3 Introducing AI in education

3.1 *Introduction*

Having established a working understanding of AI, we will now look in more detail at how AI works in educational contexts, beginning with a brief history. However, what won't be discussed in this [chapter] but what might nonetheless have a major impact on education is the use of AI to support school administration (i.e. i.e. system-facing AI that addresses things like: class timetabling, staff scheduling, facilities management, finances, cybersecurity, safety and security...)[200]. ***Our focus is the use of AI to support learning (student- and teacher-facing AI).***

3.2 *A brief history of AI in education*

3.2.1 *Setting the stage*

Precursors of the application of AI in education can be found in the work of the psychologists Sidney Pressey, who was a professor at Ohio State University in the 1920s, and B. F. Skinner, known as the father of behaviourism, who was a professor at Harvard University from 1948 until his retirement in 1974. For Pressey, the challenge was to leverage the potential of multiple-choice tests to consolidate student learning as well as to evaluate it. Drawing on Edward Thorndike's 'law of effect'[201], he argued that, for tests to support learning, immediate feedback was essential - which is not usually possible when tests are marked by hand. However, a mechanical approach could ensure that no learning opportunities were missed:

> *Devices which at once inform a student about the correctness of his answer to a question , and then lead him to the right answer, clearly do more than test him; they also teach him*. – Pressey[202]

---

[200] Readers who are interested in the use of AI technologies to support administrative functions might like to read about the the UK's school inspection service Ofsted's use of "artificial-intelligence algorithm to predict which schools are 'less than good'", https://www.tes.com/news/ofsted-use-artificial-intelligence-algorithm-predict-which-schools-are-less-good.

[201] "What comes after a connection acts upon it to alter its strength." Edward L. Thorndike, 'The Law of Effect', *The American Journal of Psychology* 39, no. 1/4 (1927): 212–22, https://doi.org/10.2307/1415413.

[202] S. L. Pressey, 'Development and Appraisal of Devices Providing Immediate Automatic Scoring of Objective Tests and Concomitant Self-Instruction.', *Journal of Psychology; Provincetown, Mass., Etc.* 30 (1 January 1950): 417–447.

Pressey made various versions of his machine (and made several unsuccessful attempts to commercialise his idea), the most sophisticated being based on a mechanical typewriter . Inside this device was a rotating drum around which was wrapped a card printed with a list of questions and hole-punched (much like the perforated rolls used in self-playing pianos) to represent the correct answers. Meanwhile, the casing featured a small window, which showed the number of the current question, and five typewriter keys, one for each possible answer. As the student worked through a printed sheet of questions and answers, they would press one of the keys on the device to select their answer for each question. The machine was configured so that the student immediately knew whether they had made the right choice, and it prevented them from moving onto the next question until they had.

Interestingly, Pressey was also one of the first to make the case that, in addition to supporting learning, a teaching machine could make a teacher's life easier and more fulfilling – by relieving them of one of their least interesting tasks (marking tests) and giving them more time to engage with their students:

> *Lift from her* [the teacher's] *shoulders as much as possible of this burden and make her free for those inspirational and thought-stimulating activities which are, presumably, the real function of the teacher.* – Pressey[203]

Pressey's approach was later extended by Skinner, who argued that the techniques he had pioneered for training rats and pigeons (in operant conditioning chambers now known as Skinner Boxes) might be adapted for teaching people. Skinner's teaching machine, which he devised in 1958, was a wooden box with a windowed lid. Questions written on paper disks appeared in one window, and the student wrote a response on a roll of paper accessible through a second window (for later marking by a teacher). Advancing the mechanism automatically covered the student's answer, so that it could not be changed, and simultaneously revealed the correct answer. In this way, Skinner's teaching machine provided automatic, immediate reinforcement. Students were required to compose their own answers, rather than choose from a limited selection (as with Pressey's multiple-choice questions), because Skinner had found that learning is more effectively reinforced by *recalling* a correct response than by simply recognising it. This approach also gave the student the opportunity to

[203] S. L. Pressey, 'A Simple Device for Teaching, Testing, and Research in Learning', *School and Society* 23 (1926): 374.

compare their answer with the given model answer, which if properly designed by the teacher and actively undertaken by the student could also contribute to learning.

Interestingly, Skinner argued that his teaching machine in effect acted like a personal tutor, foreshadowing AI in education's *Intelligent Tutoring Systems*:

> *The machine itself, of course, does not teach ... but the effect upon each student is surprisingly like that of a private tutor.... (i) There is a constant interchange between program and student.... (ii) Like a good tutor, the machine insists that a given point be thoroughly understood ... before the student moves on.... (iii) Like a good tutor, the machine presents just that material for which the student is ready.... (iv) Like a skilful tutor, the machine helps the student to come up with the right answer.... (v) Lastly, of course, the machine, like the private tutor, reinforces the student for every correct response, using this immediate feedback ... to shape his behavior most efficiently.*[204]

Skinner's teaching machine might be thought to have also foreshadowed something else later taken up by AI in education researchers, dividing automated teaching into separate components (in Skinner's case, distinguishing between the subject content, which was pre-programmed into the machine, and the student's achievements, whether or not they had answered a question correctly). However, although in a sense Skinner's teaching machine was responsive to individual students, it could not be considered 'adaptive'. That is to say, it did not adapt either the questions, or the order in which they were presented, according to the achievements or needs of the individual students. Instead, question delivery was pre-scripted. While a student could proceed at their own pace, they went through the same list of questions as every other student and in the same order.

3.2.2 *Adaptive learning*

Also working in the 1950s, Norman Crowder, who was interested in communication rather than psychology, devised a paper-based alternative to the early teaching machines, known as intrinsic or branching programmed instruction[205]. In Crowder's system (which he developed for

[204] B. F. Skinner, 'Teaching Machines', *Science* 128, no. 3330 (1958): 969–77.

[205] Norman A. Crowder, 'Automatic Tutoring by Means of Intrinsic Programming', in *Teaching Machines and Programmed Learning: A Source Book*, ed. Arthur A. Lumsdaine and Robert Glaser, vol. 116 (Washington, D.C.: DAVI, 1960), 286–98.

training U.S. Air Force engineers to find malfunctions in electronic equipment), the user is presented with a short page of information followed by a multiple-choice question, with each possible answer directing the student to a new page. If the correct answer was chosen, the new page would present new information, building upon that already correctly understood; if an incorrect answer was chosen, the new page would contain feedback designed to help the student understand the cause of their error, based on what the student had chosen. The system might also branch through one or two additional pages of corrective materials before returning the student back to the main pages. In short, Crowder's system adapted the *pathway* through the teaching materials according to the individual student's developing knowledge, such that each student might see quite different sets of pages.

However, probably the first truly adaptive *teaching machine* was developed, again beginning in the early 1950s, by a British polymath, Gordon Pask. Known as SAKI (the self-adaptive keyboard instructor), it was designed for trainee keyboard operators learning how to use a device that punched holes in cards for data processing[206]. What distinguished SAKI from the other early teaching machines was that the task presented to a learner was *adapted* to the learner's individual performance, which was represented in a continuously changing probabilistic student 'model'.

When you interact with the system, learning which keys represent which numbers:

> *the machine is measuring your responses, and building its own probabilistic model of your learning process. That "7," for instance, you now go to straight away. But the "3," for some obscure reason, always seems to elude you. The machine has detected this, and has built the facts into its model. And now, the outcome is being fed back to you. Numbers with which you have difficulty come up with increasing frequency in the otherwise random presentation of digits. They come up more slowly, too, as if to say: "Now take your time." The numbers you find easy, on the contrary, come up much faster: the speed with which each number is thrown at you is a function of the state of your learning.*[207]

[206] Gordon Pask, 'SAKI: Twenty-Five Years of Adaptive Training into the Microprocessor Era', *International Journal of Man-Machine Studies* 17, no. 1 (1 July 1982): 69–74, https://doi.org/10.1016/S0020-7373(82)80009-6.
[207] Stafford Beer, *Cybernetics and Management* (London: The English Universities Press, 1960), 124.

3.2.3 *Computer-aided instruction*

SAKI went through many iterations, taking advantage of developments in computers and the new microprocessors, and was one of the first adaptive systems to be commercialised. However, over the following years, other than in the various iterations of SAKI, adaptive learning made few advances, and the focus shifted to what became known as computer-aided instruction (CAI) systems. The 1960s and 1970s saw many CAI systems being built, an early influential example being PLATO (Programmed Logic for Automatic Teaching Operations), which was developed at the University of Illinois. PLATO involved students accessing standard teaching materials, some of which were interactive, on a central mainframe computer via remote terminals, with as many as a thousand students working at the same time.

This system was also notable for being the first to introduce in an educational technology many tools and approaches still common today, such as user forums, e-mail, instant messaging, remote screen-sharing and multiplayer games. Around the same time, Stanford University and IBM developed a computer-aided instruction system that was made available again via remote terminals to a few local elementary schools. This system involved a linear presentation of teaching materials, for mathematics and language arts, together with drill and practice activities. A third prominent example was TICCIT (Time-shared Interactive Computer-Controlled Information Television), developed by Brigham Young University, which was used to teach freshman-level mathematics, chemistry, physics, English, and various language courses. Each subject area was broken down into topics and learning objectives, which in turn were represented as screens of information. TICCIT then provided a predetermined sequence, although learners could also use the keyboard to navigate through the screens in any order that they found helpful.

Although in other ways successful, during the 1960s and '70s only very few of these CAI systems were widely adopted, mainly due to the cost and accessibility of the university mainframes that were needed to host the software. The arrival of personal computers in the 1980s changed everything, with the number of CAI programmes quickly mushrooming. Very soon, CAI programmes addressing every aspect of learning were being widely used in schools, universities and family homes. Nonetheless, of particular relevance for our present purposes, *almost all of these systems were severely hampered by the same flaw – a lack of adaptivity*. The sequence of topics, the information provided, and the system's response to student actions, were predefined and the same for each student, ignoring the individual student's successes, misconceptions, and interests, and thus inhibiting their potential learning.

3.2.4 *The introduction of AI to CAI*

At around the same time that mainstream CAI was developing, other researchers, such as John Self[208] and William Clancey[209], were wondering how CAI could become adaptive to the needs of individual students, and whether the application of techniques from AI might be useful. The first implementation of AI techniques in computer-aided instruction is usually credited to Jaime Carbonell, who, in his 1970 PhD thesis, introduced a system called SCHOLAR. This novel system was designed to enable a student to review their knowledge of the geography of South America, by means of a dialogue using a subset of English. SCHOLAR was qualitatively different from standard CAI because, uniquely for the time, it was capable of generating individual responses to student statements by drawing from a *semantic network* (a way of representing knowledge, in this case knowledge of geography, in which concepts are linked according to their semantic relationships).

> *In* [a] *conventional CAI, the data base consists of many "frames" of specific pieces of text, questions, and anticipated answers entered in advance by the teacher. By contrast,* [SCHOLAR] *is based on the utilization of an information network of facts, concepts, and procedures; it can generate text, questions, and corresponding answers. Because* [it] *can also utilize its information network to answer questions formulated by the student, a mixed-initiative dialogue between student and computer is possible with questions and answers from both sides.*[210]

A SCHOLAR dialogue[211] might look something like this:

> Scholar: *Use one of the following: Sucre, French, Argentina, Spanish, to answer the question: What is the language in Chile?*

---

[208] John A. Self, 'Student Models in Computer-Aided Instruction', *International Journal of Man-Machine Studies* 6, no. 2 (1 March 1974): 261–76, https://doi.org/10.1016/S0020-7373(74)80005-2.

[209] William J. Clancey, 'GUIDON', *Journal of Computer-Based Instruction* 10, no. 1–2 (1983): 8–15.

[210] J. R. Carbonell, 'AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction', *IEEE Transactions on Man-Machine Systems* 11, no. 4 (December 1970): 190–202, https://doi.org/10.1109/TMMS.1970.299942.

[211] J. R. Carbonell, 'AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction', p192.

Scholar: *You are taking too much time...*

Student: **Spanich**

Scholar: *You should have written "Spanish". Very good.*

Although by today's standards the dialogue (and especially its pedagogy) appears primitive, SCHOLAR is usually considered to be the first example of what came to be known as *Intelligent Tutoring Systems*, to which we turn next.

4 Applications of AI in education

4.1 Intelligent Tutoring Systems

> *Under the best learning conditions we can devise (tutoring), the average student is 2 sigma above the average control student taught under conventional group methods of instruction. The tutoring process demonstrates that most of the students do have the potential to reach this high level of learning. I believe an important task of research and instruction is to seek ways of accomplishing this under more practical and realistic conditions than the one-to-one tutoring, which is too costly for most societies to bear on a large scale. This is the "2 sigma" problem.*[212]

So-called Intelligent Tutoring Systems (or ITS) are among the most common applications of AI in education (in any case, as we have seen, they have probably been around the longest). Generally speaking, ITS provide step-by-step tutorials, individualised for each student, through topics in well-defined structured subjects such as mathematics or physics. Drawing on expert-knowledge about the subject and about pedagogy, and in response to the individual student's misconceptions and successes, the system determines an optimal step-by-step pathway through the learning materials and activities. As the student proceeds, the system automatically adjusts the level of difficulty and provides hints or guidance, all of which aim to ensure that the student is able to learn the given topic effectively.

[212] Bloom, Benjamin S. 1984. 'The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring'. *Educational Researcher* 13 (6): p4. Note, however, that according to VanLehn, "human tutors are 0.79 sigma more effective than no tutoring and not the 2.0 sigma found in the Bloom (1984) studies" VanLehn, Kurt. 2011. 'The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems'. *Educational Psychologist* 46 (4): p209. https://doi.org/10.1080/00461520.2011.611369.

ITSs come in many shapes although typically they involve several AI models, an approach that we will unpack here. As we saw in our earlier discussion of AI technologies, AI models are highly-simplified computational representations (in semantic networks, as used by SCHOLAR, in *ontologies*[213], or in *knowledge graphs*[214]) of specific knowledge about the real world (just like a model car is a simplified representation of a real car). The models used by ITS represent knowledge specific to teaching and learning: typically, knowledge about the topic to be learned is represented in what is known as a *domain* model, knowledge about effective approaches to teaching is represented in a *pedagogical* model, and knowledge about the student is represented in a *learner* model[215]. The ITS algorithm draws on these three models in order to adapt a sequence of learning activities for each individual student. A fourth model found in some ITS is the *open learner* model, to which we will return later.

4.1.1 *The domain model*

A *domain model* represents knowledge about the subject that the ITS aims to help the students learn (much like the subject knowledge in a standard, non-educational, expert system). This might, for example, be knowledge about mathematical procedures, genetic inheritance or the causes of World War I. In fact, over the years, mathematics for primary and secondary school students has dominated ITS. Mathematics, along with physics and computer science, are AIED's low-hanging fruits because they are, at least at a basic level, well-structured and clearly defined.

4.1.2 *The pedagogy model*

The ITS pedagogy model represents knowledge about effective approaches to teaching and learning that have been elicited from teaching experts and from research in the learning sciences (although it should be acknowledged that some ITS developers falsely assume that

[213] Ontologies are a way of representing a domain's concepts, data, components, entities and properties, and the relationships between them. Sowa, John F. 1995. 'Top-Level Ontological Categories'. *International Journal of Human-Computer Studies* 43 (5): 669–85. https://doi.org/10.1006/ijhc.1995.1068.

[214] Knowledge graphs are an alternative approach to ontologies, https://ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph

[215] Luckin et al., 'Intelligence Unleashed. An Argument for AI in Education.', 18. Boulay, Benedict du, Alexandra Poulovassilis, Wayne Holmes, and Manolis Mavrikis. 2018. 'What Does the Research Say about How Artificial Intelligence and Big Data Can Close the Achievement Gap?', 4. In *Enhancing Learning and Teaching with Technology*, edited by Rose Luckin, 316–27. London: Institute of Education Press.

they have sufficient expertise in pedagogy[216]). Pedagogical knowledge that has been represented in many ITS include knowledge of instructional approaches[217], the zone of proximal development[218], interleaved practice[219], cognitive load[220], and formative feedback[221]. For example, a pedagogical model that implements Vygotsky's zone of proximal development will ensure that activities provided by the system to the student are neither too easy nor too challenging, one that implements individualised formative feedback will ensure that feedback is provided to the student whenever it might support the student's learning.

4.1.3 *The learner model*

As we have seen, some CAI effectively (although usually by another name) implemented versions of both domain and pedagogical models: knowledge of what was to be learned and knowledge of how to teach what was to be learned (for example, using linear or branching programmed instruction). However, what distinguishes AI-driven ITSs is that, as introduced ~~foreshadowed~~ by Pask's SAKI, they also include a learner model: "a representation of the hypothesized knowledge state of the student"[222]. In fact, many ITS incorporate a wide range of knowledge about the student – such as their interactions, material that has challenged the student, their misconceptions, and their emotional states while using the system – all of which can be used to inform what is being taught and how, together with what support needs to be provided and when. In fact, most ITSs go much further. The knowledge stored about the individual student is augmented with knowledge of all the students who have used the system so far, from which the system machine learns in order to predict which pedagogical approach

[216] For example, many ITS set out to address student "learning styles" (Kumar, Amit, Ninni Singh, and Neelu Jyothi Ahuja. 2017. 'Learning Styles Based Adaptive Intelligent Tutoring Systems: Document Analysis of Articles Published Between 2001. and 2016.' *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)* 5 (2): 83–98. https://doi.org/10.5937/IJCRSEE1702083K) a construct that has been widely discredited (e.g., Kirschner, Paul A. 2017. 'Stop Propagating the Learning Styles Myth'. *Computers & Education* 106 (March): 166–71. https://doi.org/10.1016/j.compedu.2016.12.006).

[217] Carl Bereiter and Marlene Scardamalia, 'Intentional Learning as a Goal of Instruction', *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, 1989, 361–392.

[218] Vygotsky, *Mind in Society*.

[219] Doug Rohrer and Kelli Taylor, 'The Shuffling of Mathematics Problems Improves Learning', *Instructional Science* 35, no. 6 (6 October 2007): 481–98, https://doi.org/10.1007/s11251-007-9015-8.

[220] R.E. Mayer and R. Moreno, 'Nine Ways to Reduce Cognitive Load in Multimedia Learning', *Educational Psychologist* 38, no. 1 (2003): 43–52.

[221] V. J. Shute, 'Focus on Formative Feedback', *Review of Educational Research* 78, no. 1 (1 March 2008): 153–89, https://doi.org/10.3102/0034654307313795.

[222] Self, 'Student Models in Computer-Aided Instruction'.

and which domain knowledge is appropriate for any particular student at any specific stage of their learning. It is the learner model that enables ITS to be adaptive, and the machine learning that makes this adaptivity especially powerful.

4.1.4 A typical ITS architecture

Figure 6 shows how the domain, pedagogy and learner models might be connected in a typical ITS.

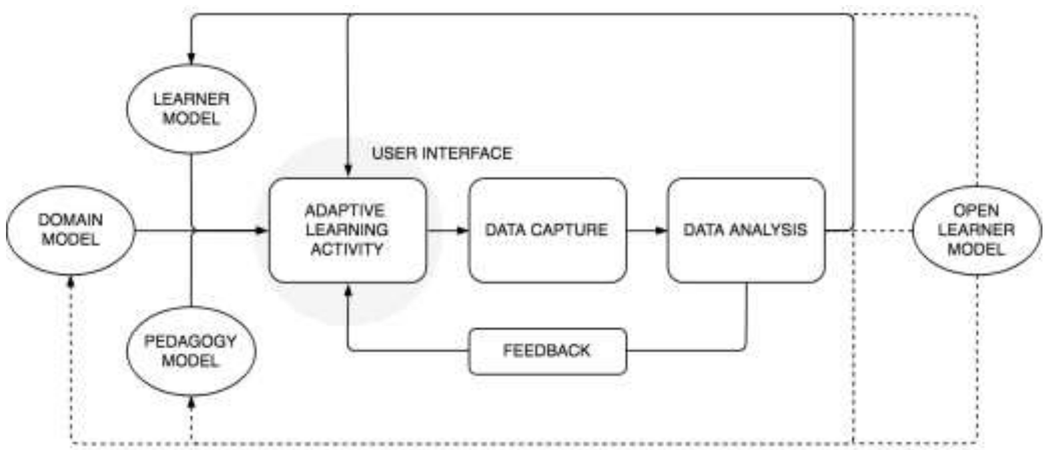


Figure 6. Flowchart representing a typical ITS architecture, including the pedagogy, domain, learner and open learner models.

In this exemplar architecture, the ITS algorithm draws on the domain, pedagogy and learner models to determine what adaptive learning activity should be presented to the individual student and how it should be adapted to that student's needs and capabilities. For example, in a mathematics ITS, the DOMAIN MODEL might contain knowledge about quadratic equations, the PEDAGOGY MODEL might contain knowledge of an effective way to teach quadratic equations, and the LEARNER MODEL might contain knowledge about the student's experience learning about quadratic equations in this ITS (for example, a misconception that they exhibited, or the fact that this topic caused them some anxiety). The learner model will also contain knowledge of all students who have ever used this ITS to learn about quadratic equations.

Drawing all of this together, the ITS algorithm will determine what ADAPTIVE LEARNING ACTIVITY to present to the student in the USER INTERFACE – in other words, which specific aspect of quadratic equations to deliver (perhaps factorising or completing the square) and what approach to use to best help the student learn about those aspects of quadratic equations (perhaps some instructional text, an image or video, or an interleaved practice activity), all of which is also dependent on the learner model (knowledge of the individual's and all students' experience of learning quadratic equations in this ITS).

While the student engages with the adaptive learning activity selected by the system, DATA CAPTURE involves the system capturing thousands of data points representing each individual interaction (what is clicked on the screen and what is typed, and possibly even how rapidly they move the mouse around the screen), together with the student's achievements (which tasks they have answered correctly or partially) and any misconceptions that they have demonstrated. Some advanced ITS also capture other data such as the student's speech, physiological responses, and an inference of their affective (emotional) state.

The next step involves DATA ANALYSIS, in which all of the captured data is automatically examined, possibly using machine learning (or a Bayesian network, an AI technique that is introduced in the [APPENDIX]), both to provide the student with individualised formative FEEDBACK (to support their learning according to their individual needs) and to update the learner model (to inform the system's decision of which adaptive learning activity to deliver next, and to contribute to the model of all students). The analysis might also, in some circumstances, update the pedagogy model (identifying which approaches have been shown to support student learning most and least effectively, in particular circumstances) and the domain model (perhaps with previously unknown misconceptions that have become apparent from the student interactions).

Over time, this ITS cycle – (a) drawing on the domain, pedagogy and learner models, (b) delivering adaptive learning activities, (c) data capture, (d) data analysis, and (e) updating the models – means that each individual student will experience their own unique personalised learning pathway through the available learning activities. If their interactions suggest that they find factorising particularly challenging, perhaps they will spend more time engaging with multiple relevant activities; whereas if their interactions suggest otherwise, perhaps they made few errors along the way, they will work through fewer activities for this topic and will more quickly move onto another topic deemed to be more appropriate for their particular needs.

Finally, as also shown in Figure 6, a few ITS also feature a fourth model, known as an *open learner model*[223]. Open learner models aim to make visible or explicit, for both the students and their teachers to inspect, both the teaching and learning that has taken place and the decisions that have been taken by the system. The *open learner model* enables learners to monitor their achievements and personal challenges, supporting their metacognition, and enables teachers to better understand each individual learner's learning (their approach, any misconceptions and their learning trajectories) in the context of the whole class, as well as potentially informing the teacher's professional development.

4.1.5 *Evaluating ITS*

Over the years there have been countless examples of ITS, many of which have been evaluated in schools or universities. Usually these evaluations have focused on learning gains, comparing one or other ITS with traditional teaching methods, such as whole-class or one-to-one teaching by a human teacher, or with CAI systems. In fact, as detailed by du Boulay and colleagues, there have also now been several meta-reviews[224] (i.e. review papers that aim to draw some general conclusions by combining and analysing the trends in several individual evaluations). For example, one meta-analysis notes: "Developers of ITSs long ago set out to improve on the success of CAI tutoring and to match the success of human tutoring. Our results suggest that ITS developers have already met both of these goals."[225]However, pooling the outcomes of the several meta-studies suggests that ITSs have not yet quite achieved parity with one-to-one teaching - when combined, the meta-reviews show an average small *negative*

---

[223] Vania Dimitrova, Gord Mccalla, and Susan Bull, 'Preface:" Open Learner Models: Future Research Directions" Special Issue of the IJAIED (Part 2).', *International Journal of Artificial Intelligence in Education*, 2007, http://psycnet.apa.org/psycinfo/2007-13116-001.

[224] Kurt VanLehn, 'The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems', *Educational Psychologist* 46, no. 4 (2011): 197–221, https://doi.org/10.1080/00461520.2011.611369; Wenting Ma et al., 'Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis.', *Journal of Educational Psychology* 106, no. 4 (2014): 901; John C. Nesbit et al., 'How Effective Are Intelligent Tutoring Systems in Computer Science Education?', in *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference On* (IEEE, 2014), 99–103, http://ieeexplore.ieee.org/abstract/document/6901409/; James A. Kulik and J. D. Fletcher, 'Effectiveness of Intelligent Tutoring Systems A Meta-Analytic Review', *Review of Educational Research*, 17 April 2015, 0034654315581420, https://doi.org/10.3102/0034654315581420; Saiying Steenbergen-Hu and Harris Cooper, *A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on K–12 Students' Mathematical Learning.* (American Psychological Association, 2013), http://psycnet.apa.org/journals/edu/105/4/970/; Saiying Steenbergen-Hu and Harris Cooper, *A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on College Students' Academic Learning.* (American Psychological Association, 2014), http://psycnet.apa.org/journals/edu/106/2/331/.

[225] Kulik and Fletcher, 'Effectiveness of Intelligent Tutoring Systems A Meta-Analytic Review', 67.

effect size of -0.19[226]. On the other hand, for ITS that are compared with whole-class teaching, the outcomes of the meta-reviews have been very positive. They show a weighted average effect size of 0.47[227] - in educational intervention research, effect sizes above 0.4 are thought to be "worth having"[228].

As we mentioned at the start of this section, ITS tend to focus on well-defined domains such as mathematics or physics. However, it is also worth noting that over recent years ITS for ill-defined problems (such as legal argumentation, intercultural skills acquisition, and dispute resolution) have also been explored[229]. One reason for the relatively low-levels of interest in ITS for ill-defined domains probably stems from the fact that ill-defined problems often require students to apply cognitively complex skills, while the contexts can be uncertain and dynamic, all of which can be challenge to model in traditional ITS. The relative lack of structure also makes it difficult to scaffold effective learning pathways without artificial constraint, to provide appropriate feedback, and to evaluate what learning is actually happening. ITS in ill-defined domains can also require additional pedagogical approaches, such as non-didactic Socratic dialogue, collaborative activities, or exploratory learning (which we look at in more detail later).

As we have seen, what AIED looks like and can do is still emerging. Accordingly, rather than trying to summarise all of AIED (a task that becomes increasingly unrealisable with every passing day), in this [chapter] we are going to introduce a wide range of AIED examples. Our list is far from definitive, but it will indicate broad areas of AIED research, and it will highlight the many possibilities and challenges introduced by the application of AI technologies in classrooms. In this section, we are discussing ITS – so we begin with some prominent current ITS examples, most of which focus on structured domains (such as mathematics): Carnegie Learning's *Mathia*, Worcester Polytechnic Institute's *Assistments* and Knewton's *alta*.

---

[226] Although one meta-review did find that ITS were "just as effective as adult, one-to-one tutoring", VanLehn, 'The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems', 214.

[227] The effect size measures how far the mean of the experimental group is from the mean of the control group measured in terms of the standard deviation of the control group scores.

[228] John Hattie, *Visible Learning*, 1st ed. (Routledge, 2008).

[229] Collin Lynch et al., 'Defining Ill-Defined Domains; a Literature Survey', in *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems*, 2006, 1–10, http://people.cs.pitt.edu/~collinl/Papers/Ill-DefinedProceedings.pdf#page=7; Beverly Woolf, 'Social and Caring Tutors ITS 2010 Keynote Address', 2010, 12–14; H. Chad Lane et al., 'Intelligent Tutoring for Interpersonal and Intercultural Skills', 2007.

4.1.6 Mathia

Building on research at Carnegie Mellon University, *Mathia*[230] (previously known as *Cognitive Tutor*) delivers AI-driven personalised mathematics instruction for K12 students. As the students work through carefully structured mathematics tasks, the system acts as a personal coach, monitoring their progress (their successes and misconceptions) and directing them along individualised learning pathways. It also provides automatic feedback that aims to explain not just why the individual student got something wrong but also how they might get it right. Interestingly, Carnegie Learning argues that *Mathia* is most effective when it is used as part of a blended learning approach (i.e. they acknowledge that, on its own, it is insufficient), which includes the use of both print and digital resources, and involves students learning collaboratively in groups as well as individually.

4.1.7 Assistments

Our second example of a current instructional ITS is *Assistments*[231], developed at Worcester Polytechnic Institute, which overall uses a similar approach to *Mathia*. However, *Assistments* also aims to address a key issue for ITS, that they by definition lead to students progressing at different rates, meaning that in any one classroom the students can be at increasingly diverging levels of attainment (potentially making the classroom teacher's job more, rather than less, challenging). Accordingly, Assistments is designed to help students catch up in the evenings, working independently at home, so that in the classroom everyone's progress remains roughly aligned. Both *Mathia*[232] and *Assistments*[233] have strong, although not definitive[234], evidence for their effectiveness.

---

[230] https://www.carnegielearning.com/products/software-platform/mathia-learning-software

[231] https://www.assistments.org

[232] John F. Pane et al., 'Continued Progress. Promising Evidence on Personalized Learning.' (Santa Monica, CA: RAND Corporation, 2015), https://www.rand.org/content/dam/rand/pubs/research_reports/RR1300/RR1365/RAND_RR1365.pdf.

[233] J. Roschelle et al., 'How Big Is That? Reporting the Effect Size and Cost of ASSISTments in the Maine Homework Efficacy Study' (Menlo Park, CA: SRI International, 2017).

[234] Holmes et al., 'Technology-Enhanced Personalised Learning. Untangling the Evidence.'

4.1.8 alta

Our third example ITS is Knewton's *alta*[235], which is doubly unusual: it is designed for Higher Education students and it focuses on a range of subjects, including mathematics, economics, chemistry and statistics. Nonetheless, like most ITS, *alta* aims to function like a 1:1 tutor, with personalised step-by-step instruction, assessment, feedback and just-in-time remediation while a student engages with an assignment. The *alta* approach clearly maps onto the typical ITS architecture outlined earlier. For each subject, it has a *domain model*, which uses open educational resources[236] (OER) and includes tutor-selectable learning objectives, together with a semantic network (or knowledge graph[237]) of relationships between the content and objectives. The domain models also include databases of relevant questions, together with data about the difficulty of those questions (based on how previous students have performed when responding to them). Alta's *pedagogy model* is based on Item Response Theory[238] (i.e., it works at the granularity of individual questions, taking into account both the question's difficulty and their representativeness of the underlying concepts). It also adopts a mastery level approach (i.e., students do not move onto new learning objectives until they have achieved mastery of earlier learning objectives). In particular, the model assumes that if a student masters one of two learning objectives that are (according to the domain model's knowledge graph) related, there is a good chance that they have also mastered the other one. Meanwhile, *alta*'s *learner model* represents a student's level of mastery in terms of the learning objectives at any given point in time. This is based on the observed history of the individual student's interactions, and of all student interactions, including which questions the students have answered correctly and incorrectly, giving more weight to an individual student's most recent responses.

4.1.9 Some final examples

In addition to these three main example of current ITSs, there are many others. Meanwhile, as we have repeatedly commented, new ones seem to appear all the time, such that any list will

[235] https://www.knewtonalta.com

[236] Jan Hylén, 'Open Educational Resources: Opportunities and Challenges', *Proceedings of Open Education*, 2006, 49–63.

[237] Heiko Paulheim, 'Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods', ed. Philipp Cimiano, *Semantic Web* 8, no. 3 (6 December 2016): 489–508, https://doi.org/10.3233/SW-160218.

[238] Susan E. Embretson and Steven P. Reise, *Item Response Theory* (Psychology Press, 2013).

always be incomplete. With this in mind, we will round out our discussion by briefly mentioning just four more, perhaps less well-known, but widely available ITS (chosen because they adopt slightly different approaches): *Area9 Learning, Dreambox*, *Toppr* and *Yixue* (we could also have chosen *ALEKS*[239], *Bjyu*[240], *Century*[241], *CogBooks*[242], *iReady*[243], *RealizeIt*[244], *Smart Sparrow*[245], *Summit Learning*[246] or... the list goes on).

Area9 Learning[247] stands out. They develop an organisation's existing learning materials into adaptive content that is delivered on their platform. Like all ITS, their approach aims to match the learning content and pathway to the needs and skill level of each individual learner, but the platform also uses an approach they call "continuous self-assessment". This involves learners rating how confident they are in their response to each question and task, which is then used to further adapt the learner's experience (even if they give a correct answer to a question, if they are not confident in that answer, they will be given additional related learning support). The diagram below shows how adaptive learning *simultaneously* reduces the time spent on learning for the bulk of the students, while allowing the slower ones to achieve mastery at their own pace.

[239] https://www.aleks.com

[240] https://byjus.com

[241] https://www.century.tech

[242] https://www.cogbooks.com

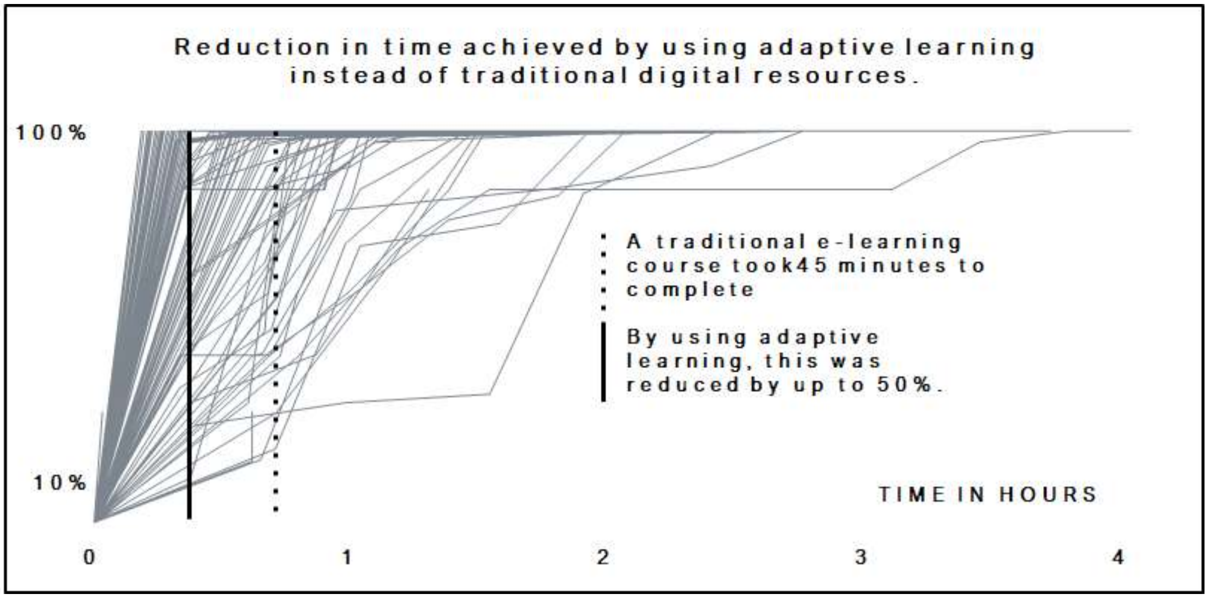
[243] https://www.curriculumassociates.com/Products/i-Ready

[244] http://realizeitlearning.com

[245] https://www.smartsparrow.com

[246] https://www.summitlearning.org

[247] https://area9learning.com



Source: Area9 Learning, private communication

Our second example, *Dreambox*[248], aims to provide students with personalised learning pathways, in K-8 mathematics: *"the right next lesson, at the right level of difficulty, at the right time"*. Again adopting a typical ITS approach, their AI-driven technology collects more than 48,000 data points every hour a student engages with the system, which it uses to evaluate the strategies the student employed to solve the problems, to adjust the lesson's level of difficulty, to sequence the lessons and to provide hints. Of the commercial ITS, *Dreambox* is fairly unusual in encouraging independent evaluations, with a recent study conducted by Harvard University finding that *"the evidence for the causal impact of DreamBox on student achievement is encouraging but mixed"*[249].

Meanwhile, Toppr[250], is an India-based company that offers personalised learning ITS mobile apps, across a wide range of school ages and subjects (from history to accounting). It uses machine learning to map out a student's strengths and weaknesses, based on their previous behaviour, in order to personalise questions and adjust the speed of presentation to make the experience optimal for each individual. This ITS prediction system is complemented by a novel AI-driven technology that is, they say, designed "to solve doubts". Students can upload an

---

[248] http://www.dreambox.com

[249] Jon Fullerton, 'Dreambox Learning Achievement Growth in the Howard County Public School System and Rocketship Education' (Boston, MA: Center for Educational Policy Research, Harvard University, 2016), https://cepr.harvard.edu/dreambox-learning-achievement-growth.

[250] https://www.toppr.com

image of a topic about which they are unsure, which a bot checks against a growing database of other uploaded “doubts” and solutions.

Our final example ITS is *Yixue*[251], which styles itself as the first intelligent adaptive education system in China. Again as a typical ITS, *Yixue* aims to simulate a teacher, providing students with a personalised learning plan and one-to-one tutoring. Drawing on standard textbooks, *Yixue* have broken each of its various subjects into around 10,000 separate “knowledge points”, which are used to benchmark an individual student’s understanding and capabilities, so that the system can predict which materials and pathway will be most effective.

4.2 *Dialogue-based Tutoring Systems*

We began our discussion of ITS with Jaime Carbonell’s SCHOLAR[252]. However, SCHOLAR is in at least one sense unlike most of the ITS that we have so far explored. Rather than presenting an individualised sequence of instructional material or learning activities (as is typical of ITS), SCHOLAR engaged students in conversations about the topic to be learned. This aspect gave rise to a version of ITS known as dialogue-based tutoring systems (DBTS). However, as with ITS, what constitutes DBTS is fuzzy-edged. Instead, we will again introduce some prominent examples : *CIRCSIM*, *AutoTutor* and *Watson Tutor*.

4.2.1 *CIRCSIM*

One of the earliest DBTSs was CIRCSIM[253], which was developed in the 1980s at Illinois Institute of Technology in partnership with Rush Medical College. It was designed for first-year medical students, to help them learn about the baroreceptor reflex control of blood pressure. CIRCSIM used one-to-one tutorial dialogues, involving some limited natural language processing and natural language generation, on the assumption that real understanding of something involves being able to talk articulately about it. It also used a rule-based expert systems approach, implementing conditional rules such as:

> *If the student answer is correct, then proceed.*
>
> *If the student answer is partially correct, then give acknowledgement and proceed.*

[251] http://www.classba.cn

[252] Carbonell, ‘AI in CAI’.

[253] Martha Evens and Joel Michael, *One-on-One Tutoring by Humans and Computers* (Psychology Press, 2006).

If the student answer is a 'near miss', then introduce a nested method.

If the student answer is "don't know", then give answer and proceed.[254]

Interestingly, CIRCSIM was not designed to introduce students to the topic. Instead, students were expected to have already acquired the facts and concepts from their readings and lectures. Their dialogue with the system helped the students explore in depth, better understand and consolidate what they had already learned. With this aim, students were asked to solve problems while engaging in an iterative typed dialogue. They would begin with a mandatory guided virtual experiment. The programme then directed the students step-by-step through a sequence of eight procedures, guiding them to predict outcomes based on supplied data, and to develop a simplified model for the homeostatic baroreceptor reflex system. Throughout, the emphasis was on the need for developing a chain of causal reasoning in solving this and similar problems.

4.2.2 AutoTutor

Our second example, AutoTutor[255], which has been extensively researched for over twenty years, is probably the most influential DBTS. Developed at the University of Memphis, it simulates a tutorial dialogue between human tutors and students as they work step-by-step through online tasks (most often in computer science topics, but also in physics, biology and critical thinking). The aim is to encourage students to develop detailed responses and an in-depth understanding, rather than the short responses and shallow knowledge that can be the outcome of some step-by-step instructional ITS.

AutoTutor uses a statistical technique known as Latent Semantic Analysis (LSA) to compare student written speech with a multi-dimensional matrix of concepts drawn from a large corpus of relevant textbooks[256]. This matrix of concepts and a curriculum script (comprising example

---

[254] Evens and Michael, 45.

[255] Arthur C. Graesser et al., 'Intelligent Tutoring Systems with Conversational Dialogue', *AI Magazine* 22, no. 4 (2001): 39.

[256] Arthur C. Graesser et al., 'Using Latent Semantic Analysis to Evaluate the Contributions of Students in AutoTutor', *Interactive Learning Environments* 8, no. 2 (1 August 2000): 129–47, https://doi.org/10.1076/1049-4820(200008)8:2;1-B;FT129. LSA, developed by Thomas Landauer (University of Colorado) originally for indexing documents for information retrieval, is "both a computational model of human knowledge representation and a method for extracting semantic similarity of words and passages from text". Peter W. Foltz, Darrell Laham, and Thomas K. Landauer, 'The Intelligent Essay Assessor: Applications to Educational Technology', *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning* 1, no. 2 (1999), http://www.imej.wfu.edu/articles/1999/2/04/printver.asp.

questions, problems, diagrams, declarative knowledge, and good and bad responses), effectively constitutes AutoTutor's ITS domain model.

Meanwhile, its version of a pedagogy model comprises Socratic tutoring principles (probing with questions rather than providing instruction) and classroom-based tutorial practices (based on analyses of dialogue from more than 100 hours of face-to-face human tutorials). Its adaptive learning activities involve engaging the student in a tutorial dialogue, a developing conversation in which the student is guided towards discovering for themselves a correct solution for the current task.

An AutoTutor tutorial dialogue typically comprises five steps[257]: (1) AutoTutor (sometimes represented by an online animated character) poses a question or problem, (2) the student attempts to answer, typing their response into the system (or, with some versions, speaking aloud), (3) the tutor determines whether the student understands the target concept, by assessing how closely their contribution matches the concept as expressed in the textbook corpus (the LSA approach means that the student does not have to match specific wording but only the underlying concept), (4) if the tutor decides the student has not sufficiently understood the concept, the tutor provides feedback (in the form of corrective statements, questions, prompts, hints or summaries), perhaps to correct a misunderstanding or to encourage the student to extend their answer, and (5) steps two to four are repeated, with the aim of helping the student improve their answer. Finally, AutoTutor's version of a student model comprises the student's contributions to the dialogue, together with the *goodness of fit* (the closeness of the student's answer to the corpus of texts) of each of those student inputs[258].

An AutoTutor dialogue[259] might look something like this:

> AutoTutor <Main Question>: *Suppose a boy is in a free-falling elevator and he holds his keys motionless in front of his him and he lets go. What will happen to the keys? Explain why.*

---

[257] Arthur C. Graesser, 'Learning, Thinking, and Emoting with Discourse Technologies', *American Psychologist* 66, no. 8 (November 2011): 746–57, https://doi.org/10.1037/a0024974.

[258] Peter Wiemer-Hastings et al., 'The Foundations and Architecture of AutoTutor', in *Intelligent Tutoring Systems* (Springer, 1998), 334–343, http://link.springer.com/chapter/10.1007/3-540-68716-5_39.

[259] Benjamin D. Nye, Arthur C. Graesser, and Xiangen Hu, 'AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring', *International Journal of Artificial Intelligence in Education* 24, no. 4 (December 2014): 434, https://doi.org/10.1007/s40593-014-0029-5.

Student: *The keys would not move but hover at the start point.*

AutoTutor <Pump for information>: *I bet you can explain this a little bit more.*

Student: *The velocity of the keys is the same due to gravity.*

AutoTutor <Hint>: *What about the acceleration of the objects involved?*

Student: *What is acceleration again?*

AutoTutor <Answer>: *Let me try to answer your question. Acceleration is the rate that velocity changes over time, like getting faster or getting slower. Acceleration is a vector quantity, so it needs to have a direction.*

AutoTutor <Hint>: *Let's try this again. What about the acceleration of the objects involved?*

Student: *The acceleration of the objects in the elevator are the same as the acceleration of the elevator which is the earth's gravitational acceleration.*

AutoTutor <Short Feedback>: *Good job!*

AutoTutor <Prompt>: *Try this. What direction are the objects going? The objects are falling* <the animated character gives a downwards hand gesture to cue the student to express something>?

Student: *Downward.*

AutoTutor <Short Feedback>: *Right*.

AutoTutor <Prompt-completion>: *The objects are falling downward.*

Over its two decades, AutoTutor has been evaluated in multiple studies[260]. In summary, these have shown that it can achieve average learning gain effect sizes of approximately 0.8, with higher gains for deep learning (understanding concepts) than for shallow learning (knowing facts), suggesting that it is more effective than many non-expert human tutors and might in some situations be as effective (at least in terms of learning gains) as a typical expert tutor[261].

---

[260] Sidney D'Mello and Art Graesser, 'AutoTutor and Affective AutoTutor: Learning by Talking with Cognitively and Emotionally Intelligent Computers That Talk Back', *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, no. 4 (2012): 23.
[261] VanLehn, 'The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems'; Nye, Graesser, and Hu, 'AutoTutor and Family'.

4.2.3 Watson Tutor

A more recent DBTS is the Watson Tutor[262] being developed collaboratively by the global corporations IBM and Pearson, which at the time of writing is being evaluated before being released as a commercial product (integrated into a Pearson higher education courseware product called REVEL). Watson Tutor is a dialogue-based tutorial system using natural language conversation that guides students through a review session – in other words, like AutoTutor and CIRCSIM, it does not set out to introduce new knowledge but to enable a deeper understanding of existing knowledge[263]. As the students engage with the Tutor, it provides supportive content (such as text, images and videos), tracks their progress and adapts the conversation based upon a classification of the student's responses and an assessment of their subject mastery.

The Watson Tutor draws heavily on the approach developed by the AutoTutor researchers, although its domain model, the formalisation of the knowledge and skills to be learned, is derived from a single textbook. It comprises a set of learning objectives and enabling objectives (sub-learning objectives that support a main learning objective), a knowledge graph of the relationships between the learning objectives, main questions and main answers, assertions (knowledge components of the main answers), and hint questions to elicit assertions (typically 600 main and hint questions are derived from the single textbook) and fill in the blank questions (based on the student assertions). All of this is derived from the chosen textbook by means of an automated analysis undertaken by the IBM Watson toolset (the AI as a service mentioned earlier, which includes natural language understanding and classification tools). This approach to the domain model means that it can, at least in theory, straightforwardly be adapted to any academic domain and its textbook (in its initial implementation it covered developmental psychology, physiological psychology, abnormal psychology, sociology, communication, and government).

The pedagogy model again adopts a Socratic approach, implemented as a dialogue manager that controls when and which responses are delivered to a student. Starting from a learning objective, Watson Tutor begins by asking the student a main question, with the aim of eliciting

[262] https://www.ibm.com/watson/education

[263] Matthew Ventura et al., 'Preliminary Evaluations of a Dialogue-Based Digital Tutor', ed. Rosé Carolyn (Artificial Intelligence in Education, London, 2018), 480–83.

from the student the main answer. The student's responses are classified by the IBM Watson system, enabling the system to respond with appropriate feedback or a hint question to help the student work towards an ideal answer. For main questions, the classifier matches the student's response to all the knowledge components of the main answer, using a supervised machine learning approach that requires only 35 scored student answers per question. If the student struggles on a main learning objective, the Tutor will move to an enabling objective, repeating the dialogue approach.

The Watson Tutor's learner model includes an estimation of each student's proficiency over the course of the conversation, in terms of both the learning and enabling objectives. This also functions as an open learner model, in that the student's estimated proficiencies and developing performance can be inspected, at both a high-level and a detailed granularity, by the individual students and their teachers.

A Watson Tutor dialogue might look something like this[264]:

> Watson Tutor <Learning objective main question>: *Let's review this topic… Try answering this question: What are some variations in the definition of family?*
>
> Student: *A family can be defined as people who live together*.
>
> Watson Tutor <Feedback on student assertion>: *Excellent. Kinship systems consist of the social links and boundaries, defined by biology and social custom, that establish who is related to whom. Let's take it a bit further.*
>
> Watson Tutor <Learning objective hint question>: *How do social scientists define the family?*
>
> Student: *Social scientists define the family as a group of related people that reside together.*
>
> Watson Tutor <Feedback on student assertion>: *That looks partly correct. There's a little more to the answer. Let's work this out together.*

[264] Based closely on Ventura et al., 482.

Watson Tutor <Fill in the blank hint>: *Can you fill in the missing words? Social scientists typically define family as a group of people who live together in a household and share ________ and/or _____ ties.*

4.3 *Exploratory learning environments*

An alternative to ITS step-by-step instruction and DBTS step-by-step dialogue is provided by a third type of AIED, known as Exploratory Learning Environments (ELEs). ELEs adopt a constructivist approach. Rather than following a dynamically generated step-by-step sequence, students are encouraged to actively construct their own knowledge by exploring and manipulating elements of the learning environment.

In fact, exploratory and discovery learning have been around for a long time[265], but they remain controversial[266]. Critics argue that, because there is no explicit instruction and students are expected to discover domain principles for themselves, it causes cognitive overload and leads to poor learning outcomes. This is where AI comes in, with many recent ELEs including AI-driven automatic guidance and feedback, addressing misconceptions and proposing alternative approaches, to support the student while they explore.

As we have seen, delivering effective AI-driven support requires a learner model. However, building learner models for unstructured environments like ELEs can be challenging: *"The unconstrained nature of the interaction and the lack of easily definable correct behaviors make it difficult to know a priori what behaviors are conducive for learning."*[267] Nonetheless, student models are usually an important component of ELEs.

Again, we will briefly explore[268] four examples, each of which does include a student model, that use different AI-driven approaches to provide the necessary support: *Fractions Lab* (which delivers automated feedback according to a student's affective state), *Betty's Brain* (which

---

[265] e.g., J. S. Bruner, 'The Act of Discovery.', *Harvard Educational Review* 31 (1961): 21–32.

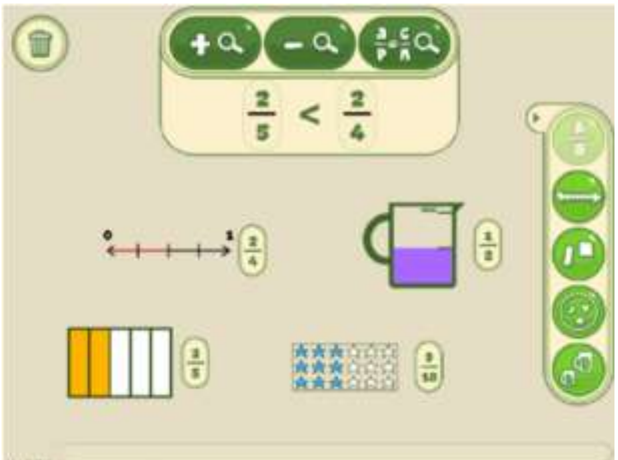
[266] P. Kirschner, J. Sweller, and R. E Clark, 'Why Minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching', *Educational Psychologist* 41, no. 2 (2006): 75–86.

[267] Lauren Fratamico et al., 'Applying a Framework for Student Modeling in Exploratory Learning Environments: Comparing Data Representation Granularity to Handle Environment Complexity', *International Journal of Artificial Intelligence in Education* 27, no. 2 (1 June 2017): 321, https://doi.org/10.1007/s40593-016-0131-y.

[268] Building on Benedict du Boulay et al., 'What Does the Research Say about How Artificial Intelligence and Big Data Can Close the Achievement Gap?', in *Enhancing Learning and Teaching with Technology*, ed. Rose Luckin (London: Institute of Education Press, 2018), 316–27.

involves a teachable agent), *Crystal Island* (which uses a games-based approach), and *ECHOES* (which is designed to support children on the autism spectrum).

4.3.1 *Fractions Lab*



Fractions Lab, which was developed by an EU-funded research project[269], is designed to help students develop conceptual knowledge, the underlying principles, of fractions. In this ELE, students can choose and manipulate fraction representations (for example, having chosen a rectangle, jug or number line to represent a particular fraction, they can create the fraction by changing the numerator and denominator), the aim being to solve a given fractions problem (such as "Use the Fractions Lab tools to add together $\frac{2}{3}$ and $\frac{2}{6}$.”). To avoid cognitive overload while they address the given task, Fractions Lab uses AI-techniques to provide the students with adaptive support – that is to say, feedback or guidance specific to where they are in their attempted solution (such as "To add the two fractions together, you first need to make them equivalent. How do you need to adjust the denominator?"). However, in addition to providing this context-specific guidance, the feedback also aims to enhance the student's affective states – i.e., to move students from nominally negative affective states (such as frustration or boredom) into nominally positive affective states (which are usually assumed to be more conducive for learning).

This is achieved by means of Bayesian networks (an AI technique that is introduced in the [APPENDIX]) trained with data from classroom studies, one of which determines the most appropriate type of formative feedback to be given to the student. For example, if the student is confused, the Bayesian network determines that an affect boost (such as "Well done. You're working really hard!") or specific instructive feedback (such as "Use the comparison box to compare your fractions.") is most effective.

[269] http://www.italk2learn.eu

Other feedback provided by the system includes Socratic feedback (such as "What do you need to do now, to complete the fraction?"), reflective prompts (such as "What do you notice about the two fractions?"), affirmation prompts (such as "The way that you worked that out was excellent."), and sequence prompts (such as "Are you sure that you have answered the task fully? Please read the task again.").

Fractions Lab's pedagogy and domain models comprise both the overall constructivist ELE approach, and the information used to determine the content of the formative feedback. Meanwhile, the learner model includes data about the student's inferred affective state, their progress with the current task, their interactions with the learning environment (whether a representation has been created, selected or manipulated), the type of feedback they have received, and the specific message, and whether or not the student follows the feedback. While the student interacts with the fractions representations to answer the tasks, the learner model is constantly being updated with information, both about those interactions and with the feedback that has been provided to the student.

As part of a larger project[270], Fractions Lab was evaluated in schools in Germany and the UK, comparing the effectiveness of Fractions Lab in combination with an ITS. Although the tools were only used for a short time, the outcomes showed that the combination of ELE and ITS achieved a learning gains effect size of 0.7 (compared with the ITS alone), suggesting that AI-supported ELEs can offer a useful approach to learning.

4.3.2 *Betty's Brain*

An iconic ELE is Betty's Brain[271], which involves an AI-driven teachable agent. It was designed to facilitate the learning of scientific conceptual understanding, using river ecosystems as a use-case exemplar. What distinguishes Betty's Brain is that, through their engagement with the system, students are encouraged to teach a fellow student, in fact a virtual agent, called Betty. This approach (which is the foundation of the system's pedagogy model) has been adopted because learning by teaching has been shown to be effective – it is known to help students

[270] Nikol Rummel et al., 'Combining Exploratory Learning with Structured Practice to Foster Conceptual and Procedural Fractions Knowledge', in *ICLS* (ICLS, Singapore, 2016).
[271] Krittaya Leelawong and Gautam Biswas, 'Designing Learning by Teaching Agents: The Betty's Brain System', *International Journal of Artificial Intelligence in Education* 18, no. 3 (2008): 181–208.

structure, reflect on and develop a more in-depth understanding of whatever is being learned[272].

Within an overarching narrative (helping Betty to join a science club), students are supported to teach Betty, then to query Betty to see how much she has understood, and finally to quiz Betty to see how well she does on questions generated automatically by the system, many of which the student may not have considered.

The mechanism used for teaching Betty centres on a Concept Map Editor, which represents what the student has taught Betty. Drawing on a range of provided reading materials and using the available editing tools, the student builds a conceptual map of the river ecosystem (the relationships between a river's plants, animals, microorganisms, chemical components, and physical characteristics), attaching nodes (representing particular knowledge components) via edges (representing causal and other links between the various components). In effect, the students build their own semantic network, which forms the core of the system's learner model (i.e. it represents the student's current knowledge and understanding). The learner model also includes a record of the student's interactions with the system. Interestingly, the Concept Map, because it is visual and open to inspection by the student and the teacher, also functions as an open learner model.

Once Betty has been taught, the student can ask her a question (such as, "if macroinvertebrates increase what happens to bacteria?"). In response, Betty reasons using the concept map to generate an answer (such as, "An increase in macroinvertebrates causes no change in bacteria."). The student can also ask Betty for an explanation, which Betty gives while highlighting the causal paths in the Concept Map.

The system also uses the Concept Map, together with the system's domain model, to generate the quiz questions which are administered by Mr Davis, a virtual teacher. Betty's responses to the quiz questions draw directly from the concept map, while Mr Davis's feedback also uses the domain model to suggest how the student might edit the concept map in order to help Betty achieve a higher quiz score – in other words, to correct any mistakes made by Betty and to identify any misconceptions. Mr Davis also makes suggestions at the meta-cognitive level,

[272] Gautam Biswas et al., 'Learning by Teaching: A New Agent Paradigm for Educational Software', *Applied Artificial Intelligence* 19, no. 3–4 (9 March 2005): 363–92, https://doi.org/10.1080/08839510590910200.

for example about making better use of the reading materials, in an effort to help the learner develop good meta-learning strategies (study skills).

Betty's Brain has been evaluated in multiple studies[273], although in its original configuration student outcomes were often split, almost half and half, between those who made very good progress with the system and those who struggled. The researchers went on to develop a new version, which they used to investigate different learning behaviour profiles[274].

4.3.3 Crystal Island

Crystal Island[275] emerged from research at North Carolina State University, takes an immersive first-person computer game approach, with students playing the role of a detective investigating a mysterious disease on a remote island. This games-based learning approach[276] functions as Crystal Island's pedagogy model. In solving the mystery, students use and develop their literacy skills while gaining experience of professional scientific inquiry methods (including evidence gathering, hypotheses testing, and data analysis), all of which together constitute the domain model. Meanwhile, the students' developing knowledge, their affective states and their skills, are automatically modelled (in the ELE's learner model) and they receive automated supportive feedback. In addition, throughout the game-play, they engage with AI-driven autonomous non-player characters (companion agents), which build on AI techniques developed over many years in mainstream computer games[277].

4.3.4 ECHOES

Our fourth example ELE, *ECHOES*[278], again involved a games-based approach, but this time to support young children who are on the autism spectrum. *ECHOES* was a virtual environment, a 'magic' garden, in which the child interacted with an intelligent child agent called Andy. The

---

[273] Gautam Biswas, James R. Segedy, and Kritya Bunchongchit, 'From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain', *International Journal of Artificial Intelligence in Education* 26, no. 1 (2016): 350–364.

[274] Hogyeong Jeong et al., 'Using Hidden Markov Models to Characterize Student Behaviors in Learning-by-Teaching Environments', in *Intelligent Tutoring Systems*, Lecture Notes in Computer Science (International Conference on Intelligent Tutoring Systems, Springer, Berlin, Heidelberg, 2008), 614–25, https://doi.org/10.1007/978-3-540-69132-7_64.

[275] http://projects.intellimedia.ncsu.edu/crystalisland

[276] Wayne Holmes, 'Digital Games-Based Learning. Time to Adoption: Two to Three Years?', in *Education and New Technologies: Perils and Promises for Learners*, ed. Kieron Sheehy and Andrew J. Holliman, 2017.

[277] Georgios N. Yannakakis and Julian Togelius, *Artificial Intelligence and Games* (Cham: Springer International Publishing, 2018), https://doi.org/10.1007/978-3-319-63519-4.

[278] Bernardini, Sara, Kaśka Porayska-Pomsta, and Tim J. Smith. 2014. 'ECHOES: An Intelligent Serious Game for Fostering Social Communication in Children with Autism'. *Information Sciences* 264 (April): 41–60. https://doi.org/10.1016/j.ins.2013.10.027.

child’s teacher (not AI) selected one of twelve learning activities, led by Andy, each of which was designed to enhance the child’s capacities for joint attention and to help them develop their social communication skills.

The magic garden was displayed on a large touchscreen monitor, allowing the child and Andy to interact with each other and with objects in the garden. Sometimes, when touched, the garden objects transformed in unusual ways. For example, tapping the petals of a flower could turn it into a floating bubble or a bouncy ball. The system also included an eye-tracking camera, allowing Andy to ‘know’ where the child was looking.

Andy was designed to function as an artificially intelligent social partner who could act both as a peer and as a tutor. Its implementation was based on a well-established AI agent architecture called FAtiMA[279], which enabled it to be autonomous, proactive, reactive, and socio-emotionally competent. In particular, Andy was designed to be a positive and supportive character. For example, he always greeted the child by name, gave positive feedback when the child participated in an interaction, and tried to re-engage the child if they appeared distracted. Andy also used facial expressions and gestures to indicate his emotional responses. For example, he smiled and gave a thumbs up when the child initiated an activity.

ECHOES also included a pedagogy model, which monitored the developing interactions between the child and Andy, to help ensure that the learning objectives were achieved, and a user model that aimed to monitor the cognitive and emotional state of the child, so that Andy could give appropriate real-time feedback.

4.3.5 *Summary*

As we have seen, because ELEs are unstructured and open-ended learning environments which students can explore as they like, there is no clear definition of ‘correct behaviours’ which makes it difficult to model the student and to provide the necessary guidance. With this in mind, over several years, Conati and colleagues have developed and researched a multi-layered student modelling framework[280], which they implemented in CCK and which potentially

[279] Dias, João, and Ana Paiva. 2005. ‘Feeling and Reasoning: A Computational Model for Emotional Characters’. In *Progress in Artificial Intelligence*, 127–40. Springer, Berlin, Heidelberg .https://doi.org/10.1007/11595014_13.
[280] Samad Kardan and Cristina Conati, ‘Providing Adaptive Support in an Interactive Simulation for Learning: An Experimental Evaluation’, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15 (New York, NY, USA: ACM, 2015), 3671–3680, https://doi.org/10.1145/2702123.2702424.

can be applied in other ELEs. This framework uses multiple logged student actions to learn which behaviours should trigger remedial guidance, and which lead to what outcomes (e.g., high- or low-achieving). It involves the use of unsupervised-learning to cluster groups of students who, based on their logged data and learning outcomes, learn 'similarly'. The logged data includes the components used (e.g., light bulb), student actions (e.g., join), and simulation outcomes (e.g., a change in light intensity). The student clusters model is then used to classify new students, and to trigger real-time adaptive support based on the logged and thus anticipated behaviours, in order to support students to achieve higher learning outcomes.

4.4 *Automatic Writing Evaluation*

The AIED applications that we have looked at so far – the step-by-step instructional and dialogue-based systems, and the exploratory learning environments – all involve students working on computers (sometimes mobile devices[281]), following individualised learning pathways while receiving immediate adaptive support. Another type of AIED, Automatic Writing Evaluation (AWE), uses natural language and semantic processing to provide automatic feedback on student writing submitted to the system[282].

Broadly speaking, there are two overlapping AWE approaches, formative (i.e., providing support to enable a student to improve their writing before submitting it for assessment)[283] and summative (i.e. automatic scoring)[284]. However, in line with the long-established automated grading of multiple choice and fill-in-the-blank tests, much of the work has been directed to scoring over feedback. It has often been driven by the desire to improve the cost, reliability and generalizability of summative assessments, whether for use in small-scale settings by teachers (for low-stakes classroom assessment) or large-scale settings by testing companies (for large-scale, statewide or national, high-stakes assessments).

> *"Efficiency is where the automated readers excel.* The e-rater *engine can grade 16,000 essays in about 20 seconds, according to ETS. An average*

---

[281] Fujitani, Satoru, and Kohei Minemura. 'An Analysis of Expectations for Artificial Intelligence-Supporting Software in Mobile Learning', 2017, 6.

[282] According to John Behrens (Pearson), automated essay grading is one area that *"machine learning is starting to make progress"*, quoted in *What Can Machine Learning Really Predict in Education?* (2018) by Sydney Johnson. https://www.edsurge.com/news/2018-09-26-what-can-machine-learning-really-predict-in-education

[283] One example being M-Write: https://lsa.umich.edu/sweetland/m-write.html

[284] One example being Gradescope: https://www.gradescope.com.

teacher might spend an entire weekend grading 150 essays, and that efficiency is what drives more education companies to create automated systems."[285]

This is why automated essay feedback and scoring is probably the best funded area of AIED research, and why it has spawned a large number of commercial systems[286]. There are so many available AWE systems[287], with different approaches and limitations, that we will again simply introduce some prominent examples.

4.4.1 *PEG*

The beginnings of AWE can be traced to the development in 1966 of Project Essay Grade (also known as PEG) by Ellis Page at Duke University. The original version of PEG used correlational statistics to compare submitted essays with a training set of up to 400 comparable essays that had already been marked by teachers, an approach that was shown in various studies to achieve predicted scores comparable to the human markers. However, the PEG system was criticised for focusing on indirect measures of writing skill (the surface features of essays such as the number of sentences, use of punctuation and grammar) rather than on the meaning of the sentences, the writing style or how the writer had developed their arguments (in other words, it was criticised for focusing on the form rather than the content of the essays). For this reason, PEG was incapable of providing meaningful formative feedback, to enable students to improve the academic (rather than the surface) quality of their essay, and instead provided only a summative mark. The effectiveness of the system was also dependent on the selection of the training essays, and the quality of the assessments made by the training set's human markers. More recently, PEG was re-engineered to include computational linguistics, machine learning,

---

[285] https://www.npr.org/sections/alltechconsidered/2012/04/24/151308789/for-automatic-essay-graders-efficiency-trumps-accuracy?t=1542533112695

[286] Semire Dikli, 'An Overview of Automated Scoring of Essays', *The Journal of Technology, Learning and Assessment* 5, no. 1 (2006), https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640. Raczynski, Kevin, and Allan Cohen. 2018. 'Appraising the Scoring Performance of Automated Essay Scoring Systems—Some Additional Considerations: Which Essays? Which Human Raters? Which Scores?' *Applied Measurement in Education* 31 (3): 233–40. https://doi.org/10.1080/08957347.2018.1464449. See also Hubert's "AI In Education — Automatic Essay Scoring", https://medium.com/hubert-ai/ai-in-education-automatic-essay-scoring-6eb38bb2e70.

[287] Stevenson, Marie, and Aek Phakiti. 'The Effects of Computer-Generated Feedback on the Quality of Writing'. *Assessing Writing* 19 (2014): 51–65.

and natural language processing techniques. It was also included in the Hewlett Foundation-sponsored Automated Student Assessment Prize competition[288].

4.4.2 *Intelligent Essay Assessor*

An alternative early AWE approach, Intelligent Essay Assessor (IEA), used Latent Semantic Analysis (LSA), the statistical technique that we introduced when discussing the dialogue-based tutoring system AutoTutor. LSA enables IEA to infer the meanings of words and sentences, by considering the context in which they occur, and to calculate the semantic relatedness of a target document to the training corpus: *"The underlying idea is that the meaning of a passage is very much dependent on its words and changing even only one word can result in meaning differences in the passage. On the other hand, two passages with different words might have a very similar meaning"*[289].

Landauer and colleagues applied this approach to also develop what they called the Intelligent Essay Assessor (IEA), which calculated similarity scores for essays when compared with a set of training texts, which included a large number of pre-scored student essays, expert model essays, and knowledge source materials (such as text books and academic papers). The system determined the essay's mark by averaging the similarity scores. However, the comparison with the key domain-representative texts also allowed the system to provide diagnostic and evaluative formative feedback across six areas: ideas and content, organisation, sentence fluency, word choice, conventions and voice. IEA was also able to detect some plagiarism (i.e., passages replicating text in the knowledge source materials) and collusion (similar passages appearing in more than one essay in a cohort), both of which (if only because of the scale involved) can be difficult for human markers to identify. IEA was also included in the Automated Student Assessment Prize competition[290]

---

[288] Shermis, Mark D. 2014. 'State-of-the-Art Automated Essay Scoring: Competition, Results, and Future Directions from a United States Demonstration'. *Assessing Writing* 20 (April): 53–76. https://doi.org/10.1016/j.asw.2013.04.001.

[289] Dikli, 'An Overview of Automated Scoring of Essays', 5.

[290] Shermis, Mark D. 2014. 'State-of-the-Art Automated Essay Scoring: Competition, Results, and Future Directions from a United States Demonstration'. *Assessing Writing* 20 (April): 53–76. https://doi.org/10.1016/j.asw.2013.04.001.

4.4.3 *WriteToLearn*

Over recent years, the IEA approach has been further developed by the international education company Pearson and incorporated in their product WriteToLearn[291]. The current system draws on a broad range of AI techniques in order to provide both in-depth formative feedback and summative scoring. Essays are assessed against one or more rubrics, using a supervised machine-learning approach involving a training set of around 300 representative essays that have been scored by humans. The rubrics involve traits such as focus, development of ideas, organization, language and style, voice, sentence correctness, and sentence fluency. The system is also able to detect a variety of errors in the submitted essay, and thus to provide a series of specific prompts in terms of narrative, exposition, description, and persuasion, all scaffolded on the student's writing performance and designed to enable the student to improve their next draft. In addition, the system also gives a score to the writing, by assessing it against a rubric designed to represent characteristics of high-quality writing: ideas, organisation, conventions, sentence fluency, word choice, and voice. A secondary component in the software can be used to assess and provide feedback based on students' summarisations of given texts, which has been shown to help students build their reading comprehension skills.

WriteToLearn has been evaluated in a number of studies. In one state-wide study[292], involving more than 20,000 students and 70,000 assignments, the students submitted an average of around four revised drafts (more than is typical in a traditional classroom setting) and improved their overall scores by almost one point out of a maximum of 6 (effect sizes are not given). The improvements were shown to be in both basic writing skills and higher-level traits like ideas and voice.

4.4.4 *e-Rater*

A third AWE approach, originally known as *e-Rater*, which was developed by the Educational Testing Service, has been widely used (for example, in GMAT tests and, in a more recent

[291] https://www.writetolearn.net

[292] Peter W. Foltz and Mark Rosenstein, 'Tracking Student Learning in a State-Wide Implementation of Automated Writing Scoring', in *NIPS Workshop on Data Driven Education*, 2013.

version, for Common Core Standards)[293]. Like the earlier systems, e-Rater analyses a large set of linguistic features (syntactic variety, topic content, and lexical and syntactic cues) that it automatically extracts from essays using Natural Language Processing techniques. Algorithms then assign values for every feature identified in an essay, which are computed using linear regression and compared to a training set of essays scored by human experts, to predict a final score. ETS claim the psychometric validity of *e-rater* scores, while accounting for cultural and second-language differences, across a range of subject areas.

4.4.5 *Revision Assistant*

Three short final AWE examples are Revision Assistant, OpenEssayist and AI grading. Perhaps best known for their antiplagiarism software, which automatically checks student writing against billions of internet documents and journal papers, the Turnitin corporation also now offers *Revision Assistant*. This system is designed to evaluate and provide formative feedback on short student essays (between 200 and 700 words) in a range of genres, which it achieves by using both supervised ML (involving training essays that have been scored by at least two human teachers) and unsupervised ML (involving a collection of thousands of unscored essays collected from students who have already used the system in classrooms).

The Turnitin analysis works by representing the submitted essays in terms of large numbers of low-level textual features (such as word n-grams and essay length), and it uses multiple methods to compute a prediction. For example, one core analytical technique is to delete a sentence from the submitted essay, to determine how that edit affects the predicted score. If the predicted score increases, the system infers that the sentence is strong for the particular trait being evaluated. This allows the system to automatically provide sentence-level formative feedback specific to the rubric in question, drawn from a pool of more than a thousand feedback comments authored by content experts, for each draft that a student submits.

Classroom observations (at the time of writing there have been no published efficacy studies) have suggested that the system generates automated feedback that is both well-received by the students and aligned with improving scores. Turnitin argue that their automated essay feedback system not only provides more frequent formative feedback to students but also

---

[293] Karen Kukich, 'Beyond Automated Essay Scoring', ed. M.A. Hearst, *IEEE Intelligent Systems and Their Applications* 15, no. 5 (September 2000): 22–27, https://doi.org/10.1109/5254.889104.

"allows the teacher to step back from the sometimes-adversarial red pen and engage with their class as guides and readers, modelling the interpretation of feedback alongside their students."

4.4.6 OpenEssayist

OpenEssayist[294], which was developed by The Open University and Oxford University in the UK, takes a somewhat alternative approach, again using Natural Language Processing but focussing more on how the feedback is presented to the students so that it is easily actionable. The system's linguistic analysis engine uses unsupervised algorithms to cluster key words, phrases and sentences from the student's essay. It then generates several external representations. For example, key words and phrases are presented in a word cloud, which can be explored and organised into groups. It also uses automatically-generated animations and interactive exercises to encourage the student to reflect on the content of their essay. This aims to help the student improve their writing, while also promoting higher-order learning processes: self-regulated learning, self-knowledge, and metacognition.

4.4.7 *AI grading*

Our final, un-named, example was developed to address the problem of marking essays for thousands of students on the EdX MOOC (Massive Open Online Courses) platform[295]. This system again uses an innovative machine learning algorithm, which was trained with hundreds of teacher-graded essays, and configured with teacher-written rubrics[296]. However, the EdX system warrants only a brief mention because it does not appear to be currently available, and details are difficult to find. However, it is included here for two reasons: because it is likely that the further developments in MOOC approaches to teaching and learning will require some way to assess student contributions at scale, and because it helped catalyse a critical reaction to the whole project of automatic essay scoring. Criticisms have been neatly summarised, and comprehensively referenced, on the *Professionals Against Machine Scoring Of Student Essays In High-Stakes Assessment* website:

---

[294] Denise Whitelock et al., 'OpenEssayist: An Automated Feedback System That Supports University Students as They Write Summative Essays', 2013, http://oro.open.ac.uk/41844/.

[295] https://www.edx.org

[296] Erin Dawna Reilly et al., 'Evaluating the Validity and Applicability of Automated Essay Scoring in Two Massive Open Online Courses', *The International Review of Research in Open and Distributed Learning* 15, no. 5 (3 October 2014), http://www.irrodl.org/index.php/irrodl/article/view/1857.

> *"Studies show that by its nature computerized essay rating is: trivial (rating essays only on surface features such as word size, topic vocabulary, and essay length), reductive (handling extended prose written only at a grade-school level), inaccurate (missing much error in student writing and finding much error where it does not exist), undiagnostic (correlating hardly at all with subsequent writing performance), unfair (discriminating against minority groups and second-language writers), and secretive (testing companies block independent research into their products)."*[297]

Finally, we should pose the diametrically reversed question. What happens when students have access to AI technologies that are capable of automatically writing (generating) high-quality essays (leading inevitably to an arms race for supremacy between automatic writing and automatic assessment[298])? For now a moot point, but almost certainly not for long.

4.5 *What other AIED is out there?*

As we saw in our discussion of ITS, any survey of applications of AI in education will always be incomplete, because new AIED applications using new AIED techniques are being launched every day. This follows the upsurge in general interest in AI and the many recent advances made possible by faster computer processors, large amounts of educational big data, and new computational approaches. In fact, because education has become a key focus for many AI developers[11] (as we noted at the outset, the market for AIED is predicted to be worth $6 billion by 2024[15]), a quick Google search will identify hundreds of AI products claiming to support students and improve learning outcomes. The EdTech consultancy GettingSmart recently published just such a search with a long list of *thirty-two* types of commercial *"applications that are (or soon will be) making good use of machine learning to support better education"*[299].

In fact, as summarised in Table 2, most existing AIED applications may be categorised in terms of five complementary dimensions: (i) the type of learners for which the AIED is designed, (ii) the learning domain that it covers, (iii) the learning approaches it facilitates, (iv) the learning support that it provides, and (v) the teaching support that it provides. AI might also be found at

---

[297] http://humanreaders.org/petition/index.php

[298] Much like the ongoing arms race between AI-generated fake news (see e.g., https://www.technologyreview.com/s/610635/fake-news-20-personalized-optimized-and-even-harder-to-stop) and AI tools to identify fake news (see e.g., http://adverifai.com).

[299] http://www.gettingsmart.com/2018/08/32-ways-ai-is-improving-education

an institutional level (i.e. outside of learning), both in learning management systems (such as MOOCs) and school management platforms (to deal with class timetabling, staff scheduling, facilities management, finances, cybersecurity, safety and security, and e-authentication). However, as we noted earlier, these administrative uses of AI in education are beyond the scope of this book.

Table 2: Dimensions of AIED applications

Dimension	Examples
type of learners	● early years ● K12 ● higher education ● informal ● professional ● students who have additional needs ● ...
learning domain	● From maths and physics, to language learning and music, and beyond...
learning approach	● step-by-step instructional adaptive learning ● dialogue-based adaptive learning ● exploratory learning ● writing analysis ● ...
learning support	● learning diagnostics ● mentoring ● assessment ● network connectors ● chatbots ● ...
teaching support	● automatic learner profiles ● smart gradebooks ● ...

One key distinction that should be noted is between AIED technologies that are designed to support students directly (i.e., student-facing tools such as the ITS, DBTS, ELE and Automatic Writing Evaluation systems that we have discussed) and AIED technologies that are designed to support teachers to support students (i.e., teacher-facing tools). We will return to this distinction later.

Here, however, we will conclude our look at applications of AI in education with six further sets of tools or technologies, some of which build on the AIED approaches that we have already mentioned, while others adopt alternative AI techniques. We begin with what we can only think to call ITS+, then AI-supported language learning, chatbots, augmented and virtual reality, and learning network orchestrators.

4.5.1 *ITS+: ALT School, ALP and Lumilo*

Our first example ITS+, by which we mean approaches that augment or extend standard ITS functionalities, is *ALT School*[300], a Silicon Valley venture founded by a former Google executive and funded by the Chan Zuckerberg Initiative. What sets ALT School apart from conventional schools is their use of a Big Data-driven ITS approach to deliver individualised learning to students throughout the whole school. Each week, all ALT School students are given an automatically-generated individual 'playlist' of activities that are designed to develop student mastery, and while they engage with the activities a vast range of data about their interactions is recorded and analysed. The results, which include each student's strengths, weaknesses, and progress, are made available to the teachers. Meanwhile, video footage of student activities, captured by classroom wall-mounted cameras, is also analysed using AI techniques, to provide indicators of student engagement. Interestingly, *ALT School* appears recently to have pivoted their business model. Perhaps inspired by our next example, they now offer their technologies to other schools rather than focusing on running their own.

Our second ITS+ is *ALP* (*Adaptive Learning Platform*), by *Kidaptive*[301], which offers an 'AIED engine as a service' to the developers of educational technologies that are not themselves enabled with AI. Partner companies connect their EdTech products to *ALP*, using client- or server-side software development kits (SDKs), which then analyses their user data in real time. Data streams from a variety of learning contexts can be aggregated to create in-depth psychometric profiles (i.e., learner models) of each individual student's interactions, preferences, and achievements. It then uses an Item-Response-Theory psychometric framework to determine the student's optimal next challenge, instructional material, or activity, which is then delivered to the student by the partner's EdTech product. *ALP* also provides

[300] www.altschool.com.
[301] http://kidaptive.com

personalised insights and recommendations to teachers and parents about the best ways in which they can support individual learners.

We have left possibly the most intriguing ITS+ for last: *Lumilo*[302]. So far only a research project, *Lumilo* uses mixed-reality smart glasses to enable teachers to access a student's real-time ITS data simply by looking at the student. In other words, *Lumilo* enables teachers to take advantage of ITS-driven adaptive learning and analytics, at the same time that they observe and engage with their classrooms as they would in a world without computers.

The tool emerged from research that suggested teachers using ITS "*wanted to be able to instantly see when a student is "stuck" (even if that student is not raising her/his hand), to instantly detect when a student is off-task or otherwise misusing the software, and to be able to see students' step-by-step reasoning, unfolding in real-time."*[303] However, while typical ITS might provide much of this information, they are not capable of registering or highlighting the many subtle cues exhibited by students that experienced teachers pick up on and use all the time.

Accordingly, the *Lumilo* transparent smart glasses superimpose real-time indicators of students' behavioural and learning states on top of the teachers' view of their classroom (in other words it functions as an Augmented Reality system, which we discuss in more detail later). As the teacher looks around the classroom, observing their students' ITS[304] activities, summary information appears, floating above each student's head. Looking at a particular student, and clicking a handheld clicker or making a specific hand-gesture, brings up a live feed of the student's screen or more detailed information (such as the number of errors they have made or the number of hints they have requested). By combining these two types of data, ITS data and teacher observations, the *Lumilo* researchers aim to enable teachers to intervene appropriately with their students, as and when they decide.

---

[302] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven, 'Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms', in *Artificial Intelligence in Education*, ed. Carolyn Penstein Rosé et al., vol. 10947 (Cham: Springer International Publishing, 2018), 154–68, https://doi.org/10.1007/978-3-319-93843-1_12.

[303] Kenneth Holstein et al., 'The Classroom as a Dashboard: Co-Designing Wearable Cognitive Augmentation for K-12 Teachers', in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18* (the 8th International Conference, Sydney, New South Wales, Australia: ACM Press, 2018), 2, https://doi.org/10.1145/3170358.3170377.

[304] Lumilo has been researched with an ITS, called *Lynette (*Vincent Aleven et al., 'Example-Tracing Tutors: Intelligent Tutor Development for Non-Programmers', *International Journal of Artificial Intelligence in Education* 26, no. 1 (2016): 224–269.), which has been designed to teach linear equations. Lynnette adaptively selects pathways using Bayesian Knowledge Tracing, and provides step-by-step guidance and feedback.

4.5.2 *Language learning: Babbel and Duolingo*

Another application of AI in education, which has recently seen considerable growth, is language learning. However, to digress briefly, potentially the most transformative application of natural language processing in recent times was the 2017 introduction of the Google Pixel Buds[305], which emerged from the research that we discussed earlier into statistical approaches to natural language processing. The Pixel Buds' algorithms, although far from perfect, are capable of translating between two spoken languages in real time, enabling two speakers who do not share a language to have a proper conversation (and finally making real some long-anticipated science fiction gadgets: remember the *Universal Translator*[306] from *Star Trek*, or the *Babel Fish*[307], from *The Hitchhiker's Guide to the Galaxy*?).

Nonetheless, currently there remain good reasons for learning another language[308]. Two of the most prominent AI-driven language learning commercial products, although there is little evidence that any are used extensively in formal educational settings, are *Babbel* and *Duolingo* (we could easily have chosen the similar *Memrise*[309], *Rosetta Stone*[310], *Mondly*[311] or many others).

Our first example is Babble[312], which has been using AI-driven speech recognition (along with typical ITS personalisation algorithms) to support language learning for about a decade. Their

[305] https://www.blog.google/products/pixel/pixel-buds

[306] http://www.startrek.com/database_article/universal-translator

[307] http://hitchhikers.wikia.com/wiki/Babel_Fish

[308] Which of course calls into question whether language learning might become as quaint as massive memorization? the CCR will be tracking this issue closely. At present its recommendation is that language acquisition matters for three reasons:
Communication, which may be superseded by translation technology, at least for conversational applications, perhaps not the ones requiring fluency
Cultural understanding, which may be taught via other mechanisms
Cognitive benefits, which are unclear through research. Will this become as indefensible as memorization after the invention of the alphabet?
While this situation unfolds, CCR's recommendation is that, given the sensitive period of language acquisition, one can easily set the foundation for multiple language acquisition through basic exposure to multiple languages. It was suggested that two languages of close linguistic distance could be mastered early (for instance for native English speakers: English at both for the first 2 years + Spanish or French by age 2-3 - both Indo-European, but one Germanic- and one Latin-based), and that maximum gains would come from a third that is very distant linguistically from a different linguistic family, and script-wise (e.g. Mandarin or Arabic) taught by age 7.
The key issue of time can be helped by technology as described herein.

[309] https://www.memrise.com

[310] https://www.rosettastone.co.uk

[311] https://app.mondly.com

[312] https://www.babbel.com

approach involves comparing student spoken words with samples of speech recorded by native speaking course editors, and giving immediate feedback to help the student improve their pronunciation. There are two main steps, recognising words and evaluating pronunciation, both of which can be challenging. To recognise words, the system first has to detect when the user starts and stops speaking, which in a typical environment means filtering out the ambient noises (such as other people speaking in the background or an airplane flying overhead). The words are then compared with the database of speech samples, first to recognise the word and then to check its pronunciation, taking into account that different people (male/female, young/old) have quite different voices (i.e., they speak at different frequencies and tempos).

With Duolingo[313], which also uses speech recognition, we will focus on their use of ITS-style personalisation. Duolingo's approach draws on two principles that are well-established in the learning sciences: the *spacing effect*[314] (we remember things more effectively if we use spaced practice, short study periods spread out over time, rather than massed practice, or 'cramming') and the *lag effect*[315] (we learn even better if the spacing between practices gradually increases)[316]. Accordingly, an algorithm (based on the Leitner Box method[317], in which flashcards answered incorrectly remain at the front of a containing box to be encountered again after only a short interval, while those answered correctly are sent to the back of the box thus delaying when they will be seen again) predicts the best time to deliver to the student a word for practice. It does so by inferring the probability that the student will correctly recall the word as a function of the lag time since the word was last practised and the half-life of the word (the 'strength' of the word in the learner's long-term memory. This is based on a student model that incorporates trace data of the learner's previous learning experiences (the number of times a student has seen the word, the number of times it was correctly recalled, and the number of times incorrect). A 2012 independent study found that students using Duolingo improved their Spanish language abilities by the equivalent of one college semester of

[313] https://www.duolingo.com

[314] David P. Ausubel and Mohamed Youssef, 'The Effect of Spaced Repetition on Meaningful Retention', *The Journal of General Psychology* 73 (1965): 147–50, https://doi.org/10.1080/00221309.1965.9711263.

[315] Arthur W. Melton, 'The Situation with Respect to the Spacing of Repetitions and Memory', 1970.

[316] Duolingo is not unique in doing this but they it is notable in having conducted various studies to optimise the approach.

[317] Sebastian Leitner, *So Lernt Man Lernen: Angewandte Lernpsychologie–Ein Weg Zum Erfolg* (Freiburg: Herder, 1995).

standard language classes (however, the study did not compare Duolingo with standard language classes or with any similar product, and no effect sizes are given).

4.5.3 *Chatbots: Ada and Freudbot*

Earlier, we met *ELIZA*, the first computer programme that appeared able to converse in natural language, in other words the first AI *chatbot*. Now, after fifty years of development, chatbots are becoming mainstream[318], with the launch of digital assistants from tech's Big Five: Amazon (*Alexa*[51]), Apple (*Siri*[2]), Facebook (*Messenger*[319]), Google (*Assistant*[50]), and Microsoft (*Cortana*[47]). Nonetheless, progress has not always been straightforward (remember the racist rants tweeted by Microsoft's chatbot *Tay*[320]?), and no chatbot has yet convincingly passed the Turing Test (if a human cannot decide whether a computer is human or a computer, the computer is said to have passed the Turing Test[321]). Having said that, recently the Google *Duplex* chatbot was presented making a restaurant reservation and booking a hairdresser appointment (however, the demonstration was clearly carefully orchestrated and more than a little controversial[322]). Nonetheless, today, one *chatbots-as-a-service* company alone[323] claims that more than 300,000 bots have been created using its toolsets, and this is just one of many such platforms[324].

In general, chatbots are designed to respond automatically to messages (SMS texts, website chats, social messaging posts, and voice), either using rules and keywords to select from pre-programmed scripted responses (as with *ELIZA* and most current simple bots), or adaptive

---

[318] e.g., Robert Dale, 'The Return of the Chatbots', *Natural Language Engineering* 22, no. 5 (2016): 811–817, and 'Everything you ever wanted to know about chatbots (but were afraid to ask)' https://www.jisc.ac.uk/blog/everything-you-ever-wanted-to-know-about-chatbots-but-were-afraid-to-ask-08-oct-2018

[319] https://messenger.fb.com

[320] M. J. Wolf, K. Miller, and F. S. Grodzinsky, 'Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications', *SIGCAS Comput. Soc.* 47, no. 3 (September 2017): 54–64, https://doi.org/10.1145/3144592.3144598.

[321] The Turing Test, or more correctly the 'Imitation Game', was devised by Alan Turing (who is regarded by many as the father of both modern computing and artificial intelligence), to determine whether we might consider a computer intelligent: *"I believe that in about fifty years' time it will be possible to programme computers… to make them play the imitation game so well that an average interrogator will not have more that 70 per cent chance of making the right identification after five minutes of questioning."* Turing, Alan, 'Computing Machinery and Intelligence', Mind 59, no. 236 (1950): 433–460.

[322] See both https://www.extremetech.com/computing/269030-did-google-duplexs-ai-demonstration-just-pass-the-turing-test and https://www.extremetech.com/computing/269497-did-google-fake-its-google-duplex-ai-demo

[323] https://home.pandorabots.com

[324] https://www.techworld.com/picture-gallery/apps-wearables/platforms-for-developers-build-chatbots-3639106

machine learning algorithms to generate unique responses (as with the more sophisticated bots such as *Siri*, *Duplex* and *Tay*). Chatbots are quickly becoming ubiquitous, for everything from booking a flight[325], to ordering food[326], as a doctor[327] or a financial adviser[328], in recruitment[329] and accounting[330], to help with shopping[331], as a personal 'companion'[332], and to support young people who are suffering from anxiety[333].

Chatbots are also increasingly being used in educational contexts, again for a variety of purposes. For example, students making initial enquiries about courses may find themselves conversing with a bot whose job it is to direct them to the information that they want, or otherwise to a human adviser[334]. In some configurations[335], they can also provide ongoing student support and guidance, in academic services, accommodation, facilities, examinations, IT, health and more[336]. And in some situations, they can also be used to directly support learning – indeed, the DBTS that we met earlier (including *AutoTutor*[142] and *Watson Tutor*[149]), may be considered special cases of educational chatbots). For example, chatbots might provide feedback to support student reflection and self-efficacy[337], while already some language learning apps use chatbots for embarrassment-free speaking practice in simulated real-life situations[338].

However this is not to suggest that chatbots are an educational silver bullet. For example, a student's *"willingness to communicate in a foreign language... seems to decline rapidly over*

---

325 https://bb.klm.com/en

326 https://www.tacobell.com/Tacobot

327 https://www.your.md

328 https://www.rbs.com/rbs/news/2016/03/rbs-installs-advanced-human-ai-to-help-staff-answer-customer-que.html

329 https://hiremya.com

330 https://www.sage.com/en-gb/products/pegg

331 https://bots.kik.com/#/vspink

332 https://www.pandorabots.com/mitsuku

333 Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile, 'Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial', *JMIR Mental Health* 4, no. 2 (6 June 2017): e19, https://doi.org/10.2196/mental.7785.

334 https://www.virtualspirits.com/chatbot-for-university.aspx

335 https://www.slu.edu/alexa/index.php

336 e.g., Deakin University are using IBM Watson to run a student services support chatbot: http://www.deakin.edu.au/about-deakin/media-releases/articles/ibm-watson-helps-deakin-drive-the-digital-frontier, while the Open University of Hong Kong have launched *The i-Counseling System* https://library.educause.edu/resources/2012/5/case-study-9-the-open-university-of-hong-kong-the-icounseling-system

337 Karsten O. Lundqvist, Guy Pursey, and Shirley Williams, 'Design and Implementation of Conversational Agents for Harvesting Feedback in ELearning Systems', in *European Conference on Technology Enhanced Learning* (Springer, 2013), 617–618.

338 http://bots.duolingo.com

time as students lose interest in chatbots as language partners compared to human learning partners. This could happen because of a simple novelty effect or simply the weaker value of chatbots compared to human assistants"[339].

Typical education chatbots (in addition to the DBTS discussed earlier) are *Ada* and *Freudbot*. *Ada*[340] has been developed by a UK community college, using IBM's Watson Conversation platform. In short, *Ada* demonstrates how education chatbots have been implemented using limited resources and *AI-as-a-service* technologies. Available on multiple devices (desktop, mobile and kiosk), and named after the computer pioneer Ada Lovelace, *Ada* is able to respond to a spectrum of student enquiries, delivering personalised and contextualised responses that draw on data such as the student's courses, their progress, their goals and their individual targets. *Ada* is already able to respond to questions about the library, student services, finance, accommodation, transport, careers, and examinations – and it learns more with every interaction. For example, a student might ask about their lessons that morning, or where tomorrow's exam is happening, or what mark they achieved in a recent assignment. Meanwhile, a teacher can ask for a list of professional development workshops they have recently attended, or about a specific student's academic performance.

An earlier chatbot (and in effect a primitive DBTS) called *Freudbot*[341] engaged students in a conversation about an educational topic (rather than being designed to provide students with information about their institution and studies, as in the case of *Ada*). It took on the role and persona of Sigmund Freud, chatting in the first person with introductory psychology students about his theories and life. *Freudbot* was developed before the availability of easily-accessed machine learning techniques, and so used rules and recognised keywords to select from pre-programmed scripted responses, drawing on a university resource comprising explanations of Freudian terms and concepts together with an open-source biography. And when a student's question or response was outside its rule-base, *Freudbot* would default to a backstop strategy such as asking for clarification or suggesting a new topic for discussion, always with the aim of leading the user back to the core discussion of Freudian topics.

---

339 Rainer Winkler and Matthias Soellner, 'Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis', *Academy of Management Proceedings* 2018, no. 1 (July 2018): 17, https://doi.org/10.5465/AMBPP.2018.15903abstract.
340 http://www.aftabhussain.com/ada.html
341 Bob Heller et al., 'Freudbot: An Investigation of Chatbot Technology in Distance Education', in *EdMedia: World Conference on Educational Media and Technology* (Association for the Advancement of Computing in Education (AACE), 2005), 3913–3918.

4.5.4 *Augmented and virtual reality*

Virtual reality (VR) and augmented reality (AR) are two innovations that many have been trying to apply in educational contexts[342] (for example, Google have developed for educational contexts more than a 1000 VR and AR *Expeditions*[343]). VR provides an immersion experience that shuts out the physical world, enabling users who are wearing appropriate VR goggles to be transported into a vast range of real-world or imagined environments such as the International Space Station, an operating theatre[344], or Hogwarts castle[345]. AR, on the other hand, as we saw earlier with *Lumilo*, overlays computer-generated images on top of the user's view of the real world (much like a fighter pilot's heads up display[346]), when seen through a smartphone or other similar device – the aim being to enhance or mediate the user's view of reality. For example, when a smartphone's camera is pointed at a mountain range, the names of the mountains and their elevations might be superimposed[347], while pointing at a particular QR code might reveal a 3D human heart that can be explored in detail[348], and in a particular street location a Pokémon character might be found waiting to be caught[349]. Although not traditionally thought of as AI technologies, we have mentioned VR and AR here because both are frequently combined with AI machine learning, image recognition and natural language processing, with the aim of further enhancing the user experience[350].

While the thought of thirty students all wearing goggles and immersed in another world might strike fear in the heart of a classroom teacher, and while *"VR does not intrinsically make every experience better in terms of motivation and learning"*[351], used judiciously both VR and AR do have potential to become useful tools in the educator's toolbox. To give a few brief examples: VR has been used effectively to give breast cancer patients an anxiety-relieving virtual

---

342 e.g., http://www.classvr.com
343 https://edu.google.com/expeditions
344 http://ossovr.com
345 https://www.pottermore.com/news/new-expanded-fantastic-beasts-and-where-to-find-them-vr-experience-announced
346 https://www.youtube.com/watch?v=Ay6g66FbkmQ
347 https://www.peakfinder.org
348 https://medmovie.com/augmented-reality-heart
349 https://www.pokemongo.com/en-gb
350 e.g., https://www.apple.com/uk/ios/augmented-reality, https://www.samsung.com/global/galaxy/galaxy-s9/augmented-reality and https://ametroslearning.com
351 Chris Dede et al., 'Virtual Reality as an Immersive Medium for Authentic Simulations', 2017, 133–56, https://doi.org/10.1007/978-981-10-5490-7_8.

experience of the radiotherapy process, tailored to each individual patient[352], VR simulations have been used extensively to provide training for neurosurgical residents on a variety of neurosurgical procedures[353] and to enable students to interact directly with historical characters[354], while a VR classroom has been used to provide trainee teachers with an *"absorbing, realistic and interactive virtual classroom, allowing them to engage in realistic interactions with virtual students."*[355] Researchers have also proposed the use of VR to enhance student experiences in immersive simulations such as *EcoMUVE*[356], an immersive multi-user virtual environment and associated inquiry-based curriculum developed at Harvard University. *EcoMUVE* was designed to enable school students to learn about ecosystems by playing the role of a scientist, exploring and collecting data in a virtual ecosystem in order to answer research questions. Although a VR interface might make some tasks more difficult, the researchers suggest that it has the potential to make the simulation more 'realistic', by increasing the students' feeling of being present in the simulated environment, which in turn is likely to enhance transfer of the learning from the virtual to the real world[357].

AR, on the other hand, has been used to enable students to explore and manipulate three dimensional models of organic molecules in order to enhance their understanding of chemistry[358], to help primary school students learn about history[359], and in an AR-enabled

---

[352] Yobelli A. Jimenez et al., 'Patient Education Using Virtual Reality Increases Knowledge and Positive Experience for Breast Cancer Patients Undergoing Radiation Therapy', *Supportive Care in Cancer* 26, no. 8 (August 2018): 2879–88, https://doi.org/10.1007/s00520-018-4114-4.

[353] Laura Stone McGuire and Ali Alaraj, 'Competency Assessment in Virtual Reality-Based Simulation in Neurosurgical Training', in *Comprehensive Healthcare Simulation: Neurosurgery* (Springer, 2018), 153–157.

[354] Baierle, Ivan Luis Feix, and João Carlos Gluz. 'Programming Intelligent Embodied Pedagogical Agents to Teach the Beginnings of Industrial Revolution'. In Intelligent Tutoring Systems, edited by Roger Nkambou, Roger Azevedo, and Julita Vassileva, 10858:3–12. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-91464-0_1.

[355] Kalliopi Evangelia Stavroulia et al., 'Designing a Virtual Environment for Teacher Training: Enhancing Presence and Empathy', in *Proceedings of Computer Graphics International 2018 on - CGI 2018* (Computer Graphics International 2018, Bintan, Island, Indonesia: ACM Press, 2018), 273, https://doi.org/10.1145/3208159.3208177.

[356] http://ecolearn.gse.harvard.edu

[357] See Dede et al. (2017), in which they discuss in depth the potential and limitations of VR for simulated environments, and suggest some useful principles for effective implementation.

[358] Derek Behmke et al., 'Augmented Reality Chemistry: Transforming 2-D Molecular Representations into Interactive 3-D Structures', *Proceedings of the Interdisciplinary STEM Teaching and Learning Conference* 2, no. 1 (1 January 2018), https://doi.org/10.20429/stem.2018.020103.

[359] Irene Efstathiou, Eleni A. Kyza, and Yiannis Georgiou, 'An Inquiry-Based Augmented Reality Mobile Learning Approach to Fostering Primary School Students' Historical Reasoning in Non-Formal Settings', *Interactive Learning Environments* 26, no. 1 (2 January 2018): 22–41, https://doi.org/10.1080/10494820.2016.1276076.

digital games-based learning environment to support students' reading comprehension[360]. These few examples only touch the surface of the research that has investigated VR and AR in education[361].

4.5.5 *Learning Network Orchestrators: Third Space Learning and Smart Learning Partner.*

Learning Network Orchestrators (LNO)[362], are tools or approaches that enable and support networks of people engaged in learning (e.g., students and their peers, students and their teachers, or students and people from industry), in order to promote and enhance the learning. LNOs typically match participants based on their availability, the subject domain, and the participants' varied expertise, and can facilitate coordination and cooperation between them: *"participants can interact with one another, share their learning experiences, build relationships, share advice, give reviews, collaborate, co-create and more"* [363]. AI techniques are slowly being introduced to LNO products, to automate much of this orchestration, opening up network possibilities previously unachievable.

For example, in a novel approach, Third Space Learning connects UK primary school children who are at risk of failure in mathematics with mathematics tutors in India and Sri Lanka. The system supports individual tutoring, with tutors and students communicating with each other by means of a secure online virtual classroom that has two-way audio and a shared interactive whiteboard. AI is being introduced to automatically monitor every session, thousands of hours of teaching and learning every week, which generates huge quantities of data. Algorithms then aim to guide the tutor with real-time feedback, ensuring that teaching broadly follows an outline script that adopts well-supported learning sciences principles, identifying where students have misconceptions not picked up by the tutor (by comparing individual and overall student models), and empowering the tutors to constantly improve their teaching skills.

---

[360] Hendrys Tobar-Muñoz, Silvia Baldiris, and Ramon Fabregat, 'Augmented Reality Game-Based Learning: Enriching Students' Experience During Reading Comprehension Activities', *Journal of Educational Computing Research* 55, no. 7 (December 2017): 901–36, https://doi.org/10.1177/0735633116689789.

[361] Iulian Radu, 'Augmented Reality in Education: A Meta-Review and Cross-Media Analysis', *Personal and Ubiquitous Computing* 18, no. 6 (2014): 1533–1543.

[362] e.g., Nepris (https://www.nepris.com) and Educurious (https://educurious.org), which both support schools to connect with experts from around the world, to bring an industry perspective into the classroom. Possibilities include interactive question and answer sessions, virtual field trips, and project mentorships.

[363] Holmes et al., 'Technology-Enhanced Personalised Learning. Untangling the Evidence.'

The *Smart Learning Partner*[364], on the other hand, uses much simpler AI technologies to put students more in control of their own learning. It is the result of a collaboration between Beijing Normal University's Advanced Innovation Center for Future Education and Tongzhou district of Beijing, in China. A key component of the *Smart Learning Partner* system is an AI-driven platform that enables students to connect with a human tutor via their mobile phones. The platform uses AI somewhat like a dating app – except it matches students and tutors according to student queries and tutor areas of expertise, together with the tutor's availability and ratings given to them by other students whom they have already tutored. The student uses the app to search for a tutor, to ask what they want to know about any school topic, and they then receive twenty minutes of one-to-one online tuition (sharing audio and screens only).

5 What else is possible?

Much of the AIED that we have discussed so far involves the application of AI techniques to mainstream learning approaches, and tends to reflect (or automate) existing educational assumptions and practices. In addition, although we have seen some notable exceptions, much AIED has been designed (whether intentionally or not) to supplant teachers or to reduce them to a functional role[365], and not to assist them to teach more effectively. This approach, while potentially useful in contexts where teachers are few and far between, clearly undervalues teachers' unique skills and experiences, as well as learners' needs for social learning and guidance. However, instead of just automating the teaching of students sat at computers, conceivably AI might help open up teaching and learning possibilities that are otherwise difficult to achieve, that challenge existing pedagogies, or that help teachers to be more effective. Here we will speculate on some possibilities, some of which have been foreshadowed by the AIED tools we have already discussed, while others are both novel and complex to achieve, and all of which raise interesting social questions. We begin with AI to support collaborative learning, then AI-driven student forum monitoring, AI to support

---

[364] http://slp.bnu.edu.cn (NB Only accessible to students and faculty who have an account.)

[365] Worryingly, one of the developers we have mentioned has suggested that the sophistication of their AIED means that teachers only need to play an auxiliary role, working like fast-food chefs (*"KFC-like"*) to strictly regulated scripts.