Artificial Intelligence in Intelligent Tutoring Robots: A Systematic Review and Design Guidelines

Jinyu Yang [1] and Bo Zhang [2],*

1   School of Computers, National University of Defense Technology, Changsha 410072, China;
    dr.jinyu.yang@outlook.com
2   Artificial Intelligence Research Center, National Innovation Institute of Defense Technology,
    Beijing 100071, China
*   Correspondence: bo.zhang.airc@outlook.com

Abstract: This study provides a systematic review of the recent advances in designing the intelligent tutoring robot (ITR) and summarizes the status quo of applying artificial intelligence (AI) techniques. We first analyze the environment of the ITR and propose a relationship model for describing interactions of ITR with the students, the social milieu, and the curriculum. Then, we transform the relationship model into the perception-planning-action model to explore what AI techniques are suitable to be applied in the ITR. This article provides insights on promoting a human-robot teaching-learning process and AI-assisted educational techniques, which illustrates the design guidelines and future research perspectives in intelligent tutoring robots.

Keywords: artificial intelligence; intelligent tutor system; robot; knowledge graph; decision-making; scene construction

1. Introduction

The recent advances in artificial intelligence (AI) techniques have attracted great contributions from the academia and industry. Powered by the dramatically increased computational power and available data, the deep neural network-based machine learning techniques are prospering and are being applied to facilitate more intelligent robots [1]. Robots have been applied widely in our daily lives, and the number of service robots has already surpassed that of industrial robots in 2008 [2]. Robots are slowly beginning to integrate in daily life. Social robots have played an even more important role in children and young people's lives since robots can be applied to promote their development and intellect.

Education is primarily related to understanding and supporting teaching and learning. It focuses on how to teach and learn since both of them are impacted by communication, course, curriculum design, assessment, and motivation. The continuous advent of new technology and rapid advancement of AI techniques can contribute to improve and enrich educational methods. AI techniques may find ways to enhance the acquisition, manipulation, and utilization of knowledge and the conditions where learning takes place. Hence, they may help educators improve their effective teaching and promote students' individualized learning. Therefore, it is essential to understand what and how AI techniques can be used to achieve educational goals, which produce accessible, affordable, efficient, and effective teaching [3].

With the aid of AI techniques, the robots have been explored and gained increasing popularity in the educational domain [4]. In recent years, many researchers have worked on the specific parts of the intelligent tutoring robots (ITRs). For example, much of the research has focused on personalizing interactions of ITRs to specific users [5]. ITRs adopt these techniques in on-screen tutoring systems to

help tailor the complexity of problems to the capabilities of the student by providing more complex problems only when easier problems have been mastered [6,7]. Papamitsiou and Economides from Reference [8] provide an overview of empirical evidence for understanding the current knowledge on learning analytics and educational data mining and its impact on adaptive learning. Reference [9] reviews the adaptive system in education, and compare intelligent tutoring systems with adaptive hypermedia systems, and show examples of the implementation of the two kinds of systems. Truong has reviewed articles in integrating learning styles and an adaptive e-learning system [4]. These research studies may be helpful in understanding specific aspects of ITRs, but it is difficult for educational researchers to grasp the general framework and understand how to connect educational needs with AI techniques in ITRs.

However, there is a limited survey to comprehensively review the educational robots with the state-of-art AI techniques in the literature. Woolf has worked on a book to provide a comprehensive overview on intelligent tutors in the field of robots [3]. However, the robots and techniques in their book are not state of the art since the AI techniques have been developing rapidly in the past 10 years. Reference [10] presents a review of the field of robots in education from five dimensions: learning activity, location of the activity, the role of the robot, types of robots, and types of robotic behavior. They show that robots are primarily used to provide language, science, or technology education and that a robot can take on the role of a tutor, tool, or peer in the learning activity. The recent survey in Reference [5] has reviewed social robots for education based on four research fields: efficacy of robots in education, robot appearance, behavior, and roles, with a focus on robot design to deliver the learning experience through a social interaction with learners.

Against the background, the aim of this article is to provide a comprehensive overview on ITRs with the state-of-the-art AI techniques and build a bridge between education and robots. This article contributes to help educational researchers acquire knowledge about AI techniques used in the ITRs, and robotics researchers understand the needs in teaching and learning. The main contributions of this article are as follows.

- First of all, we analyze the teaching-learning process between human teachers and students and discuss teacher competences in teaching activities to motivate the exploration of building ITRs and describe the relationship model with the following four factors involved: teacher, student, curriculum, and social milieu.
- Second, we propose a framework based on the relationship model for analyzing and designing ITR with AI techniques, which can integrate different dimensions of teaching activities into this agent framework. Specifically, we use the perception-planning-action framework to fill up the gap between the teaching-learning relationship analysis in the education domain and the ITR design in an AI and robotic domain.
- Third, since AI techniques have been developing rapidly and applied widely in recent years, we provide guidance on applying the state-of-the-art AI techniques in the intelligent tutoring robot, which may provide insights on the research area of intelligent tutor robotics.
- Lastly, we use a case study to illustrate that the proposed perception-planning-action framework may be incorporated for analyzing and designing practical ITRs. We use the case study to show how to further improve the ITR.
- The rest of this paper is organized as follows. The relationship model of the teaching-learning process is analyzed and reviewed in Section 2, based on which the perception-planning-action framework of the ITR is proposed in Section 3, along with reviews and design guidelines of the recent AI techniques that can be used in the systems. Section 4 puts the framework and the techniques together into a case study. The final section provides insights on future research on ITRs with AI techniques.

2. Relationship Model of the Teaching-Learning Process

This section presents three dimensions of teacher competencies in teaching activities, where we can use AI techniques to represent intelligent tutoring robots. How can the ITR replace a human teacher? How can these AI techniques improve teaching? In order to find the answers, we investigated ITRs from the perspective of human teachers.

Heated debates abound on "what teachers should know" from a century ago as a theoretical and empirical discussion. In 1897, John Dewey suggested that the educational process is complicated with both psychological and sociological sides, and educators are required to understand and have knowledge of multiple domains of teaching and learning [11].

Drawing on Schwab's "commonplaces of educating," the analysis of the elements of teacher competence has formed interaction groups [12]. Schwab believes that teaching activities are composed of four factors including teachers, students, and the curriculum, which are all influenced by the social environment where education takes place. These four elements contribute to the experience of teaching and learning, and "none of these is reducible to any other . . . each must be considered in educating" [13] (p. 6). These elements interact and develop into five commonplace groups: The teacher-self, the teacher-social milieu, the teacher-curriculum, the student-curriculum, and the teacher-student [14]. In addition, Krim from Reference [15] places the teacher competence into a classroom context and puts forward 10 factors. These factors include pedagogy, content, curriculum, pedagogy content knowledge, interpersonal, intrapersonal, knowledge of students, adaptive expertise, and social responsibility [11,14–18]. While teacher capacity shows the knowledge, skills, and dispositions that a successful teacher must have, the commonplaces of education demonstrates the experience of teaching, and the events that occur when teaching [14]. These interaction groups are combined with the 10 factors of teacher capacity to form a framework of the teacher's role in teaching activities [15]. Our paper focuses on the intelligent tutor with AI techniques, so we only list the dimensions and skills that can be applied and developed in the intelligent tutor. According to References [14,17,19], human teachers not only interact with the student, the curriculum, and the social milieu, but also with themselves. Therefore, teacher capacity includes the ability for introspection and reflection, which produces confidence and a sense of self-image for teachers [14,17,19]. However, most robots or intelligent tutors did not have such a function, and it is difficult for robots to reflect on their image. In this case, ITRs need to seek guidance from the ITR designers and human tutors. Additionally, there is some overlap between these interaction groups, so we combine the interaction group, students and curriculum, with other interaction groups. Therefore, this article discusses the three dimensions of the teaching activities: teacher and student, teacher and social milieu, and teacher and curriculum.

2.1. Teacher and Student

The first two important elements in teaching activities are the teacher and the learner. These two elements are the main and most important participants in teaching activities. This intergroup describes the dynamic relationship between teachers and students. What capacity is needed in this interaction group?

- Pedagogy: First of all, pedagogy is an essential part of a teacher's professional knowledge during the educating process [17–19], which is a general body of knowledge about learning, instructing, and learners [15]. Pedagogy includes practical aspects of teaching, curricular issues, and the theoretical fundamentals of how and why learning occurs [16].
- Students diversity awareness: Teachers need to understand students and take into account the characteristics of the students, such as race, religion, physical characteristics, personal life choices (clothing, food, music, lifestyle), cultural factors (clothing, food, music, rituals), and body image, or cognitive diversity, such as personality or learning differences [14,16,19]. Additionally, there are emotional, mental, or physical diversities in students, so it is a teacher's

responsibility to understand these differences and promote tolerance, curiosity, and equity among their students [14].

- Responding to students: Teachers need to be aware of students' emotion and response during teaching, so they can respond to them accordingly. Experienced teachers can easily distinguish between active students (taking notes or preparing to make comments) and passive students (too tired or bored to participate) with a quick glance [3]. Teachers are also required to respond the current culture and community in the teaching process and make efforts to promote learning with all students, in particular those with lower scores on the tests for accountability [16]. In addition, teachers should understand learners' thinking, which are the factors that make it easy or difficult for students to learn a particular topic and the best way to teach content to students [16]. Students of different ages and backgrounds bring their own concepts and preconceptions into the process of learning the most regularly taught topics and lessons [16]. If those preconceptions hurdle the learning of new knowledge, teachers need to master the strategies that can reorganize the learners' understanding [16].
- Multiple communication methods: Without communication, teaching cannot take place. Communication is essential since teachers use it to deliver lessons, convey concepts, understand students' knowledge, and motivate students. Using communication strategies and a wide range of methods such as analyzing written work, providing explanations and drawing graphics [3], human teachers can effectively develop knowledge of students and transmit information.
- Building relationships: A human teacher is supposed to communicate with students and others and build relationships with them in order to create a community of learners [20]. This flow of information between teacher and student is key to the educational process. If the teacher fails to figure out how the information they give to students has been received and understood, and is not influenced by the information, the educational process has not been completed successfully [15].

2.2. Teacher and Social Milieu

Social milieu represents the environment where teaching happens, with the teacher, the student, and the curriculum all existing within it. This not only consists of the classroom and the school, but also refers to the relationships between students in school and out of school, the relationships between teachers, and influences from the family and the community [15]. Therefore, teachers need to consider what influence these relationships have exerted on teaching and learning. This interaction group includes factors of social responsibility and adaptive expertise. A key feature of teacher capacity is that teachers need to be able to develop themselves, be flexible, and adapt to changing situations over time [14,17]. This is because the world and its associated knowledge are changing frequently and rapidly [14]. Teachers also need to realize that learning to teach is continuous [17,18]. The "on-going" learning experience helps teachers grow and change since adaptive expertise and context-solving skills cannot be simulated [14]. Therefore, it is important for teachers to understand their social environment and their sense of responsibility as a professional in this environment. For instructor robots, this is the same situation.

In addition, teachers are required to teach with social responsibility or context [15]. This requires teachers to pay sufficient attention to those who have been shown to fall behind in standardized tests and make sure they make progresses that are necessary for them to succeed in learning. It can also be essential for those special schools and students with special needs in particular. This is also discussed in the interaction group between teachers and students.

2.3. Teacher and Curriculum

Teacher competence is based on an intensive comprehension of content or subject matter, substantive knowledge, the facts and concepts, the framework for organizing those facts and concepts, and syntactic knowledge as well as the methods and means of gaining subject knowledge [12,19,21]. Content knowledge, which is also the domain knowledge, is a necessary part of a teacher's professional

knowledge [16,17]. The knowledge represents expert knowledge, or how experts perform in the domain, including definitions, processes, or skills needed [3]. It refers to the knowledge and skills that teachers aim to teach and that students are expected to learn in a specific subject such as mathematics and the English language.

Teachers will not only have the knowledge of a domain, but also grasp the curriculum designed for students, and national and local standards to evaluate students' learning performance [21]. Researchers define curriculum as the content and norms of a course or a program of study [16]. An ideal teacher will "use a variety of open-ended, applied projects so children can practice the reflective process, link subject matter to real-life situations, and think of creative products to demonstrate their learning" [17] (p.11) [19]. Teachers also should have the ability to critically evaluate the curriculum, and look for or design teaching materials that suit their students [16]. National, State, and Local standards serve as a means for evaluating the performance or quality of work that students must achieve [21]. Competent teachers will have knowledge of these standards and be able to organize teaching within the curriculum to achieve the objectives of these standards [21].

Human teachers are also expected to grasp the ways to express and develop content in their subjects that enable learners to comprehend them [16]. This means teachers need to possess the most frequently taught topics, in which the most useful expressions of those ideas are "the most powerful analogies, illustrations, examples, and demonstrations in their subject area" [16] (p. 9, 10). Teachers are also expected to have an understanding of "the curricular alternatives and materials for instruction, the alternative texts, software, programs, visual materials, single-concept films, laboratory demonstrations, or 'invitations to inquiry' [16] (p. 10).

There are many types of knowledge (topics, misconceptions and bugs, affective characteristics, student experience, and stereotypes), and teachers are expected to teach in a variety of ways [3]. It may take many years for human teachers to develop pedagogical content knowledge. However, the ITR may develop such expertise within a much shorter time with the aid of state-of-the-art AI techniques powered by the dramatically increased data and computational resources.

3. Artificial Intelligence Techniques for Designing Intelligent Tutor Robots

Following the relationship model analyzed in the last section, this section is devoted to the analysis of applying AI techniques in building and evaluating the ITR, along with a review of the state-of-art research contributions. In order to fill up the gap between the education domain and artificial intelligence domain, this section first transforms the relationship model by describing the teaching-learning process, into the well-established perception-planning-action model in the area of AI. Then, we analyze the architecture of ITR, and discuss how AI techniques may be applied to the system architecture, module design, and system evaluation. It is worth noting that Woolf approaches the issue from a different perspective, in which ITRs encode student knowledge and domain knowledge, tutoring strategies, and communication [3]. Nevertheless, the perception-planning-action model is more favorable for building a robot system from the designed point of view [22].

Therefore, we propose the perception-planning-action model for ITRs, along with its interactions with the students and the social milieu, as illustrated in Figure 1. Within each perception-planning-action loop, ITR percepts students' activities to collect information for analysis and planning, then reshapes the social and physical milieu for students learning, before starting a new round of perception. The next paragraph provides a brief overview of the specific design and mechanism for each module in the ITR and the rest of this section will be devoted to the detailed design of the perception, planning, and action modules.

The perception module adopts multi-modal sensors to observe students' activities and uses AI techniques for learning style and knowledge mastery analysis, which serves as the input for the planning module. The planning module builds internal models for students and evaluates teaching outcomes for different teaching strategies before making a decision. According to the teaching decision, the action module constructs teaching-learning scenes to generate appropriate social and physical milieus and

uses multi-modal communication channels to deliver teaching contents. Continuous feedback enables the perception-planning-action loop to perform the online adaptation and endows the ITR to learn [22]. Even so, due to the complexity and ethical issues of teaching pedagogy and knowledge structures of the curriculum, human tutors may monitor and intervene in the perception-planning-action loop during either the designing or runtime processes.
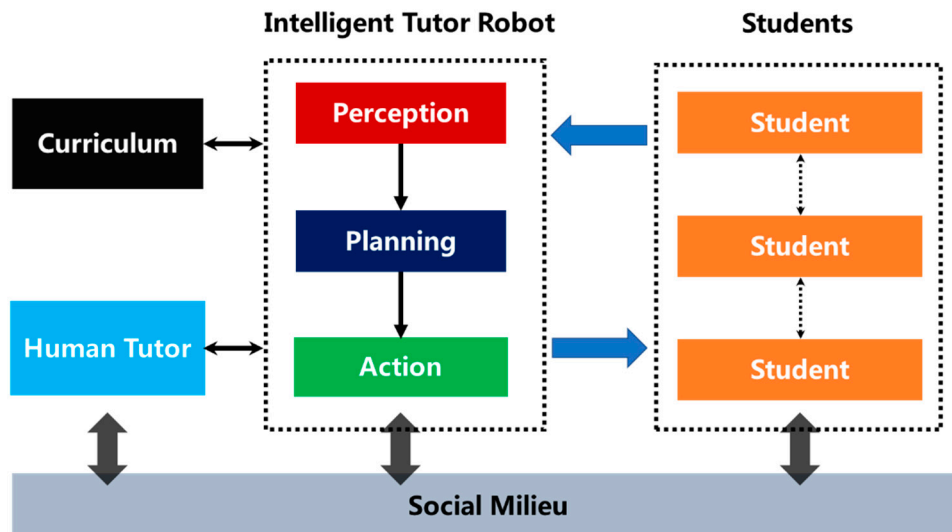

Figure 1. Perception-planning-action architecture of the intelligent tutor robot (ITR) and its interactions with the human-tutor, social milieu, curriculum, and students.

3.1. Multi-Modal Perception of Students

For the human-tutor, the foundation of effective teaching that leads to student learning is to recognize an individual student's learning style and level of knowledge [23]. As shown in Figure 2, the ITR may be equipped with visible light, acoustic, infrared, tactile, and other sensors. First, the ITR may utilize the multi-modal sensors to collect data of the external environments, from which the students' activities may be captured. Then, multi-modal data fusion is implemented to reduce the noise and interference in the data and align data from multi-modal channels. Then, the pre-processed data may be fed to a learning style analysis module and knowledge mastery module, which may apply AI techniques [4] to extract useful information concerning the students learning status, learning styles, and the knowledge level.
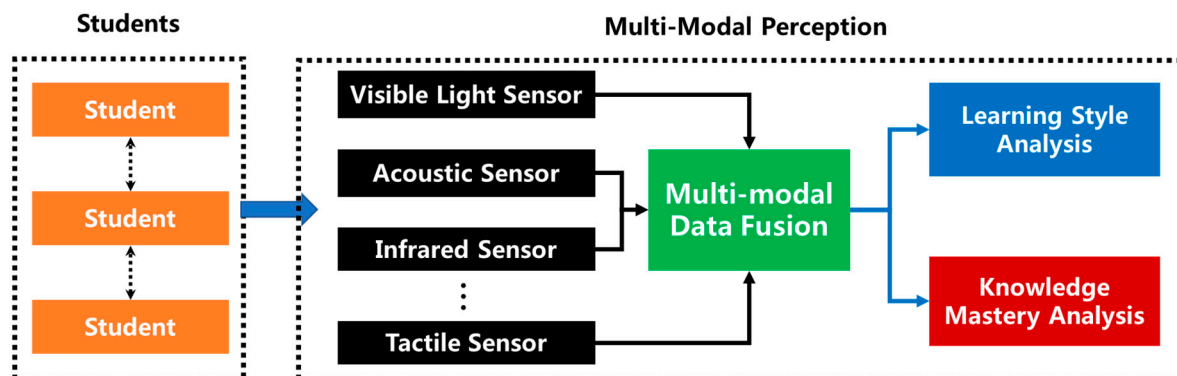

Figure 2. Architecture of multi-modal perception with data fusion and analysis.

3.1.1. Multi-Modal Data Fusion

The first phase of initiating an effective cognitive perception-planning-action loop is to acquire sufficient information concerning the environments and the students. As shown in Figure 2, the ITR may use multi-modal perception by utilizing multiple information channels, e.g., audio, visual, tactile, electromagnetic, and electroencephalograph sensors. These sensors may outperform and augment human-tutor sensory capabilities. Pre-processing steps are demanded before extracting information from the raw signals. First, it is required to remove the noise and interference, while preserving the useful information. Then the signals and data gleaned from multiple independent channels should be aligned in the spatial-temporal space to avoid incorrect correlations [24].

In the context of the ITR design, multi-modal data fusion is not specifically explored, but the topic is well-investigated and still a hot one in the context of data-mining and robotics. Due to space limitations, the rest of the section would not aim at a thorough review of the state-of-art data fusion. Instead, it reviews recent application driven surveys on data fusion and their great potential applications in the intelligent tutor design.

- Pixel-level data fusion: In Reference [25], the authors provide a review of contributions on pixel-level data fusion for multiple visual sensors, with applications to remote sensing, medical diagnosis, surveillance, and photography. Although pixel-level data fusion is seldom devoted to intelligent educational applications, the fusion methods and quality measures design may be transferred to the design of ITRs. For example, matting methods may be used for multiple image fusion of moving objects captured by the moving visual sensor mounted on the ITR in dynamic scenarios, and recent popularized depth image sensor like Kinect may capture both posture and video data of the learners by fusing multi-channel 2D visual images and depth maps, which provides more information than single 2D visual sensory data for estimating learners' status [26,27].

- Human action recognition: In Reference [28], a survey focusing on data fusion in applications of human action recognition is given. The authors compare the pros and cons of combining inertial sensory data with traditional visual sensors. Human action recognition may find its wide range of applications such as video analytics, robotics, and human-computer interaction. Therefore, it may be directly adopted in intelligent tutor design. For instance, in a complex tutoring scenario of much occlusion and moving objects, the inertial sensory data may complement the visual sensory data with a limited field of view using a support vector machine and hidden-Markov model. An example of applying human action recognition for second language ITR in Reference [29] uses hidden-Markov models and k-means algorithms to model and annotate Kinect sensory data, which is then fed to train action-recognition classifiers.

- Affective computing: In Reference [30], the authors discuss in detail the recent contributions in affective computing, and the methods are evolving from uni-modal analysis to multi-modal fusion. Affective computing is an inter-disciplinary research field devoted to endow machines with cognitive capabilities to recognize, interpret, and express emotions and sentiments. Hence, it may be applied directly in ITR for students' emotion and sentiment interpretation that helps the analysis of learning style and knowledge level. From audio-visual sensory data, the ITR may facilitate a series of metabolic variables, e.g., heart monitoring and eye tracking, to gather information about the emotion as well as the level of engagement and attention of the students. Multiple kernel learning and deep convolutional neural network methods may be adopted for the sentiment detection of students. However, the state-of-the-art work in ITRs relies on uni-modal, mostly visual data to derive affective information. A popular commercial tool for affective computing in ITR is Affdex Software Development Kit (SDK), which is adopted by References [31,32] to derive affective information autonomously from video records of children's interactions with an ITR.

The increasing amount of perceptual data may overwhelm the following stages of information processing. Therefore, it demands teaching task-oriented compression. However, as found by the

recent survey [4], there are several predictors that have been taken into account. No study manages to compare the effectiveness of different attributes in predicting learning styles. Hence, it remains an open question on how to select the appropriate set of variables. A promising way of salient feature design is the attention mechanism, which has been shown effective in handling a large amount of audio-visual data flows. This strikes a beneficial balance between information preservation and computing efficiency [33,34].

3.1.2. Learning Style and Knowledge Mastery Analysis

With the aid of multi-modal data fusion, the ITR may gather data about the students with external-assessment and self-assessment, and then implement statistical machine learning methods to analyze the students' learning style and knowledge level.

Before delving into the external and self-assessment methods, the ITR selects the criteria for a learning style. Although a wide range of learning style theories have been developed in the literature, most of them cannot be compared quantitatively so that it is hard to argue only one of them could outperform others [4]. Therefore, the human-tutor may have to analyze the curriculum and select the learning style theory for the ITR and, most importantly, the learning style theory should be projected to a well-defined model in terms of categorization and quantification. For example, the Felder-Silverman's model [35] proposes the well-known four-dimension model for learning styles: perception (Sensory/Intuitive), information input (Image or Verbal), information process (Active or Reflective), and understanding (Sequential or Global). Each dimension may be labelled in supervised learning and be quantitatively scaled.

After selecting the learning style model, an external-assessment may be implemented during the teaching process, which imitates human-tutors' capabilities of observing students' activities and affections to assess the learning status. By analyzing the heterogeneous data gathered from multi-modal sensors, machine learning techniques may extract hidden modes of the students' learning style and status affected by personality, motivation, and emotion [30]. The students' self-assessment that incorporates levels of competencies for each concept or skill may also be implemented, which allows the ITR to continuously calibrate planning and actions to better meet student needs. In this case, a simple way for self-assessment questionnaires is to use a scale that ranges from a cursory level and a factual knowledge level to a conceptual knowledge level and application level. With the aid of simple statistical machine learning techniques and visualized analytics like bar charts, radar charts, and ranking tables, information about the per-student knowledge mastery level may be extracted [36,37].

More sophisticated concept maps may be adopted to analyze students' level and organization of knowledge. The knowledge along with its structure for mastering a curriculum heavily relies on the evaluation and experience of human-tutors. Hence, the ITR may receive inputs from human-tutors, and a well-designed and formalized data structure for the human-robot tutor interface is the concept map and knowledge graph [38,39]. In a concept map or knowledge graph, each vertex is a concept or ontology and a network is formed to describe the relationship between ontologies. Based on a series of related knowledge concepts and a standard concept map designed by human experts, the ITR may request students to construct a knowledge graph [23,39] representing everything that they know about the topic, and then use machine learning techniques designed for graph networks [40,41] to evaluate the density and intensity of the students' knowledge structures. This reveals students' blind spots and weak knowledge connections that need to be strengthened during the follow-up teaching and learning process. Take an example shown in Figure 3. The knowledge graph may be compared with a target graph for the curriculum by the knowledge level analysis module, so that the disparity graph may be generated to reveal the weak and missing knowledge of a student.

The extracted information indicating students' learning style and knowledge levels may be jointly fed to the planning modules of the ITR.
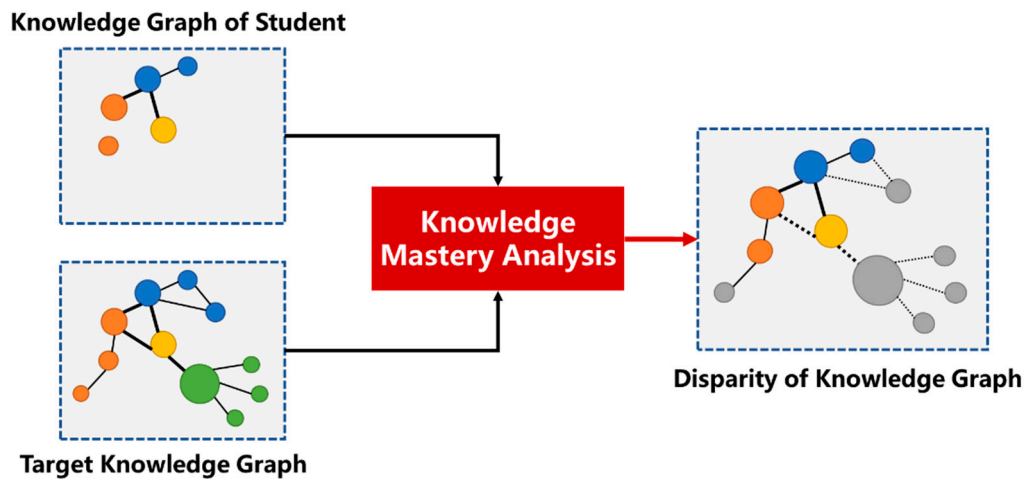
Figure 3. Disparity analysis between the knowledge graph of a student and that of an expert.

3.2. Planning of Teaching Contents and Strategies

According to the learning style and knowledge level analysis of the students, the ITR may build the student model for each student or a group of students, based on which it may predict short-term and long-term outcomes of delivering certain teaching aspects via specific teaching strategies. Then, these predictions may be fed to the decision-making process.

3.2.1. Student Model and Teaching Outcome Prediction

The ITR reasons that students cognitive and emotional information builds the information into the student model. This model includes the dimension of topics, misconceptions, affective characteristics, student experiences, and stereotypes, and may be instantly represented by operational student modules [3]. Hence, the ITR may evaluate the outcome of potential teaching actions in its own "brain" as seen in Figure 4, before performing actions in the real world [22,42].
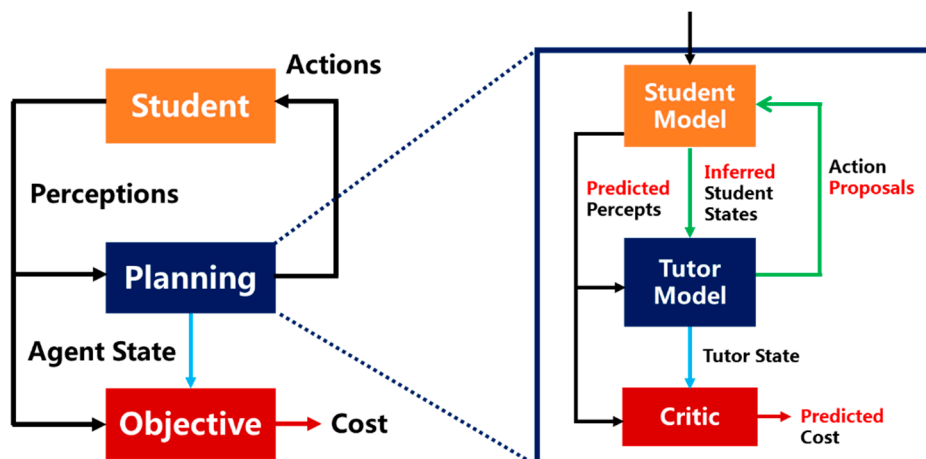

Figure 4. Illustrations of internal models of the planning module.

The analytical results from the perception module form the data basis of a student model, but the method of using these data in building a student model may be diversified. First, the model tracing method assumes that students may be modelled as rule-based agents, and the execution trace of these rules is available for the ITR to infer the student status [43]. Second, the constraint-based model method assumes that learning cannot be fully recorded and only errors can be recognized by an ITR, which may build an annotated domain model indicating the gap between students' knowledge and expert's

knowledge in the context of the curriculum, as well as a bug library indicating the misconceptions and missing knowledge of the students [44]. Third, the machine learning method avoids necessities on the full model of student behavior and such techniques used for modeling student knowledge most often include statistical inference, and have been used in the ITR to predict how and when student responds and whether the response is expected to be correct [3,45]. However, machine learning methods lack causality analysis and interpretability [20,43,45].

The distinctions between the model-tracing, constraint-based and machine learning methods are vanishing. For example, a classical model-tracing based ITR called PAT represents human declarative knowledge by modular units called chunks and human procedural knowledge by if-then production rules to capture the students' weakness and misconceptions [3]. In comparison to hand-crafted modelling in PAT, a more recent model-tracing based ITR in Reference [46] adopts a deep recursive neural network, which is a machine learning model that avoids explicit encoding of human domain knowledge and can capture complex representations of student knowledge.

Based on the built student model, the ITR may simulate, test, and predict teaching outcomes by interacting with student models. As shown in Figure 4, the student model may be instantiated as a virtual agent that reacts to ITR's actions, which allows the ITR to perform and compare different teaching actions. Based on the interactions between the student models, the ITR may adjust its strategies to the students' particular feedback.

3.2.2. Teaching Decision Making

The ITR may interact with the student's model with multiple candidate plans of teaching contents and strategies and make predictions on the outcome of each candidate plan. The next step is to compare these candidate plans and make a decision.

In general, there is no optimal plan since the solution may serve contradictory objectives in multiple dimensions. For example, the plan may select easier problems for motivating students to enjoy short-term success, while it may lead to smaller learning gains in the long run. As shown in Figure 5, the ITR has to take at least three factors into account, e.g., long-term education objective, short-term curriculum objective, and a student personal objective. The long-term education objectives include facilitating students into self-directed learners as well as equipping students with general learning tools so that the students excel in the life-long learning process. The short-term curriculum objectives usually aim to promote students to master knowledge in the context of the curriculum. The students' personal objectives may be further diversified according to their own interests [23].
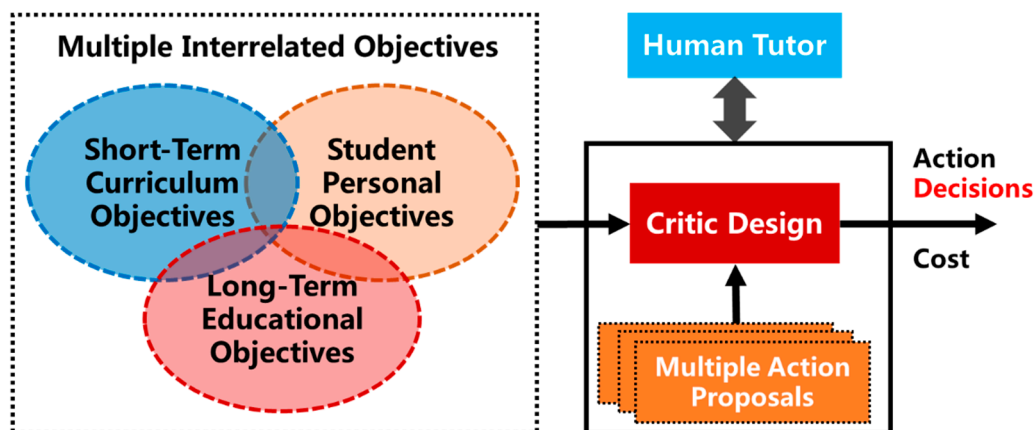

Figure 5. Critic design needs to take into account multiple interrelated objectives and may benefit from human tutor supervision. The critic may select from multiple action proposals before making a decision.

In many cases, it is a challenge for the ITR to formulate the multiple-objective optimization problems by itself. As shown in Figure 5, human tutors and curriculum designers may intervene in and design quantitative and assessable cost functions to facilitate the decision-making process for the ITR. As long as the cost function is properly designed for the criteria, the machine learning techniques may compute the decision actions from multiple action proposals. In Reference [47], the authors propose a selection algorithm for adaptive tests based on multi-criteria decision models integrating expert knowledge by fuzzy linguistic information. In Reference [48], the authors propose a four-stage decision-making plan and a synthesis mechanism for cognitive maps of knowledge diagnosis.

In general, the decision plan may be interpreted in a form of action sequence to intervene in and impact students' learning [3]. The action sequence is fed to the action modules of the ITR. The actions have multiple forms in order to adapt to a variety of teaching-learning situations. Specifically, presentation of examples reduces the cognitive load for students in complex problem-solving tasks, and timely feedback provides information for correcting students' errors and misconceptions, meanwhile it enhances motivations for higher levels of effort [23]. The specific design of action sequence relies heavily on selecting tutoring strategies. The strategy may be inferred from either mimicking human teaching such as apprenticeship [49], or a variety of learning theories such as cognitive learning [50], constructivism [51], and situated learning [52].

3.3. Action through Multi-Modal Communications

After receiving the plan result, the action module resolves the action sequence, which indicates what, when, and how teaching factors are to be delivered to the students. Even with the best students and teaching plan, an ITR is of limited value without effective communicative strategies. Therefore, the action module needs to form an appropriate scenario for the students, as well as provide effective communication channels for transmitting the teaching content and receiving feedback from students. Various techniques and demonstrations have been proposed to enable scene construction and multi-modal communication channels in the action module.

3.3.1. Scene Construction

Physical or virtual scenes generated by the ITR may allow students to immerse themselves in the learning process and use multi-modal perception to acquire information, which may lead to more effective learning. As shown in Figure 6, the virtual scene generation is of much higher flexibility and relatively low costs. Therefore, it becomes a major research direction.
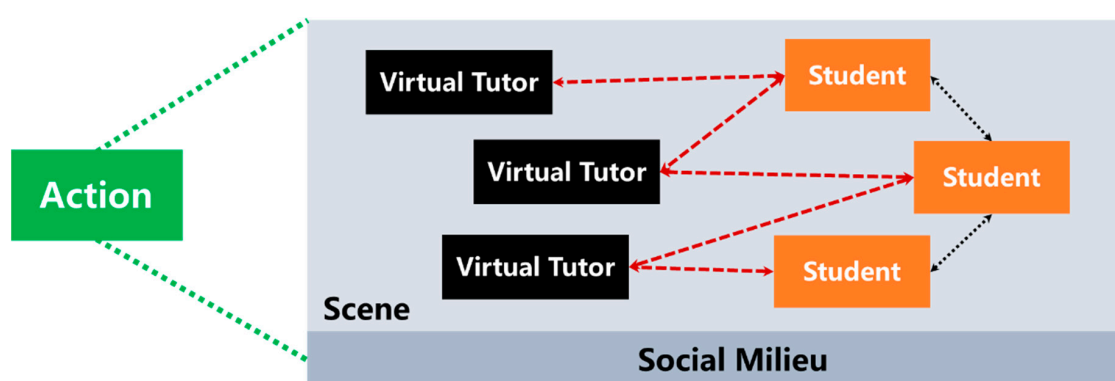

Figure 6. Action module constructs an immersive scene for students with multiple virtual tutors.

The action module may generate virtual-reality and augmented-reality environments, including the social milieu and physical milieu powered by a physical engine. In Reference [53] and Reference [54], the authors provide a thorough review of using virtual reality techniques to build virtual environments for laboratory and training facilities. Recent advances in applying augmented reality techniques in

education have been reviewed in References [55–57]. Furthermore, a comparison of the pros and cons in using virtual-reality and augmented-reality for learning system design is given by Reference [58].

Then, the action module may inhabit virtual tutors in the virtual-reality environment. The action module may imitate human tutors who have natural behavior such as natural languages, gestures, and facial expressions to interact with students [59]. Furthermore, the action module may enable collaboration and communication in ways that are impossible with traditional tutoring because of ethical issues, e.g., virtual patients [60].

It is important to note that scene construction relies heavily on the characteristics of the learner and subjects in order to provide immersive environments for the learners. Additionally, there is a natural trade-off between scene construction fidelity and costs. Therefore, scene construction usually involves case-by-case analysis and customized design. For example, the survey of virtual reality-based scene construction for ITRs in Reference [54] categorizes the applications into medical, industrial, commercial, collaborative and massive online open courses (MOOCs), serious games, and rehabilitation. The ITR used for medical [61] and industrial skills training [62] often demand haptic devices to enable real-time physical procedures, which is not always a pre-requisite in some MOOCs such as language learning [63] and mathematics [64].

3.3.2. Multi-Modal Communication Channel

Within the physical or virtual scenes generated by the ITRs, multi-modal forward channels are built for delivering the action sequence. It has been found that rather than passively receiving knowledge from the tutor, the students construct their own structures and organize their own knowledge. Therefore, the ITR should promote critical thinking, self-directed learning, and self-explanation via social communication exploiting both student affection and facial features.

By simulating human communicative strategies, graphic communication and natural language are the conventional communication channels for social communication. Although a vast number of contributions have been devoted [3,20], it is worth noting that recent work on generated adversarial networks allow high-fidelity generation of facial expression [65,66] and natural language [67,68].

The physically embodied robots may endow physical and emotional interactions, increase cognitive and affective outcomes, and achieve outcomes similar to those of human tutoring on restricted tasks [5]. The authors of References [69,70] assess the effect of the physical presence of a ITR in an automated tutoring interaction, showing that physical embodiment and personalization can yield significant benefits in educational human-robot interactions, which produces measurable learning gains. The authors of References [71,72] explore the social robot capability of improving students' curiosity, while those of Reference [73] use hints and distractions of curious facts to improve students' learning performance.

4. Intelligent Tutor Robot Design: A Case Study

In the last section, the perception-planning-action framework for ITRs is proposed, along with AI techniques applied in implementations of the perception, planning, and action modules. In this section, a case study is discussed to illustrate the practical ITR design workflow for certain learners and learning subjects. Specifically, we select Tega [71], which is a one-to-one ITR for children's second language skills. We analyze the current implementation of the Tega system in the context of the proposed perception-planning-action framework and discuss potential AI techniques to improve the Tega system.

4.1. Design Analysis of the Tega System

The tutoring goal of the Tega system is to help children learn new words in a second language, which includes a tablet to facilitate virtual scene construction and interaction, as well as a physical robot to convey physical communications, as shown in Figure 7. The perception-planning-action framework is used to decompose the Tega system and the implementation of each module is analyzed below.

- Multi-modal perception: The Tega system is equipped with a visual sensor and an acoustic sensor to capture the video and voice streams of the child, along with a tactile sensor to capture the interactions between the child and the virtual game environment over the tablet screen. A real-time facial expression detection and analysis algorithm was implemented to extract the child's emotions, e.g., smile, brow-furrow, brow-raise, and lip-depress, and to analyze the child's valence and engagement.

- Planning of teaching contents and strategies: The reinforcement learning (RL) technique is adopted to learn the personal affective policy for each child. The input of the RL is the valence and engagement status of the child as well as the child's task action within the virtual game, acquired from the multi-modal perception module. The critic design relies on measuring the reward as a weighted sum of child affective performance and task performance. The RL may then facilitate online training and adapt proper verbal and non-verbal actions through the virtual game and the physical robot. Meanwhile, the student characteristic is implicitly modelled in the RL.

- Action through multi-modal communications: The scene construction is implemented by a tablet, where a virtual traveling game is synthesized to allow the child to interact and practice with a virtual animated character. The physical robot may perform head up/down, waist-tilt left/right, waist-lean forward/back, full body up/down, and full body left/right, expressing non-verbal gestures to attract and guide the attention of the child, along with the verbal natural language utterances.
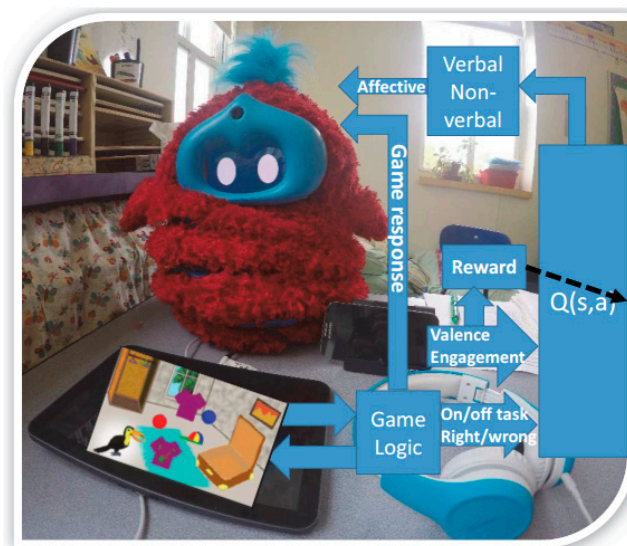

Figure 7. The implementation of the Tega system, including a tablet and a physical robot [72].

4.2. Potential Enhancement of the Tega System

It is shown that the perception-planning-action framework may be used to decompose and analyze the Tega system. Although several AI techniques have been implemented, other AI techniques in Section 3 may be adopted in the perception, planning, and action modules to improve Tega.

- Multi-modal perception: The current implementation of the Tega system uses a single visual sensor to capture and analyze the child's facial expression. It may be extended to multiple visual sensors from the physical robot and the tablet may capture the child's video streams from multiple perspectives. In this case, pixel-level data fusion may be adopted to synthesize the child's activities in the 3D environments, where both action recognition and facial expression detection may be used for improving the child's valence and engagement analysis.

- Planning of teaching contents and strategies: The current implementation of Reinforcement learning (RL) in the Tega system used case-by-case training and a fixed weighted reward. Two methods may be adopted to improve the learning and decision-making performance.

First, the Tega system may gather information learned from multiple children to generate transferable student models, which may be deployed in new Tega system to reduce training time. Second, long-term child's engagement metric may be designed and included in the critic design, which helps adapting weights in the reward function to incorporate a balanced tradeoff between short and long-term learning objectives.

- Action through multi-modal communications: The current implementation of the Tega splits the virtual actions in the 2D tablet game and physical actions by the physical robot, where the 3D actions are limited by the degree-of-freedom of the physical robot. Augmented or virtual reality techniques may be adopted to synthesize a virtual robot for more vivid actions and affections in the interactive game.

5. Discussion and Conclusions

This paper transforms the relationship model describing the teaching-learning process into the well-established perception-planning-action ITR model in the area of AI.

- The first phase of initiating an effective perception-planning-action loop of ITR is to acquire sufficient information concerning the environments and the students, so it introduces multi-modal perception that utilizes multiple communication channels. Then, the ITR leads the students to do a variety of activities for external-assessment and self-assessment, so it can gather information, and then implement statistical machine learning methods to analyze the students' learning style and knowledge level.
- According to students' learning style and knowledge level, the second phase of ITR builds student models for each student or a group of students, based on which it can predict short-term and long-term outcomes of delivering certain teaching contents via specific teaching strategies. Then, the ITR may interact with the student models with multiple candidate plans of teaching contents and strategies and make predictions on the outcome of each candidate plan. Human tutors and curriculum designers may also intervene in the decision-making process.
- After receiving the plan result, the third phase is activated by the action module of the ITR that resolves the action sequence indicating what, when, and how teaching contents are to be delivered to the students. The action module needs to form an appropriate scenario for the students, as well as to provide effective communication channels for transmitting the teaching contents and receiving the feedback from students, using scene construction and multi-modal communication channels in the action module.
- With the rapid progress of AI techniques, many open research areas may be defined for ITRs with the aid of the perception-planning-action model.
- Perception: Multi-modal data fusion is not fully researched in the context of ITR design, even though the topic is well-investigated and still in heated discussion in the field of data-mining and robotics. Additionally, little work has compared the effects of different factors that indicate learning styles. Hence, it remains an open question on how to select the appropriate set of variables to predict a students' learning process.
- Planning: In the context of student modelling, the model tracing, constrained model, and machine learning methods have their own application scenarios and limitations. Therefore, the recent advances of explainable AI may be incorporated in ITRs. In terms of the decision-making process, limited research has contributed to the problem formulation that takes into account the multi-objective problems such as jointly considering the long-term education objective, short-term curriculum objective, and student personal objective.
- Action: The action module field turns out to be the most researched area with state-of-art AI techniques. Even so, the applications of advanced AI techniques in ITRs such as generated adversarial networks are underway, while the design of physically embodied robots is far from mature.

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef]
2. IFR, Statistical Department. *World Robotics Survey*; IFR: Frankfurt, Germany, 2008.
3. Woolf, B.P. *Building Inteligent Interactive Tutors, Student-Centered Strategies for Revolutionizing E-Learning*; Morgan Kaufman: Burlington, MA, USA, 2008.
4. Truong, H.M. Computers in Human Behavior Integrating Learning Styles and Adaptive E-Learning System: Current Developments, Problems and Opportunities. *Comput. Hum. Behav.* **2016**, *55*, 1185–1193. [CrossRef]
5. Belpaeme, T.; Kennedy, J.; Ramachandran, A.; Scassellati, B.; Tanaka, F. Social Robots for Education: A Review. *Sci. Robot.* **2018**, *3*, eaat5954. [CrossRef]
6. Lopez, T.; Chevaillier, P.; Gouranton, V.; Evrard, P.; Nouviale, F.; Barange, M.; Bouville, R.; Arnaldi, B. Collaborative virtual training with physical and communicative autonomous agents. *Comput. Anim. Virtual Worlds* **2014**, *25*, 485–493. [CrossRef]
7. Johnson, W.L.; Lester, J.C. Face-to-face interaction with pedagogical agents, twenty years later. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 25–36. [CrossRef]
8. Papamitsiou, Z.; Economides, A.A. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *J. Educ. Technol. Soc.* **2014**, *17*, 49–64.
9. Wilson, C.; Scott, B. Adaptive systems in education: A review and conceptual unification. *Int. J. Inf. Learn. Technol.* **2017**, *34*, 2–19. [CrossRef]
10. Mubin, O.; Stevens, C.J.; Shahid, S.; Al Mahmud, A.; Dong, J.J. A Review of the Applicability of Robots in Education. *Technol. Educ. Learn.* **2013**, *1*. [CrossRef]
11. Dewey, J. My Pedagogic Creed. *Sch. J.* **1897**, *54*, 77–80.
12. Schwab, J.J. *Science, Curriculum, and Liberal Education*; University of Chicago Press: Chicago, IL, USA, 1978.
13. Novak, J.D.; Gowin, D.B. *Learning How to Learn*; Cambridge University Press: Boston, MA, USA, 1984.
14. Howard, T.C.; Aleman, G.R. Teacher capacity for diverse learners: What do teachers need to know? In *Handbook of Research on Teacher Education: Enduring Questions in Changing Contexts*, 3rd ed.; Cochran-Smith, M., Feiman-Nemser, S., McIntyre, D.J., Demers, K.E., Eds.; Routledge: Abingdon, UK, 2008; pp. 157–174.
15. Krim, J.S. Critical Reflection and Teacher Capacity: The Secondary Science Pre-Service Teacher Population. Ph.D. Thesis, Montana State University, Bozeman, MT, USA, July 2009.
16. Shulman, L.S. Knowledge and teaching: Foundations of the new reform. *Harv. Educ. Rev.* **1987**, *57*, 1–23. [CrossRef]
17. Collinson, V. *Reaching Students: Teachers Ways of Knowing*; Corwin Press, Inc.: Thousand Oaks, CA, USA, 1996.
18. McDiarmid, G.W.; Clevenger-Bright, M. Rethinking teacher capacity. In *Handbook of Research on Teacher Education: Enduring Questions in Changing Contexts*, 3rd ed.; Cochran-Smith, M., Feiman-Nemser, S., McIntyre, D.J., Demers, K.E., Eds.; Routledge: Abingdon, UK, 2008; pp. 134–156.
19. Turner-Bisset, R. *Expert Teaching: Knowledge and Pedagogy to Lead the Profession*; David Fulton Publishers: London, UK, 2001.
20. Wenger, E. *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*; Morgan Kaufmann: Burlington, MA, USA, 2014.
21. Grossman, P.L. *The Making of a Teacher: Teacher Knowledge and Teacher Education*; Teachers College Press: New York, NY, USA, 1990.
22. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Pearson Education Limited: Kuala Lumpur, Malaysia, 2016.
23. Ambrose, S.A.; Bridges, M.W.; DiPietro, M.; Lovett, M.C.; Norman, M.K. *How Learning Works: Seven Research-Based Principles for Smart Teaching*; John Wiley & Sons: Hoboken, NJ, USA, 2010.

24. Hall, D.L.; Llinas, J. An introduction to multi sensor data fusion. *Proc. IEEE* **2002**, *85*, 6–23. [CrossRef]

25. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [CrossRef]

26. Ritschel, H. Socially-aware reinforcement learning for personalized human-robot interaction. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, Stockholm, Sweden, 10–15 July 2018; pp. 1775–1777.

27. Tsiami, A.; Koutras, P.; Efthymiou, N.; Filntisis, P.P.; Potamianos, G.; Maragos, P. Multi3: Multi-Sensory Perception System for Multi-Modal Child Interaction with Multiple Robots. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–26 May 2018; pp. 1–8.

28. Chen, C.; Jafari, R.; Kehtarnavaz, N. A survey of depth and inertial sensor fusion for human action recognition. *Multimed. Tools Appl.* **2017**, *76*, 4405–4425. [CrossRef]

29. Kose, H.; Akalin, N.; Yorganci, R.; Ertugrul, B.S.; Kivrak, H.; Kavak, S.; Ozkul, A.; Gurpinar, C.; Uluer, P.; Ince, G. iSign: An architecture for humanoid assisted sign language tutoring. In *Intelligent Assistive Robots*; Springer: Cham, Switzerland, 2015; pp. 157–184.

30. Poria, S.; Cambria, E.; Bajpai, R.; Hussain, A. A review of affective computing: From unimodal analysis to multimodal fusion. *Inf. Fusion* **2017**, *37*, 98–125. [CrossRef]

31. Spaulding, S.; Gordon, G.; Breazeal, C. Affect-aware student models for robot tutors. In Proceedings of the International Conference on Autonomous Agents & Multiagent Systems, Singapore, 9–13 May 2016; pp. 864–872.

32. Park, H.W.; Gelsomini, M.; Lee, J.J.; Breazeal, C. Telling stories to robots: The effect of backchanneling on a child's storytelling. In Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2016; pp. 100–108.

33. Sharma, S.; Ryan, K.; Ruslan, S. Action Recognition Using Visual Attention. 2015. Available online: https://arxiv.org/abs/1511.04119 (accessed on 12 December 2018).

34. Lu, J.; Xiong, C.; Parikh, D.; Socher, R. Knowing when to look: Adaptive attention via a visual sentinel for image captioning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; Volume 6, p. 2.

35. Felder, R.M.; Silverman, L.K. Learning and teaching styles in engineering education. *Eng. Educ.* **1988**, *78*, 674–681.

36. Chou, C.Y.; Tseng, S.F.; Chih, W.C.; Chen, Z.H.; Chao, P.Y.; Lai, K.R.; Chan, C.L.; Yu, L.C.; Lin, Y.L. Open student models of core competencies at the curriculum level: Using learning analytics for student reflection. *IEEE Trans. Emerg. Top. Comput.* **2015**, *5*, 32–44. [CrossRef]

37. Epp, C.D.; Bull, S. Uncertainty Representation in Visualizations of Learning Analytics for Learners: Current Approaches and Opportunities. *IEEE Trans. Learn. Technol.* **2015**, *8*, 242–260.

38. Amadieu, F.; van Gog, T.; Paas, F.; Tricot, A.; Mariné, C. Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learn. Instr.* **2009**, *19*, 376–386. [CrossRef]

39. Ehrlinger, L.; Wöß, W. Towards a Definition of Knowledge Graphs. In Proceedings of the 12th International Conference on Semantic Systems (SEMANTiCS 2016), Leipzig, Germany, 12–15 September 2016.

40. Battaglia, P.W.; Hamrick, J.B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv* **2018**, arXiv:1806.01261.

41. Zhang, D.; Yin, J.; Zhu, X.; Zhang, C. Network representation learning: A survey. *IEEE Trans. Big Data* **2018**, arXiv:1801.05852. [CrossRef]

42. Zhang, B.; Chen, B.; Yang, J.; Yang, W.; Zhang, J. An Unified Intelligence-Communication Model for Multi-Agent System Part-I: Overview. *arXiv* **2018**, arXiv:1811.09920.

43. Li, Q.; Lau, R.W.; Popescu, E.; Rao, Y.; Leung, H.; Zhu, X. Social Media for Ubiquitous Learning and Adaptive Tutoring [Guest editors' introduction]. *IEEE MultiMed.* **2016**, *23*, 18–24. [CrossRef]

44. Mitrovic, A. Fifteen years of constraint-based tutors: What we have achieved and where we are going. *User Model. User-Adapt. Interact.* **2012**, *22*, 39–72. [CrossRef]

45. Abyaa, A.; Idrissi, M.K.; Bennani, S. Learner modelling: Systematic review of the literature from the last 5 years. *Educ. Technol. Res. Dev.* **2019**. [CrossRef]

46. Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L.J.; Sohl-Dickstein, J. Deep knowledge tracing. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2015; pp. 505–513.

47. Badaracco, M.; MartíNez, L. A fuzzy linguistic algorithm for adaptive test in Intelligent Tutoring System based on competences. *Expert Syst. Appl.* **2013**, *40*, 3073–3086. [CrossRef]

48. Uglev, V. Implementation of Decision-making Methods in Intelligent Automated Educational System Focused on Complete Individualization in Learning. *AASRI Procedia* **2014**, *6*, 66–72. [CrossRef]

49. Maclellan, C.J.; Harpstead, E.; Patel, R.; Koedinger, K.R. The Apprentice Learner architecture: Closing the loop between learning theory and educational data. In Proceedings of the 9th International Conference on Educational Data Mining, Raleigh, NC, USA, 29 June–2 July 2016; pp. 151–158.

50. Cao, S.; Qin, Y.; Zhao, L.; Shen, M. Modeling the development of vehicle lateral control skills in a cognitive architecture. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *32*, 1–10. [CrossRef]

51. Bremgartner, V.; de Magalhães Netto, J.F.; de Menezes, C.S. Adaptation resources in virtual learning environments under constructivist approach: A systematic review. In Proceedings of the 2015 IEEE Frontiers in Education Conference (FIE), Washington, DC, USA, 21–24 October 2015; pp. 1–8.

52. Ketelhut, D.J.; Dede, C.; Clarke, J.; Nelson, B. Studying Situated Learning in a Multi-User Virtual Environment. In *Assessment of Problem Solving Using Simulations*; Baker, E., Dickieson, J., Wulfeck, W., O'Neil, H., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2007; pp. 37–58.

53. Potkonjak, V.; Gardner, M.; Callaghan, V.; Mattila, P.; Guetl, C.; Petrović, V.M.; Jovanović, K. Virtual laboratories for education in science, technology, and engineering: A review. *Comput. Educ.* **2016**, *95*, 309–327. [CrossRef]

54. Vaughan, N.; Gabrys, B.; Dubey, V.N. An overview of self-adaptive technologies within virtual reality training. *Comput. Sci. Rev.* **2016**, *22*, 65–87. [CrossRef]

55. Bacca, J.; Baldiris, S.; Fabregat, R.; Graf, S. Augmented reality trends in education: A systematic review of research and applications. *Educ. Technol. Soc.* **2014**, *17*, 133–149.

56. Radu, I. Augmented reality in education: A meta-review and cross-media analysis. *Pers. Ubiquitous Comput.* **2014**, *18*, 1533–1543. [CrossRef]

57. Akçayır, M.; Akçayır, G. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educ. Res. Rev.* **2017**, *20*, 1–11. [CrossRef]

58. Harley, J.M.; Poitras, E.G.; Jarrell, A.; Duffy, M.C.; Lajoie, S.P. Comparing virtual and location-based augmented reality mobile learning: Emotions and learning outcomes. *Educ. Technol. Res. Dev.* **2016**, *64*, 359–388. [CrossRef]

59. Robb, A.; Kopper, R.; Ambani, R.; Qayyum, F.; Lind, D.; Su, L.M.; Lok, B. Leveraging virtual humans to effectively prepare learners for stressful interpersonal experiences. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 662–670. [CrossRef]

60. Conradi, E.; Kavia, S.; Burden, D.; Rice, A.; Woodham, L.; Beaumont, C.; Savin-Baden, M.; Poulton, T. Virtual patients in a virtual world: Training paramedic students for practice. *Med. Teach.* **2009**, *31*, 713–720. [CrossRef] [PubMed]

61. Wijewickrema, S.; Copson, B.; Ma, X.; Briggs, R.; Bailey, J.; Kennedy, G.; O'Leary, S. Development and Validation of a Virtual Reality Tutor to Teach Clinically Oriented Surgical Anatomy of the Ear. In Proceedings of the 31st International Symposium on Computer-Based Medical Systems, Karlstad, Sweden, 18–21 June 2018; pp. 12–17.

62. Gavish, N.; Gutiérrez, T.; Webel, S.; Rodríguez, J.; Peveri, M.; Bockholt, U.; Tecchia, F. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interact. Learn. Environ.* **2015**, *23*, 778–798. [CrossRef]

63. Hassani, K.; Nahvi, A.; Ahmadi, A. Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interact. Learn. Environ.* **2016**, *24*, 252–271. [CrossRef]

64. Patterson, R.L.; Patterson, D.C.; Robertson, A.M. Seeing Numbers Differently: Mathematics in the Virtual World. In *Emerging Tools and Applications of Virtual Reality in Education*; IGI Global: Hershey, PA, USA, 2016; pp. 186–214.

65. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8789–8797.

66. Nojavanasghari, B.; Huang, Y.; Khan, S. Interactive Generative Adversarial Networks for Facial Expression Generation in Dyadic Interactions. *arXiv* **2018**, arXiv:1801.09092.

67. Press, O.; Bar, A.; Bogin, B.; Berant, J.; Wolf, L. Language generation with recurrent generative adversarial networks without pre-training. *arXiv* **2017**, arXiv:1706.01399.

68. Shetty, R.; Rohrbach, M.; Hendricks, A.L.; Fritz, M.; Schiele, B. Speaking the same language: Matching machine to human captions by adversarial training. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4135–4144.

69. Leyzberg, D.; Spaulding, S.; Toneva, M.; Scassellati, B. The physical presence of a robot tutor increases cognitive learning gains. In Proceedings of the Annual Meeting of the Cognitive Science Society, Sapporo, Japan, 1–4 August 2012; p. 34.

70. Leyzberg, D.; Spaulding, S.; Scassellati, B. Personalizing robot tutors to individuals' learning differences. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, Bielefeld, Germany, 3–6 March 2014; pp. 423–430.

71. Gordon, G.; Breazeal, C.; Engel, S. Can children catch curiosity from a social robot? In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015; pp. 91–98.

72. Gordon, G.; Spaulding, S.; Westlund, J.K.; Lee, J.J.; Plummer, L.; Martinez, M.; Das, M.; Breazeal, C. Affective Personalization of a Social Robot Tutor for Children's Second Language Skills. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 3951–3957.

73. Blancas-Muñoz, M.; Vouloutsi, V.; Zucca, R.; Mura, A.; Verschure, P.F. Hints vs Distractions in Intelligent Tutoring Systems: Looking for the proper type of help. *arXiv* **2018**, arXiv:1806.07806.