	Proposed Control
Hallucination	Learner develops incorrect understanding	Retrieval and response verification
Outdated information	Obsolete learning content	Curated and regularly updated knowledge base
Bias	Unequal or inappropriate learning experience	Bias evaluation and diverse resources
Overconfident response	Learner assumes uncertain information is correct	Confidence and uncertainty signalling
Pedagogical inconsistency	Ineffective instructional intervention	Pedagogical rules and orchestration
Context error	Incorrect personalization	Context validation and relevance filtering
Assessment error	Invalid learner evaluation	Human review for high-stakes assessment
Excessive assistance	Reduced independent reasoning	Progressive scaffolding

4.6 Human-AI Collaborative Tutoring

The proposed architecture positions the LLM as an augmentation mechanism rather than a complete replacement for teachers. Human educators remain particularly important for curriculum design, high-stakes assessment, intervention decisions, learner motivation, and interpretation of complex educational circumstances [12], [15].

Teachers can use the system to monitor learner progress, identify recurring misconceptions, review AI-generated instructional interactions, and intervene when automated tutoring is insufficient. The system can also provide learner analytics indicating concepts with persistent difficulties, frequency of assistance requests, progress across learning objectives, and areas requiring instructor attention.

This human-AI collaboration is especially important because educational quality cannot be reduced to response accuracy. Teachers can evaluate contextual factors that may not be adequately represented in computational learner models. A human-supervised architecture consequently provides a stronger basis for trustworthy deployment.

Table 10. Human-AI Division of Tutoring Responsibilities

Activity	LLM/ITS Role	Teacher Role
Routine explanations	Primary	Review when necessary
Practice-question generation	Primary	Approve or modify
Immediate feedback	Primary	Monitor
Learner progress tracking	Primary	Interpret
Misconception detection	Assist	Validate important cases
Curriculum design	Support	Primary responsibility
High-stakes assessment	Assist	Primary responsibility
Complex learner intervention	Recommend	Primary responsibility
Ethical and disciplinary decisions	No autonomous authority	Primary responsibility

4.7 Expected Advantages of the Proposed Framework

The integration of LLMs with established ITS principles provides several potential advantages. First, natural-language generation enables more flexible interaction than conventional predefined dialogue systems. Second, contextual learner information allows the same instructional content to be adapted to different learners. Third, dynamic question and feedback generation can expand the instructional variety available to learners. Fourth, knowledge grounding can provide greater control over factual content. Fifth, integration with learner modelling allows the system to move beyond generic conversational assistance toward continuous instructional adaptation.

At the same time, the proposed architecture emphasizes that these advantages depend on appropriate system design. LLM capability alone does not guarantee personalization or educational effectiveness. The strongest architecture is therefore a **hybrid model in which learner modelling provides personalization, pedagogical orchestration provides instructional control, retrieval provides knowledge grounding, and the LLM provides flexible natural-language interaction**.

Table 11. Conventional ITS, Generic Educational LLM and Proposed LLM-ITS Comparison

Feature	Conventional ITS	Generic Educational LLM	Proposed LLM-ITS
Natural-language interaction	Moderate	High	High
Learner modelling	High	Limited	High
Context retention	Limited–Moderate	Moderate	High
Dynamic content generation	Low–Moderate	High	High
Pedagogical control	High	Variable	High
Knowledge grounding	Structured	Variable	Structured and retrieval-based
Personalized feedback	Moderate	High	High
Adaptive difficulty	High in defined domains	Limited	High
Hallucination risk	Low in closed systems	High	Reduced through grounding and validation
Scalability of content	Limited	High	High
Teacher oversight	High	Variable	High
Longitudinal personalization	Moderate	Limited	High
Multidomain potential	Limited	High	High
Suitability for controlled education	High	Variable	High

The proposed framework consequently bridges the principal strengths of conventional ITS and modern LLMs. Conventional ITS contribute explicit learner modelling, pedagogical structure, and controlled adaptation, while LLMs contribute conversational flexibility, generative capacity, and contextual language interaction. Their integration provides a foundation for personalized and context-aware tutoring that is more flexible than conventional ITS while remaining more pedagogically structured and controllable than an unrestricted educational chatbot.

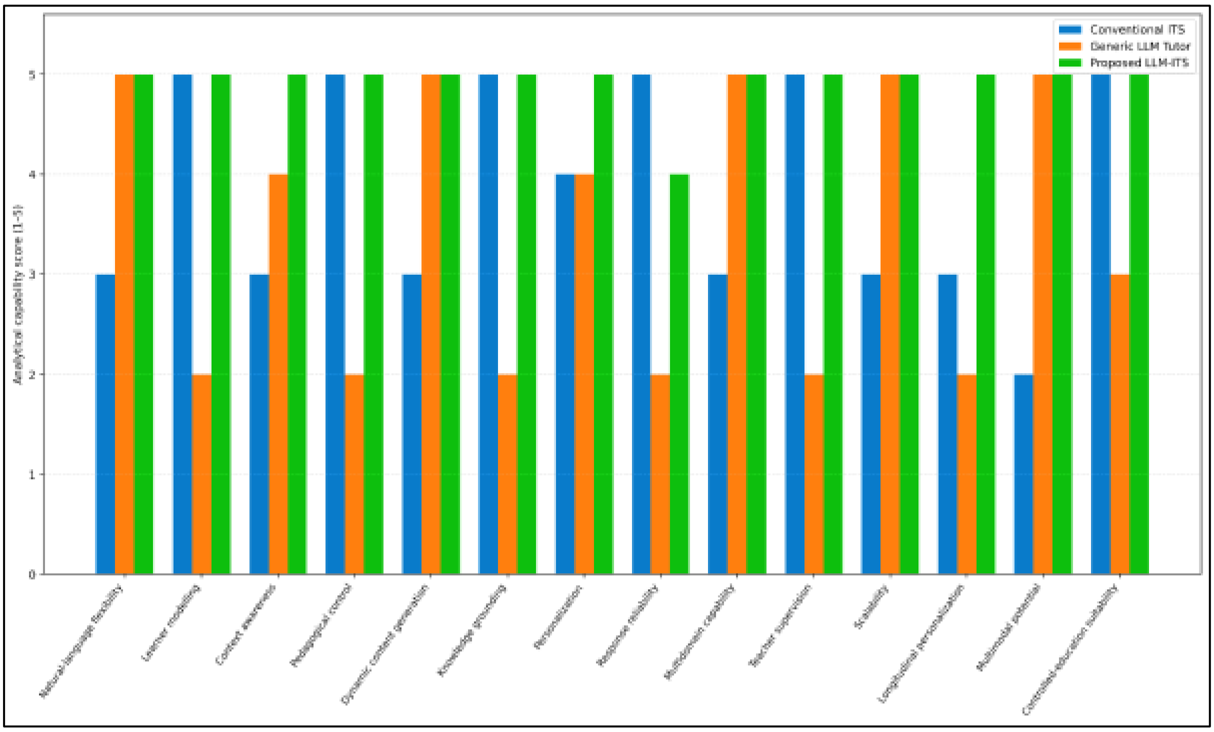

Fig. 1. Comparative analytical assessment of conventional intelligent tutoring systems, generic large language model tutors, and the proposed LLM-enhanced intelligent tutoring system across key functional, pedagogical, personalization, reliability, and scalability dimensions.

5. Results and Discussion

5.1 Personalization Capability of LLM-Enhanced ITS

The proposed framework indicates that LLM-enhanced ITS can provide a substantially more flexible form of personalization than conventional rule-based tutoring systems. Traditional ITS generally personalize instruction by selecting from predefined instructional paths, whereas LLMs can dynamically modify explanations, examples, hints, questions, and feedback according to learner characteristics and immediate interaction context [1], [5], [10]. The effectiveness of this personalization, however, depends on the accuracy of the underlying learner model. If learner proficiency, misconceptions, previous attempts, and learning objectives are represented correctly, the LLM can transform these variables into individualized instructional responses.

Personalization should therefore be understood at multiple levels rather than simply as customized language. Content selection, instructional difficulty, explanation depth, feedback intensity, questioning strategy, and learning sequence can all be adapted. A beginner may receive foundational explanations and concrete examples, while an advanced learner can receive analytical problems requiring higher-order reasoning. This capability supports the principle that tutoring should respond to differences in learner knowledge rather than impose identical instructional experiences [2], [5].

Table 12. Expected Impact of Personalization Mechanisms

Personalization Mechanism	Expected Effect	Educational Value
Adaptive explanation	Better conceptual accessibility	High
Difficulty adjustment	Appropriate cognitive challenge	High
Personalized examples	Improved contextual relevance	High
Progressive hints	Reduced dependence on direct answers	High
Adaptive questioning	Increased active participation	High
Remedial content	Addressing individual knowledge gaps	High
Personalized revision	Reinforcement of weak areas	High
Learning-path adaptation	More efficient progression	High

5.2 Context Awareness and Learning Continuity

Context awareness represents one of the strongest potential advantages of integrating LLMs with learner modelling. A conventional chatbot may treat each question independently, whereas an LLM-driven ITS can

incorporate information from previous interactions, current learning objectives, task difficulty, and learner performance. This allows the tutor to maintain continuity during multi-step learning activities.
The proposed framework distinguishes immediate conversational context from longitudinal learner context. Immediate context supports interpretation of references and follow-up questions, whereas longitudinal context enables the tutor to recognize persistent misconceptions, previously mastered concepts, and historical learning patterns. Such integration can make tutoring more coherent and reduce repetitive explanations.
However, context should be selectively retrieved rather than indiscriminately passed to the LLM. Excessive or irrelevant context may increase response complexity and introduce inappropriate personalization. Effective context management therefore requires relevance filtering, temporal prioritization, and regular updating of learner information.

5.3 Pedagogical Effectiveness and Feedback Quality

LLM-based tutoring provides opportunities to improve formative feedback because responses can be generated according to the learner's specific error rather than relying solely on generic correctness statements. Instead of informing a learner that an answer is incorrect, the system can identify the conceptual step associated with the error, provide a targeted hint, and encourage another attempt.
The framework supports progressive scaffolding in which assistance increases only when required. This is preferable to immediately revealing complete solutions because it encourages learner participation and problem-solving. Cognitive tutors have previously demonstrated the importance of linking instructional feedback to underlying knowledge components [3], while knowledge tracing provides mechanisms for estimating learner mastery over time [6].
LLMs can extend these principles by converting structured diagnostic information into natural and individualized explanations. Nevertheless, pedagogical effectiveness should be evaluated independently from linguistic quality. A fluent response is not necessarily an effective teaching intervention, and empirical learning outcomes remain essential for validating LLM-based tutoring [14].

5.4 Comparative Performance of Conventional and LLM-Enhanced Tutoring

The conceptual comparison indicates that LLM-enhanced ITS can combine the structured adaptation of conventional ITS with the linguistic flexibility of generative AI. Conventional systems generally offer greater control in narrowly defined domains, whereas LLMs offer greater flexibility and broader knowledge coverage. The proposed hybrid architecture attempts to obtain both advantages through explicit learner modelling, knowledge grounding, pedagogical orchestration, and response validation.

Table 13. Comparative Analytical Assessment

Criterion	Conventional ITS	Generic LLM Tutor	Proposed LLM-ITS
Natural-language flexibility	Moderate	Very High	Very High
Learner modelling	High	Low–Moderate	Very High
Context awareness	Moderate	High	Very High
Pedagogical control	Very High	Moderate	Very High
Dynamic content generation	Moderate	Very High	Very High
Knowledge grounding	High	Variable	Very High
Personalization	High	Moderate–High	Very High
Response reliability	High in closed domains	Variable	High with validation
Multidomain capability	Moderate	Very High	Very High
Teacher supervision	High	Variable	High
Scalability	Moderate	Very High	Very High

The comparison suggests that the proposed hybrid approach is theoretically better positioned for educational deployment than an unrestricted LLM tutor. Its primary advantage lies not in replacing established ITS mechanisms but in augmenting them with generative and conversational capabilities.

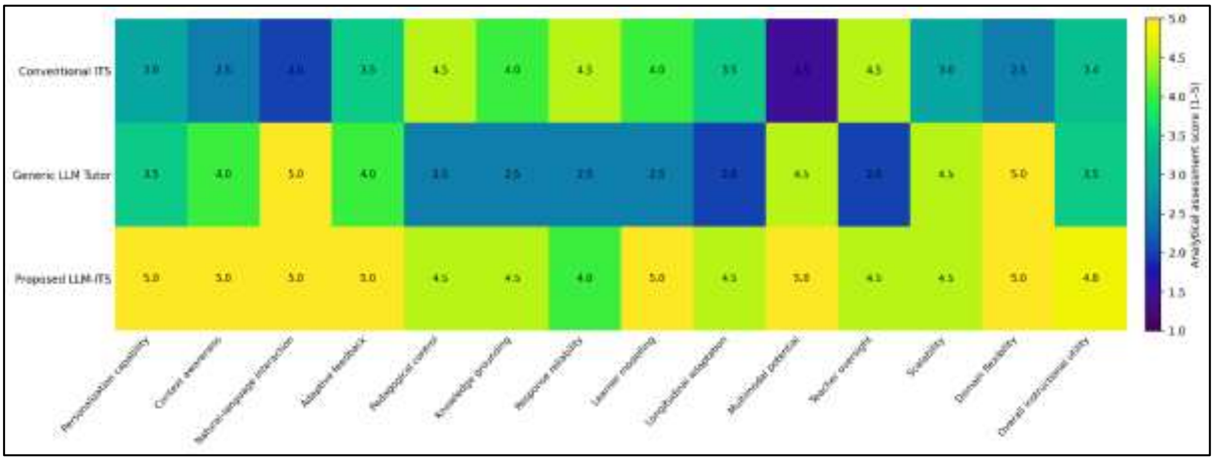

Fig. 2. Detailed heatmap-based comparative assessment of conventional intelligent tutoring systems, generic large language model tutors, and the proposed LLM-enhanced intelligent tutoring system across personalization, contextual awareness, adaptive feedback, pedagogical control, knowledge grounding, learner modelling, reliability, scalability, and overall instructional utility.

5.5 Reliability, Hallucination and Educational Risk

Reliability remains the principal technical challenge for LLM-based tutoring. LLMs can produce responses that are linguistically coherent but factually inaccurate [12]. In educational settings, such errors can be particularly damaging because learners may internalize incorrect explanations. The problem becomes more significant in technical subjects where minor inaccuracies can propagate into subsequent reasoning.

Knowledge grounding, retrieval-augmented generation, structured prompting, response validation, and human oversight can substantially reduce this risk. The proposed framework therefore places validated educational resources between the learner's query and unrestricted model generation. Responses should also be evaluated according to factual consistency, instructional relevance, and alignment with the intended learning objective.

Table 14. Risk–Mitigation Framework

Risk	Severity	Recommended Mitigation
Hallucination	Very High	Retrieval and verification
Incorrect assessment	Very High	Human validation
Learner over-reliance	High	Progressive scaffolding
Bias	High	Bias auditing and diverse datasets
Privacy leakage	Very High	Data minimization and access controls
Context contamination	Medium–High	Relevance filtering
Outdated knowledge	High	Curated knowledge repository
Inappropriate feedback	High	Pedagogical constraints
Reduced critical thinking	High	Question-led tutoring
Excessive automation	Medium–High	Teacher supervision

5.6 Overall Discussion

The analysis demonstrates that the principal contribution of LLMs to ITS is their ability to transform structured learner information into flexible, conversational instructional experiences. Conventional ITS already provide the conceptual foundation for learner modelling and adaptive instruction [1], [4]. LLMs extend this foundation by enabling dynamic language generation and broader interaction capabilities [8], [9].

The most effective architecture is therefore unlikely to be an LLM operating independently. Instead, educationally reliable systems should combine **explicit learner modelling, pedagogical orchestration, knowledge grounding, contextual memory, LLM generation, and validation**. Such integration addresses the central tension between flexibility and control.

The findings are also consistent with broader research suggesting that AI can contribute meaningfully to personalized education while requiring stronger evidence concerning long-term learning outcomes, teacher involvement, and responsible deployment [10], [13], [15]. Future empirical research should therefore evaluate whether increased conversational personalization translates into measurable improvements in knowledge acquisition, retention, engagement, and independent problem-solving.

6. Challenges, Ethical Considerations and Future Directions

6.1 Technical and Pedagogical Challenges

Despite substantial potential, LLM-enhanced ITS face several unresolved challenges. The first is maintaining factual reliability. The second concerns learner-state estimation because personalization is only as effective as the information used to model the learner. The third involves pedagogical consistency. An LLM may produce different responses to similar educational situations unless appropriate instructional policies are enforced.

Another challenge is maintaining continuity over long learning periods. Learner models must be updated without accumulating irrelevant, outdated, or incorrect information. Furthermore, different academic disciplines require different pedagogical strategies; an approach appropriate for language learning may not be suitable for mathematics, engineering, medicine, or professional education.

6.2 Privacy and Data Governance

Personalized tutoring requires learner information, potentially including performance records, interaction histories, assessment responses, and behavioural patterns. Consequently, privacy becomes a central consideration. Systems should collect only information necessary for instructional purposes and should establish clear policies governing storage, access, retention, and deletion.

Educational institutions should also distinguish between information required for personalization and information that is unnecessarily intrusive. Transparent consent, controlled access, data minimization, and secure processing should form essential components of LLM-based ITS governance.

6.3 Bias, Fairness and Accessibility

LLM-based tutoring can reproduce biases present in training data or generate explanations that unintentionally disadvantage particular learner groups. Bias may appear through examples, language, cultural assumptions, difficulty selection, or evaluation behaviour. Continuous auditing is therefore required to determine whether different groups receive comparable instructional opportunities.

Accessibility is an important counterbalancing opportunity. Conversational interfaces can support learners who experience difficulties with conventional text-heavy interfaces, while multilingual generation can potentially expand access to educational resources. However, accessibility benefits must be validated carefully rather than assumed solely from language-generation capability [10], [15].

6.4 Explainability and Trust

Learners and teachers need to understand why an ITS provides a particular recommendation, question, hint, or difficulty level. Excessive opacity can reduce trust, whereas excessive confidence can cause users to accept incorrect information without verification.

The proposed architecture can improve explainability by maintaining explicit learner-state variables and pedagogical decisions outside the LLM. For example, the system can identify that a learner received remedial instruction because repeated responses indicated insufficient mastery of a prerequisite concept. This is more transparent than allowing the LLM to independently determine instructional decisions without an interpretable learner model.

6.5 Human Oversight and Institutional Governance

Teachers should remain central to high-stakes educational decisions. LLM-based ITS can automate routine explanation, formative feedback, question generation, and progress monitoring, but decisions concerning final assessment, academic progression, disciplinary action, and complex learner needs should retain meaningful human oversight.

Institutional governance should establish policies for acceptable AI use, assessment integrity, data protection, model evaluation, incident reporting, and teacher intervention. Such governance is consistent with the broader argument that AI in education should augment educational professionals rather than eliminate their role [8], [10].

6.6 Future Research Directions

Future research should move from conceptual demonstrations toward controlled empirical evaluation. Large-scale studies should compare LLM-enhanced ITS with conventional ITS, human tutoring, and generic educational chatbots using common learning-outcome measures. Evaluation should extend beyond immediate test performance to include retention, transfer, critical thinking, self-regulated learning, and independent problem-solving.

Multimodal tutoring represents another important direction. Future ITS may combine text, diagrams, images, speech, video, code execution, simulations, and interactive problem environments. Agentic architectures could allow tutoring systems to retrieve resources, construct learning activities, evaluate responses, and update learner models through coordinated tools.

Longitudinal learner modelling is particularly promising. Instead of adapting only within a single session, future systems could develop evolving representations of learner mastery across courses and semesters. Such systems would require strong privacy safeguards and transparent mechanisms for correcting inaccurate learner information.

Table 15. Future Research Priorities

Research Area	Current Limitation	Future Requirement
Long-term personalization	Limited longitudinal evidence	Multi-semester learner studies
Pedagogical evaluation	Linguistic quality often emphasized	Learning-outcome-based evaluation
Hallucination control	Residual factual errors	Automated and human verification
Learner modelling	Incomplete learner-state representation	Dynamic multimodal learner models
Multimodal tutoring	Predominantly text-based interaction	Text, speech, image and simulation integration
Explainability	Limited transparency	Interpretable adaptation mechanisms
Fairness	Uneven evidence across learner groups	Systematic fairness auditing
Privacy	Increasing learner-data requirements	Privacy-preserving personalization
Teacher integration	Variable institutional adoption	Human-AI collaborative workflows
Benchmarking	Lack of standardized measures	Common ITS-LLM evaluation benchmarks

The future of LLM-enhanced ITS should therefore focus on **trustworthy personalization rather than unrestricted automation**. The objective should be to create systems that understand learner needs, select pedagogically appropriate interventions, communicate naturally, verify information, and remain accountable to educators and institutions.

7. Conclusion

LLM-enhanced Intelligent Tutoring Systems provide a promising pathway toward personalized and context-aware education. Their combination with learner modelling, contextual memory, pedagogical orchestration, knowledge grounding, and adaptive assessment can overcome several limitations of conventional ITS while preserving instructional control. However, hallucination, privacy, bias, explainability, and learner over-reliance remain substantial concerns. The most effective approach is therefore a human-supervised hybrid architecture rather than an autonomous conversational model. Future research should emphasize longitudinal evaluation, standardized benchmarks, multimodal interaction, reliable knowledge grounding, and measurable learning outcomes. Properly designed LLM-based ITS can complement teachers and deliver scalable, responsive, and individualized learning experiences across diverse educational environments.

References

1. Woolf, B. P. (2010). *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*. Morgan Kaufmann.
2. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
3. Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook: Vol. 3. Application to Learning and Teaching*. American Psychological Association.
4. Nkambou, R., Bourdeau, J., & Mizoguchi, R. (Eds.). (2010). *Advances in Intelligent Tutoring Systems*. Springer.
5. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4(2), 167–207.
6. Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253–278.
7. Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26, 582–599.
8. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson.
9. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
10. 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., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., & Kasneci, G. (2023). ChatGPT for good? On

opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

11. Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15.
12. 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, 39.
13. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., 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. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
14. OpenAI. (2023). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*.