Review on Explainable AI–Driven Intelligent Tutoring System for Hyper-Personalized Lifelong Learning

Shivani Sharma
Department of Computer Science & Informatics,
University of Kota, Kota, Rajasthan, India

Om Prakash Rishi, PhD
Department of Computer Science & Informatics,
University of Kota, Kota, Rajasthan, India

ABSTRACT
Intelligent Tutoring Systems (ITS) have undergone dramatic development over the last 40 years. They now inhabit a unique area where the fields of artificial intelligence, cognitive science, and education converge. This paper reviews the design and development of an explainable AI-based ITS for hyper-personalized lifelong learning. It explores why explainability is crucial, what algorithms power such systems, and how they are architected. The review also draws on evidence regarding the impact that these systems can have on student learning in both formal and informal learning environments. Bloom's mastery learning, Vygotsky's zone of proximal development, and the constructivist approach to learning are key theoretical frameworks discussed. The algorithms examined include Bayesian Knowledge Tracing, Deep Knowledge Tracing, LIME, SHAP, collaborative filtering, and reinforcement learning. Based on foundational studies and recent empirical research, the review concludes that XAI-driven ITS can have a significant impact on learning outcomes when hyper-personalization is tailored to the needs of the individual learner and made transparent to both the learner and the teacher. Finally, the paper lists practical challenges and directions for future exploration.

Keywords
Intelligent tutoring systems, explainable AI, hyper-personalization, lifelong learning, knowledge tracing, student modelling, Bayesian knowledge tracing, deep learning, academic performance.

1. INTRODUCTION
Today's education is different from what it was even 20 years ago. Learning is an ongoing process, regardless of grade level. Learners acquire new things at work, from the Internet, through informal mentoring, and via self-study. Researchers term this shift lifelong learning. However, traditional educational systems were not designed for it. They are rigid and operate under the assumption that all students learn in the same manner and at the same rate. Intelligent Tutoring Systems (ITS) offer something different. They can adapt to each learner in real time by recording students' knowledge, identifying gaps, and selecting the next best piece of content—all without waiting for a teacher's feedback [1]. Over the past 40 years, ITS have evolved from simple, rule-based programs to advanced AI-powered platforms serving millions of users around the world. But a fundamental issue remains: most ITS operate as black boxes. They make decisions (what to teach next, what difficulty level to set, when to give a hint) without providing explanations for these choices. This opacity frustrates students and teachers [9].

For instance, when a student is asked to repeat three modules without being told why, they are left wondering: Are there unmastered concepts? Is it a memory issue? Are they consistently making sloppy mistakes? With no explanation provided, the student is left to guess. Trust erodes, and engagement drops. Explainable Artificial Intelligence (XAI) addresses this issue. It encompasses approaches and strategies that aim to make AI decisions understandable to humans [15]. When XAI is integrated into an ITS, the system can share its reasoning with users. A student can understand the reason behind the selection of a given exercise, while a teacher can review the reasoning behind a system's recommendation. This clarity fosters trust and creates a sense of meaning in the learning process rather than perceived randomness. Hyper-personalization goes one step further. While standard personalization in ITS customizes content based on the student's knowledge, hyper-personalization is much more comprehensive. It takes into account learning styles, emotional state, previous learning patterns, preferred pace, prior knowledge, and future objectives [19]. Relying on more enriched data and more complex algorithms, it generates a truly individualized learning path. The outcome resembles 1-on-1 tutors more closely than any other technology has ever achieved.

The impact of XAI in ITS is strong, yet it presents technical challenges. It demands careful system design, thoughtful algorithm selection, and rigorous evaluation. This paper provides a state-of-the-art review spanning concepts, the historical development of ITS, theories and algorithms of explainability and personalization, system architecture, and the reported effects on students' academic performance.

2. KEY CONCEPTS
2.1 Intelligent Tutoring Systems
An ITS is a software system which enable the individualized instruction and feedback without the presence of human teacher [13]. It models the student's knowledge state, selects appropriate instructional content, and delivers feedback in real time. Early ITS focused on narrow domains such as algebra or grammar. Modern systems can handle multiple subject areas and support complex reasoning tasks.

A canonical ITS has four core components [13]. The domain model contains the subject-matter knowledge the system aims to teach. The student model maintains a dynamic representation of what the learner knows and does not know. The pedagogical model decides how to teach—choosing strategies, sequencing content, and timing feedback. The user interface manages all interactions between the system and the learner.

2.2 Explainable Artificial Intelligence
XAI refers to AI models and methods that can produce human-understandable explanations for their outputs [15]. The term gained prominence following DARPA's XAI program in the mid-2010s [23]. Standard AI models, especially deep neural networks, are difficult to interpret. They can achieve high

accuracy but offer little insight into how they arrived at a prediction. XAI tools like LIME and SHAP were developed to fill this gap [6, 7]. They generate explanations for individual predictions or for the overall behaviour of a model.

In educational settings, XAI has two primary beneficiaries. Students benefit from understanding why they are receiving a specific learning activity. Teachers and administrators benefit from being able to audit and trust the system's decisions [9].

2.3 Hyper-Personalization

Personalization in ITS is defined as the way to adjusting the content and teaching methods to the learner's actual knowledge. Hyper-personalization takes a more comprehensive approach. It adapts across various dimensions at once: knowledge state, learning pace, modality preference, error patterns, engagement level, emotional context and long-term goals [19]. It is based on ongoing data collection and frequent updates to its models. The goal is to make the learning experience feel actually customized to the person – not just "customized" in terms of level.

2.4 Lifelong Learning

Lifelong learning is an alternative way to acquire knowledge and skills throughout a person's life on a voluntary basis [18]. It includes formal training, workplace training, informal training and learning and community-based training and learning. An ITS that supports lifelong learning must be flexible not only on a session-by-session basis, but over the course of years of interaction. The student model should develop over the course of learner's development. The content library needs to be expanded accordingly. The pedagogical strategies also have to change to meet the changing priorities and constraints of an adult learner.

3. EVOLUTIONS OF INTELLIGENT TUTORING SYSTEMS

3.1 Early Beginnings (1970s)

ITS has its origins in the early '1970s. The initial systems were very simple. Carbonell's SCHOLAR (1970) was one of the first AI based instructional programs. It involved students in discussions in their native language concerning South American geography. WEST and SOPHIE came later in the decade. WEST mentored students in a board game math activity. SOPHIE was a simulation of an electronic circuit environment for practice in troubleshooting. These systems were pioneering, yet they were not very adaptive. The AI with in them was limited and the student models were basic.

3.2 Rule-Based Expert Systems (1980s)

In the 1980s, Rule Based Expert Systems became a popular approach for creating AI systems. Most of the ITS that lasted through the 1980's were based on expert system architectures and production rules. Anderson and his colleagues at Carnegie Mellon University modeled human problem-solving in the ACT-R theory in the Cognitive Tutor [21]. It modelled expert knowledge in the form of production rules. Positive reinforcement was given to a student when his/her step matched a rule. If it didn't, the system provided focused feedback. The results of large-scale assessments of the Cognitive Tutor indicated that learning gains were substantial compared with the learning that occurred in the traditional classroom setting [21]. However, the system was very inaccessible for the users.

3.3 Probabilistic and Bayesian Models (1990s)

Probabilistic student models were introduced in the 1990s. The paradigm that emerged to dominate was called Bayesian Knowledge Tracing (BKT) and was introduced by Corbett and Anderson in 1994 [2]. BKT assumes that students' knowledge of a skill is a latent variable that can be estimated from the student's responses. This was more principled than was rule matching. It enabled ITS to monitor learning over time and to make a probability prediction on future performance. It was also somewhat interpretable, as each parameter was easy to interpret from an education perspective.

3.4 Machine Learning and Educational Data Mining (2000s–2010s)

The period from the 2000s to the 2010s is characterized by the adoption of Machine Learning and Educational Data Mining. Machine learning methods were introduced in the research field of ITS in the 2000s. Educational Data Mining became a formal field of study [24]. Scientists started to examine huge log files **from** online learning environments **to find** patterns that human designers hadn't **anticipated**. **Collaborative filtering from** recommendation systems **was** adapted to recommend learning resources, **and** adaptive testing modules were built using Item Response Theory [19]. It was a rapidly expanding field as more and more data **became** available via learning management systems and MOOCs. In 2015, Piech et al. [8] published Deep Knowledge Tracing (DKT), which proved to be a turning point. It represented the students' knowledge by using Long Short-Term Memory (LSTM) networks. DKT was able to account for the nonlinear relationships between skills and time, unlike BKT. It was able to beat BKT on a number of benchmark datasets. It was also very difficult to comprehend, though. This trade-off between model accuracy and transparency would be a core issue in the subsequent work.

3.5 Recent Developments

Natural language processing, computer vision, affective computing and reinforcement learning are included in today's ITS. AI is employed to provide personalized instruction at scale on platforms like Duolingo, Khan Academy, and Carnegie Learning. Some systems are based on facial expression recognition, and they can identify whether the person is confused or bored [18]. One can use sentiment analysis to deduce the emotional state from the text inputs. The field has come far from just correctness tracking. Modern ITS attempt to model all aspects of the learner: cognitive, emotional, and motivational. The latest field is integration with large language models (LLMs). The conversational tutors can be LLMs such as GPT-4. They can describe, respond to questions, and provide comments using natural language [5]. Combining LLMs and structured student models and XAI tools is an ongoing and fast-developing research field.

4. EXPLAINABILITY IN ITS

4.1 Why Explainability Matters in Education

The student is unable to explain why. Has there been any conceptual difference from two weeks ago? Did they rush? Are they mixing up two related concepts? The recommendation without an explanation seems arbitrary so the student can disregard it. Explainability changes this. If the system can tell the student, "We believe that you should review fraction division, because you had 4 out of 5 related problems incorrect this week," then the student has some context. The recommendation makes sense. Research also indicates that pupils who understand the rationale behind decisions tend to follow the guidance of the teacher [9]. Explainability is as well a crucial aspect for teachers. It can be impractical for an instructor to review each interaction log for a group of 30 students. An XAI-powered ITS can bring to light the following insights: “This student is having trouble with a particular concept” and “This student is showing signs of disengagement” in a way that is actionable for the teacher. It makes a black-box system useful [25].

4.2 Types of Explainability

There are two axes that broadly classify XAI methods [15]. The first is scope: global methods tell us about the behaviour of a model while the local methods tell us about individual predictions. The second is dependency: model-specific methods are designed for a particular type of model (for example, decision trees are inherently interpretable) while model-agnostic methods can be applied to any black-box model.

In ITS applications, local model-agnostic explanations are most practically useful. A teacher or student typically wants to understand a specific recommendation, not the system's global logic. Tools like LIME and SHAP are well suited to this purpose [6, 7].

4.3 LIME — Local Interpretable Model-Agnostic Explanations

LIME was proposed by Ribeiro, Singh, and Guestrin in 2016 [6]. The core idea is straightforward. LIME produces a set of perturbed samples of the inputs to the model based on the data point of interest for any prediction that the model makes. It then fits a simple, interpretable model, usually a linear model or decision tree to these samples to approximate the behavior of the complex model locally. This local model explanation is provided by the features that have the highest weights. LIME can provide an explanation to address the question of why a student was identified as "at risk" for dropping out in an ITS. It could show that it was the sharp decline in log-ins, the rise in times when hints were requested, and the tendency to leave portions of assignments incomplete that most influenced it. All of these have a definite educational interpretation. A teacher can immediately take action on it. LIME is quick and adaptable. It comes with its own major drawback: instability. The same input, through which the LIME is run twice, may explain the same phenomena in slightly different ways because of the random sampling. This variability can present difficulties in high-stakes decisions in the educational field [15].

4.4 SHAP — SHapley Additive Explanations

SHAP was introduced by Lundberg and Lee in 2017 [7]. It is based on Shapley values of cooperative game theory. The concept is to consider every feature as a "player" in a game and to determine a "fair" contribution of each of them to the the final prediction. SHAP values remain stable and are accurate locally. They have a number of desirable mathematical properties which LIME cannot always ensure. In education, SHAP can demonstrate for a teacher how much each factor (attendance, time spent on task, accuracy on a quiz, and use of hints, for example) contributed to a predicted final grade. This is useful information. The teacher is able to focus attention on what matters most. In addition to case-level explanations, SHAP provides a summary of explanations across the entire dataset, to provide global insights into model behaviour.

4.5 Attention Mechanisms

Transformers-based models have built-in explainability with the attention mechanism. Attention weights show which sections of the input has the model attended to when producing a prediction. For a language-based ITS, such as a writing tutor, attention maps can indicate which sentences in a composition the model deemed weak. This makes the feedback better understood by both the student and lecturer without the need for an additional explanation module [8]. However, attention weights are not necessarily reliable explanations. A series of studies have found that, despite the high level of attention that is paid to a feature, it is not always the case that it is the feature that led to the prediction. So explanations relying on attention are suggestive, but not definitive.

4.6 Cognitive Theories Supporting XAI in Education

There are several existing cognitive theories that argue the pedagogical importance of explainability in ITS. Working memory is limited and easily overloaded, as suggested by Cognitive Load Theory [Sweller, 1988]. Too complicated or long explanations may hinder learning. This translates directly into a design implication: XAI modules in ITS should provide explanations in easy-to-understand language and be brief. Technical outputs such as SHAP scores need to be communicated to a student in language that they can understand. Constructivism, which is most closely associated with Piaget [20] is a suggestion that learners develop understanding through their own experience and reflections. An ITS that provides a rationale for the use of a specific activity facilitates the process whereby the learner builds meaning from that activity. Learning is not a passive experience, but is one in which the instruction is actively learned. According to Vygotsky's Zone of Proximal Development (ZPD) [4] is the difference between what a learner can do alone and what he can do with the help of another. An ITS that provides explanation as to scaffolding it offers, helps the learner see where he/she is within the ZPD. This ensures that the support is targeted and as such, raises the student's readiness to accept the support.

5. HYPER-PERSONALIZATION

5.1 From Personalization to Hyper-Personalization

Early ITS adapted instruction in one dimension: difficulty. If a student answered correctly, the next problem was harder. If they failed, the next was easier. This is simple adaptive instruction. Hyper-personalization is fundamentally different. It adapts across many dimensions at once [19]. These include knowledge state, learning pace, preferred content modality, prior knowledge from outside the system, error patterns, engagement signals, and long-term learning goals.

Hyper-personalization requires richer data and more computationally intensive algorithms. But the payoff is substantial. When the system's mental model of the learner is accurate and multi-dimensional, the instruction it provides is more precisely targeted. Students spend less time on content they already know. They are challenged at the right level. And they are presented information in the form that suits them best.

5.2 Bloom's 2-Sigma Problem

Benjamin Bloom's landmark 1984 study [3] found that students who received individual one-on-one tutoring performed, on average, two standard deviations better than those in a conventional classroom setting. He called this the "2-sigma problem": the challenge of delivering tutoring-quality instruction at scale.

Bloom proposed mastery learning—requiring students to demonstrate competence before moving on—as a partial solution. ITS have taken this idea further. By automating the assessment of mastery and adjusting instruction continuously, they come closer to the one-on-one tutoring ideal than any other technology. Hyper-personalized ITS are the most direct modern answer to Bloom's challenge [11].

5.3 Bayesian Knowledge Tracing (BKT)

BKT, introduced by Corbett and Anderson in 1994 [2], is the most widely used algorithm for student knowledge modeling It assumes that student's knowledge for each skill is a Hidden Markov Model with two states: known and unknown. The model has 4 parameters: P(L0), the student's initial probability that he/she knows the skill; P(T), the probability that he/she learns the skill with each instance of practice; P(G), the probability that he/she guesses correctly even though he/she does not know the skill; and P(S), the probability that he/she slips and guesses incorrectly when he/she does know the skill. BKT recalculates the probability that the student has mastered the skill after each student response. This posterior estimate is to be used in the system to decide on which content to pick. When the probability of mastery is above a threshold, the system progresses. Otherwise, it assigns more practice. In order to improve individualization, Pardos and Heffernan extended BKT to provide the possibility of having student-specific parameters [17]. BKT is transparent. It has educational interpretations for its parameters. This enables it to be friendly towards XAI goals – the system's reasoning is already comprehensible to an informed teacher. The biggest drawback is that abilities are assumed to be independent, and this is not the case for real learning.

5.4 Deep Knowledge Tracing (DKT)

DKT, which was introduced in 2015 by Piech et al. [8], models student knowledge with a recurrent neural network (RNN), specifically LSTMs. The input is followed by a series of (skill, correct/incorrect) pairs. The output is a prediction vector giving the probability of a correct response for any skill. DKT understands complex and nonlinear relationships among skills and over time. It consistently shows superior performance on common prediction tasks compared to BKT. This improvement comes at the price of being interpretable. An LSTM is a black box! Unlike BKT's parameters and its internal state doesn't have the intuitive meaning. It is at this point that the XAI tools play a crucial role. SHAP can be used to explain the output of DKT, which may show which of the past interactions affected the current prediction the most. This is connecting the accuracy – transparency gap to a usable level [9].

5.5 Collaborative Filtering

Inspired by recommender systems, the collaborative filtering approach finds identical groups of learners based on their profile and use their learning histories to make recommendations for new learners [19]. When a group of students with similar prior knowledge and error patterns learned from a specific set of videos and exercises, the ITS can suggest a set of videos and exercises to a new student whose prior knowledge and error patterns match those of the group. The well-known cold-start problem is a challenge: a new learner without any interaction history, the system has little basis for similarity calculation. This can be solved by hybrid solutions that incorporate the collaborative filtering and the content-based filtering based on descriptions of learning resources used instead of their mere use history.

5.6 Reinforcement Learning for Personalized Pedagogy

Reinforcement learning (RL) is a natural framework for adaptive instruction [18]. The ITS acts as an agent. Its actions are instructional decisions such as choosing the next problem, setting the difficulty, providing a hint, or moving to a new topic. The state represents the student's current knowledge and involvement. Reward is a learning gain measure. The key problem is to specify a good reward function. In education, the learning outcomes can be postponed. A student may have a challenging problem now that may become deeper understanding in the future. Rewarding functions that enable the learning of long-term behavior, rather than short-term behavior, is an ongoing research field. Both policy gradient methods and Q-learning have been used to apply to ITS and have shown good performance in the controlled environment.

5.7 Felder-Silverman Learning Style Model

The Felder-Silverman model categorizes learners into four categories: active/reflective, sensing/intuitive, visual/verbal, and sequential/global [22]. A framework to adapt the presentation of content is used by several ITS. The visual learner is given diagrams and charts. The verbal learner is given explanations in writing. Sequential Learner: Learner receives step by step progressions. Empirical findings about the validity of fixed learning style classifications are mixed. Some people have questioned that learning styles may be situation specific and change over time. The model has, however, proved to be a good design space for structuring multimodal delivery of content in hyper-personalized ITS, particularly when applied in conjunction with adaptation based on observed interaction data.

5.8 Item Response Theory (IRT)

IRT represents the correlation between the latent ability of a student and the probability that he or she will answer an item correctly [24]. IRT differs from simple percentage scores by considering the difficulty of the items, the discrimination of the items, and pseudo-guessing. Models with three parameters (3PL) are often utilized. IRT has been used extensively in large-scale standardized tests and embedded in adaptive testing components in ITS. An IRT-based ITS chooses the question which is most informative about the current ability estimate for the student. This is a computerized adaptive testing (CAT) method which is more efficient for assessment purposes. Less items are required to make an accurate ability estimate. This efficiency is particularly useful for a time-constrained lifelong learner.

6. DESIGN AND DEVELOPMENT OF XAI-DRIVEN ITS

6.1 System Architecture

A well-designed XAI-based ITS comprises of the following interrelated parts: the domain model, the student model, the pedagogical module, the explainability module, and the user interface. All these components are in constant dynamic interaction within a learning session. The target knowledge is modeled as a structured graph, called a domain model. Nodes are concepts or skills. The directed edges are for precedence relationships. If the student does not know a concept, he is not likely to know a related concept. Content sequencing is driven by the domain model. It guarantees that teaching route is arranged in a reasonable sequence in the knowledge graph. The student model could be a dynamic profile of the learner. It maintains the following: estimated knowledge states on each skill, recent error patterns, pace indicators, interaction history, and inferred engagement level. This model is constantly refreshed as you interact with it. In case of a system based on BKT, this update is a Bayesian revision of probabilities. It is the forward pass through the neural network in a DKT based system. The pedagogical module is fed the current student model, and provides as output pedagogical decisions. It chooses the next item of content, the difficulty level, whether it will provide a hint and when a formative assessment will be activated. With RL-based designs, this module is a learned policy. In rule-based designs, it is a collection of heuristics based on pedagogical research. The explainability module is located between the pedagogical module and the user interface. Generates explanation for the systems decision in human-readable form. Internally, the SHAP values, decision rules or attention weights are used. The results are communicated in a layperson's terms. A student reads: "This exercise is on multiplying fractions, since it seems like there are some things you need to work on there. Not SHAP scores.

6.2 Data Pipeline and Privacy

A well-functioning and privacy-aware data pipeline is essential for an XAI-based ITS [12]. Raw data includes student answers, response times, questions for hints, navigation patterns and session times. This information is then pre-processed and stored in a structured way and exploited to update student models and/or to train/fine-tune pedagogical models. Privacy is a primary concern. Student information is confidential, particularly of minors. FERPA applies to systems that are operating in the US. Any systems used in the EU must be GDPR compliant. Minimisation, limitation of purpose, anonymisation of data stored, transparency in disclosure to learners of what is collected and why are key elements of best practice. The ethical handling of data is crucial for the credibility of educational AI systems, as mentioned by Baker and Inventado [12].

6.3 Feedback Design

In educational research, feedback is one of the most important studied aspects. A thorough review of research by Shute found that feedback should be timely, specific, and task-focused, as opposed to person-focused [16]. XAI enhances all three dimensions. It allows for real-time feedback and automates the analysis of student responses. It makes feedback specific, identifying the concept or skill that needs improvement and it doesn't label the student, it has focus on the learning task, it explains some gaps in learning. Conversational feedback using natural language dialogue has been shown to be especially effective by Graesser et al [5]. Students are more likely to interact with feedback if it's like a conversation than a computer-generated grade. With the modern NLP capabilities this sort of dialogue is possible on a large scale.

6.4 Integration Challenges

There are some practical tensions in building a working XAI-driven ITS. Perhaps the most fundamental is the accuracy-explainability trade-off. Higher prediction accuracy is achieved by more complex models such as deep neural networks, ensembles, but these are more difficult to explain. Easier to understand models such as BKT, Decision Trees, and Logistic Regression might not capture the data patterns that are necessary [15]. A second challenge is that the quality of explanations. The explanations produced by the underlying student model will be misleading if the student model is not accurate. A student may be informed that he/she should revise a topic that he/she already knows. The teacher may be informed that a student is "on track" when in fact he/she is not. An explanation that is not good can cause loss of trust more than no explanation [9]. Real-time explanation generation is a third challenge. For complex models with numerous features, calculations for SHAP can be costly. If it is a large system that has thousands of simultaneous users, this computation cost needs to be carefully controlled. A common approach is to use methods such as pre-computation, caching, or approximation.

7. IMPACTS ON STUDENT ACADEMIC PERFORMANCE

7.1 Evidence from Meta-Analyses

There is a considerable and consistent body of empirical evidence of the effectiveness of ITS. Ma et al. [11] meta analyzed 107 studies that compared ITS to other control conditions. They reported a mean effect size of 0.66 standard deviations as compared to classroom instruction which is considered practically significant. The effect size was 0.31 compared to other types of computer-based instruction. This was found to be true both in subject areas and educational levels. VanLehn's exhaustive review [1] saw the average effect size of ITS at approximately 0.76 relative to no-tutoring conditions. This is close to Bloom's fabled 2-sigma effect. The effect sizes that VanLehn obtained depended on the type of feedback loop that was used. Systems giving step-level feedback (where they react to a step in a problem-solving process) gave greater gains than systems giving only answer-level feedback.

7.2 The Role of Transparency on Learning Outcomes

ITS benefits seem to be multiplied when they include explainability. A recent literature review of the application of XAI in education by Khosravi et al. [9] showed that if students could see the reasoning behind the recommendations of the system, then their engagement with the system's recommendations would increase. This helps students to use a hint effectively if they know the reason for including it. If students can see why a problem was chosen for practice, then they are more likely to make a real effort. Conati, Porayska-Pomsta, and Mavrikis [25] believed that open learner models, which make a representation of the learner's knowledge state explicit, are an effective tool to support metacognitive learning. Students monitoring and regulating themselves when they can view what the system believes they know. Self-regulation is a strong predictor of academic success in the long term. This discovery indicates that XAI in ITS not only helps in the immediate performance improvement but also in the formation of sustainable learning skills.

7.3 Personalization and Learning Efficiency

In his study on adaptive hypermedia Brusilovsky demonstrated that an adaptive system can significantly increase learning results, as well as the satisfaction of the learner, as compared to the non-adaptive system. Students who receive content that is appropriate to their level, learn more efficiently [19]. They do not take as much time on material they already know. They don't feel swamped by information that is too sophisticated. Plus, they are more satisfied with the learning experience. These effects are further compounded with hyper-personalization. The tighter the fit between the instruction provided and the learner need, the more that multiple dimensions of the learner's profile are considered in tandem. Additional findings of hyper-personalized ITS have been demonstrated when compared to basic adaptive systems, with improvements in both the time to mastery and knowledge retention found [18].

7.4 Impact in Lifelong Learning Contexts

The situation is more nuanced in situations of lifelong learning and overall positive. Zawacki-Richter et al. [10] surveyed AI systems for higher education and it was found that most of them are personalized and correlated with completion rate and satisfaction. As adult learners are likely to be balancing work, family and study, systems that value time, by eliminating irrelevant content and prioritising real gaps, that are particularly useful for them. Explainability is of even greater significance in the context of professional and vocational training. A medical professional who is taking continuing education must know the rationale for a training module. An engineer who wants to retrain for a new position should be able to find out what particular skill areas the system has identified as gaps. Leaving out explanations, the system's recommendations may be ignored. They are more likely to follow the recommended pathway and to finish it with them [23].

7.5 Equity, Access, and Bias

A well-designed ITS can be an equity tool. It can offer high quality, personalized instruction to students in under-resourced schools, in remote locations or in environments with limited access to expert teachers [18]. This potential of democracy is substantial. Even with a moderately effective ITS, a significant impact can be demonstrated in areas that have a high student to teacher ratio. But there are dangers. Training and ITS on a curriculum mostly from high-performing or well-resourced schools may lead to poor performance in students from underrepresented backgrounds. Algorithmic bias can exacerbate and reinforce preexisting inequities in education. This is a partial protection which is provided by XAI. The system's decision can be made transparent and biases can be seen. They may be identified, reported and addressed. In contrast, opaque systems can silently continue to be biased [15].

8. CHALLENGES AND FUTURE DIRECTIONS

While substantial advancements have been made, there are still challenges that need to be overcome for XAI-driven hyper-personalized ITS systems to realize their full potential. The cold start issue continues to be a problem. For a new learner, there is no history of interaction. The student model should be established through assumptions and/or short assessments of student knowledge. Personalisation is only limited by the amount of data gathered, which is not yet enough. Improved models of prior knowledge and quicker adaptation algorithms are required. The integration of multimodal data is an interesting new area. Today the majority of ITS consist mainly of text-based interactions. Learners communicate by using voice, gestures, written work, facial expression and gaze. When combined, these channels can yield a much more elaborative student model. Some early efforts to use facial analysis to determine emotions have been successful [18]. Nevertheless, the integration of multimodal introduces many issues related to privacy and needs to be part of a carefully governed process. The validation of XAI methods, especially in educational context is in an early stage [9]. The majority of the XAI techniques have been designed and tested in fields such as finance and healthcare. They are not yet fully proven for their effectiveness and reliability in education or their proper use. Further controlled experimental research is warranted, exploring comparisons of learning outcomes with and without XAI explanations, involving various student groups and topics. Longitudinal evaluation is also lacking. The majority of empirical research on ITS focuses on results over a week or a month. However, lifelong learning is a process that takes place over years and decades. It remains unknown if the benefit achieved by XAI enabled ITS can be sustained over time, applied in other areas, or extended to practical applications. Long-term cohort studies are an important priority of the field. Another practical barrier is teacher adoption. Many teachers are skeptical about AI systems and especially about not knowing how the system has come to its conclusions. It is crucial to design ITS that can actively help teachers instead of replacing them with AI. This means co-development of explanation interfaces with teachers, not only students. This also involves the development of professional development courses to support teachers to understand and use the system's results. Future challenges involve using large language models as conversational ITS tutors with transparent reasoning. Federated learning, which enables models to be trained on data stored at students while keeping their data from being sent to a central location, provides a route to scaling personalization in a way that still respects the privacy of student data. The ability to model knowledge within GNNs holds promise for more realistic representation of the interconnections among skills. In the future, Neuro-symbolic AI might provide an architecture that can deliver high accuracy while maintaining transparency, given its fusion of pattern recognition capabilities from neural networks with symbolic reasoning.

9. CONCLUSION

ITS has progressed quite a distance in the last 50 years. The technology has advanced from SCHOLAR's basic geography discussions to LSTM's deep knowledge tracing. Explainable AI and hyper-personalization are the most aspirational directions the field is on right now. The review has demonstrated that there is a solid and expanding evidence base of the effectiveness of ITS. Effect sizes are consistently meaningful in the context of meta-analyses. These benefits are enhanced with personalization. Transparency magnifies them more. In lifelong learning settings, where learners are diverse, self-driven and time limited, the ability to explain and hyper-personalize is an ideal fit to learner's needs. The problem isn't so much technical anymore. It's sensible and humane. How can we, as a society, responsibly use these systems in real educational environments? How can they be fair, accurate and trusted by students, teachers and administrators? “How do we make them work for the 60-year-old nurse who's looking to refresh herself on the latest ways to improve her clinical skills and the 16-year-old who is studying for exams?” These questions involve both interdisciplinary work and collaboration. All parties need to collaborate, including educational researchers, AI scientists, cognitive psychologists, ethicists, and policymakers. It's not just about having better software. To achieve better learning — for all learners and at all life stages.

10. REFERENCES

[1] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," Educational Psychologist, vol. 46, no. 4, pp. 197-221, 2011.

[2] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," User Modeling and User-Adapted Interaction, vol. 4, no. 4, pp. 253-278, 1994.

[3] B. S. Bloom, "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one

tutoring," Educational Researcher, vol. 13, no. 6, pp. 4-16, 1984.

[4] L. S. Vygotsky, Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA, USA: Harvard University Press, 1978.

[5] A. C. Graesser, K. VanLehn, C. P. Rose, P. W. Jordan, and D. Harter, "Intelligent tutoring systems with conversational dialogue," AI Magazine, vol. 22, no. 4, pp. 39-51, 2001.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, Aug. 2016, pp. 1135-1144.

[7] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 4765-4774.

[8] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in Advances in Neural Information Processing Systems (NeurIPS), vol. 28, Montreal, Canada, Dec. 2015, pp. 505-513.

[9] H. Khosravi et al., "Explainable artificial intelligence in education," Computers and Education: Artificial Intelligence, vol. 3, p. 100074, 2022.

[10] O. Zawacki-Richter, V. I. Marin, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education - where are the educators?" Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, p. 39, 2019.

[11] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, "Intelligent tutoring systems and learning outcomes: A meta-analysis," J. Educ. Psychol., vol. 106, no. 4, pp. 901-918, 2014.

[12] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in Learning Analytics: From Research to Practice, J. A. Larusson and B. White, Eds. New York, NY, USA: Springer, 2014, pp. 61-75.

[13] H. S. Nwana, "Intelligent tutoring systems: An overview," Artif. Intell. Rev., vol. 4, no. 4, pp. 251-277, 1990.

[14] J. A. Self, "Bypassing the intractable problem of student modelling," in Proc. Int. Conf. Intelligent Computer-Aided Instruction, Montreal, Canada, 1990, pp. 18-33.

[15] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, Feb. 2017.

[16] V. J. Shute, "Focus on formative feedback," Rev. Educ. Res., vol. 78, no. 1, pp. 153-189, 2008.

[17] Z. A. Pardos and N. T. Heffernan, "Modeling individualization in a Bayesian networks implementation of knowledge tracing," in Proc. 18th Int. Conf. User Modeling, Adaptation, and Personalization (UMAP), Big Island, HI, USA, Jun. 2010, pp. 255-266.

[18] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, Intelligence Unleashed: An Argument for AI in Education. London, UK: Pearson Education, 2016.

[19] P. Brusilovsky, "Adaptive hypermedia," User Model. User-Adapt. Interact., vol. 11, no. 1-2, pp. 87-110, 2001.

[20] J. Piaget, The Origins of Intelligence in Children. New York, NY, USA: W. W. Norton, 1952.

[21] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive tutors: Lessons learned," J. Learn. Sci., vol. 4, no. 2, pp. 167-207, 1995.

[22] R. M. Felder and L. K. Silverman, "Learning and teaching styles in engineering education," J. Eng. Educ., vol. 78, no. 7, pp. 674-681, 1988.

[23] D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," AI Mag., vol. 40, no. 2, pp. 44-58, 2019.

[24] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans. Syst. Man Cybern. C, vol. 40, no. 6, pp. 601-618, 2010.

[25] C. Conati, K. Porayska-Pomsta, and M. Mavrikis, "AI in education needs interpretable machine learning: Lessons from open-learner modelling," arXiv preprint arXiv:1807.00154, Jul. 2018.