AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling

Cristina Conati [1]   **Kaśka Porayska-Pomsta** [2]   **Manolis Mavrikis** [2]

Abstract

Interpretability of the underlying AI representations is a key raison d'être for Open Learner Modelling (OLM) – a branch of Intelligent Tutoring Systems (ITS) research. OLMs provide tools for 'opening' up the AI models of learners' cognition and emotions for the purpose of supporting human learning and teaching. Over thirty years of research in ITS (also known as AI in Education) produced important work, which informs about how AI can be used in Education to best effects and, through the OLM research, what are the necessary considerations to make it interpretable and explainable for the benefit of learning. We argue that this work can provide a valuable starting point for a framework of interpretable AI, and as such is of relevance to the application of both knowledge-based and machine learning systems in other high-stakes contexts, beyond education.

1. Introduction of the ITS field

Since the early 1970s, the field of Intelligent Tutoring Systems (ITS – also known as Artificial Intelligence in Education) has been investigating how to leverage AI techniques to create educational technologies that are personalised to the needs of individual learners, with the goal to approximate the well-known benefits of one-to-one instruction (for a recent review see du Boulay, 2016). The idea is essentially to devise intelligent pedagogical agents (IPAs) that can model, predict and monitor relevant learner behaviours, abilities and mental states in a variety of educational activities, and provide personalised help and feedback accordingly (Woolf, 2009). These IPAs need to be

able to capture and process information about three main components of the teaching process: (i) the target instructional domain (*domain model*), (ii) the relevant pedagogical knowledge (*pedagogical model*), and (iii) the student themselves (*student model*). These three components define a conceptual architecture of instructional modelling and interaction that emerged from the ITS research over the years (e.g., du Boulay, 2016). Of these components, the student model constitutes a defining characteristic of an ITS. Non-AI educational technologies, such as test-and-branch systems, generally provide instructional feedback by matching students' responses against some pre-programmed solutions (e.g., (Nesbit et al., 2014)). ITS research, on the other hand, strives to provide deeper and more precise pedagogical interventions by modelling in real-time a variety of features that are important for individualised instruction, such as students' domain knowledge and cognitive skills, as well as their meta-cognitive abilities and affective states. Here, it is noteworthy that it is the need for delivering appropriate pedagogical interventions that makes the educational context a high-stakes one for AI: such interventions may have potentially long-lasting impact on peoples learning, development and life-long functioning.

ITS research has successfully delivered techniques and systems that provide personalised support for problem solving activities in a variety of domains (e.g., programming, physics, algebra, geometry, SQL), based on an on-going student modelling assessment of the evolving student domain knowledge during interaction with the system. Formal studies have shown that these systems can foster students' learning better than practicing problem solving skills in class or smaller group contexts, and that the outcomes which are measured in terms of improvements in students' grades are comparable with those achieved through human tutoring (Schroeder et al., 2013; Nesbit et al., 2014; du Boulay, 2016; Ma et al., 2014). Some of the ITS are actively used in real-world settings e.g.,(Mitrovic et al., 2007; Koedinger & Aleven, 2016), reaching several thousand schools in the US alone and even changing traditional school curricula. The growing shortage of qualified teachers, coupled with growing numbers of students world-wide, represents a substantial societal global challenge and provides a strong motivation to continue investing in ITS re-

[1] University of British Columbia, Department of Computer Science, Canada; [2] University College London, UCL Knowledge Lab, United Kingdom. Correspondence to: Kaśka Porayska-Pomsta <K.Porayska-Pomsta@ucl.ac.uk>.

search and solutions to enhance and support human learning and development in an accountable way and at scale.

2. New developments in ITS and the need for machine learning

For reasons related to the ITS research having started in the tradition of symbolic, rule-based expert systems, as well as owing to the nature of the educational domain where inferential transparency is key to delivering pedagogically effective instruction, a large proportion of ITS is based on knowledge-based techniques. However, there is an emerging appetite and need in the field for machine learning approaches, which is fuelled by a combination of (i) the emergence of big data, e.g. in learning and teaching at scale contexts such as through Massive Open Online Courses (MOOCs), and (ii) a shift within the field towards more complex instructional domains, for which it may be harder to engineer and represent knowledge based on traditional knowledge elicitation from human experts. Specifically, in addition to ITS for problem solving, researchers have been investigating ITSs for a variety of other educational activities in more complex domains that can benefit from individualised support, such as learning from examples (?Long & Aleven, 2017), learning by exploration or games (Conati & Kardan, 2013a; Porayska-Pomsta et al., 2013; Grawemeyer et al., 2017), or learning by teaching (Biswas et al., 2005). Providing individualised support for these activities poses unique challenges, because it requires an ITS to model the activities as well as student behaviours, abilities and states that may not be as well-defined, understood or easily captured as those involved in problem solving. For instance, an ITS that provides support for exploration-based learning must be able to "know" what it means to explore a given concept or domain effectively (e.g. via an interactive simulation) so that it can monitor the students' exploration process and provide adequate feedback when needed. It might also need to capture and model the domain-independent learner abilities that foster good exploratory behaviour (e.g. self-assessment, planning).

Machine learning (ML) techniques are instrumental in addressing the challenges of these new endeavours, because they can help learn from data the knowledge and models that might be challenging to obtain from human experts, and compute predictions of students' cognitive and mental states in highly dimensional and ill-defined spaces of human behaviours. Examples of ML applications in ITS include modelling student states and abilities such as self-efficacy (Mavrikis, 2010), emotional reactions (Conati & Maclaren, 2009; Bosch et al., 2016; Monkaresi et al., 2016), predicting students' ability to successfully conduct scientific inquiries in virtual environments (Baker et al., 2016), and automatically generating hints (Stamper et al.,

2011; Conati & Kardan, 2013b; Fratamico et al., 2017). We argue that the interpretability of these techniques, and indeed any other AI techniques employed in an ITS, is critical to enabling an IPA to explain to its users its inferences and actions. The importance of such explanations is two-fold. First, they can improve an IPA's pedagogical effectiveness because they often form an integral part of the skills to be taught (e.g., to help students understand why the system deems their answers to be incorrect or a particular topic to be learned or not).

Second, as in other high-stakes contexts which employ AI for decision-making, an IPA's ability to explain its decisions relates to nurturing users' trust in such decisions (e.g., Kostakos & Musolesi, 2017). In the ITS context this includes students' trust and their consequent willingness to follow the IPA's suggestions, as well as teachers' trust, which is key for the adoption of these technologies. ITS researchers have yet to investigate systematically to whom (e.g. student or teacher, or either), why, when and to what extent interpretability and the consequent explainability of an IPA's underlying models can be beneficial. However, in the next section we discuss initial evidence pertaining specifically to the benefits of having interpretable and explainable student models, coming from a branch of ITS research known as Open Learner Modelling (OLM). This research also offers an emergent conceptual framework (outlined in the final section) for understanding the key criteria for interpretable and explainable AI in educational applications. We argue that this conceptual framework along with the examples of different approaches to OLMs may be of relevance to machine learning use in other high-stakes contexts, beyond application in education, in which interpretability, explainability and user control over AI is a requirement.

3. Open Learner Modelling and Interpretability

Open Learner Models (OLMs) are student models that allow users to access their content with varying levels of interactivity (Bull, 1995; Bull & Kay, 2016). For example, an OLM can be:

- *scrutable*, i.e. users may view the models current evaluation of relevant student's states and abilities (henceforth – assessment);

- *cooperative* or *negotiable*, where the user and the system work together to arrive at an assessment,

- *editable*, namely the user can directly change the model assessment and system's representation of their knowledge at will.

Traditionally, OLMs have been designed for students as users of an ITS, with two main purposes: one, pedagogical in nature, is to encourage effective learning skills such as self-assessment and reflection; the second is to improve model accuracy by enabling students to adjust the model's predictions or even its underlying representation when such are deemed inaccurate by the students. Clearly, even OLMs that are merely scrutable require having an underlying representation that is interpretable at some level, so that the model's assessment can be visualised for and understood by its users. However, the more interactive the OLM, the more interpretable and explainable the underlying representations may need to be, because of the increased control that the user has over the access to the different aspects of the model. For example, in a type of negotiation OLM developed by Mabbott and Bull (2006), the user can register their disagreement with the system's assessment and propose a change. At this point, the system will explain why it believes its current assessment to be correct by providing evidence to support these beliefs, e.g. samples of the learners' previous responses that may indicate a misconception. If the user still disagrees with the system, they have a chance to 'convince' it by answering a series of targeted questions that the system generates. In order to do this, the system keeps a detailed representation of the user's on-task interactions along its assessments of the user's skills.

In the rest of this section we will provide examples of existing OLMs and of their benefits on pedagogical outcomes.

The TARDIS system is an example of ITS that includes a scrutable OLM, namely an OLM that allows students to see the model assessment, but not to interact with the model. TARDIS offers a job interview coaching environment for young people at risk of social exclusion through unemployment. TARDIS includes AI virtual agents which act as recruiters in different job interview scenarios. TARDIS collects evidence from the virtual interviews, based on low-level signals such as the users gaze patterns, gestures, posture, voice activation, etc., and uses machine learning techniques to predict from this evidence the quality of behaviours known to be important for effective interviews (e.g. appropriate energy in the voice, fluidity of speech, maintenance of gaze with that of the interviewer) (Porayska-Pomsta et al., 2014).

The model's assessment over these behaviours is then visualised to the learner as shown by the pie charts in Fig. 1, as a way to provide the users with a concrete and immediate basis for reflecting on how they may improve their verbal and non-verbal behaviours in subsequent interviews with the AI agents. The learner is also shown a time-lined view of the low-level signals and the interpretation thereof.

The information in Fig. 1 is further used by human prac-

titioners to provide nuanced discussion of the learner behaviours in follow-up briefing sessions, showing the importance of interpretable models such as TARDIS's for enhancing human teaching practices. In fact, despite the relatively shallow nature of the information provided by the TARDIS's OLM, a controlled evaluation study with 28 high-risk students, aged 16-18, (14 in OLM and 14 in no OLM condition) showed significant improvements in key behaviours for the OLM condition (Porayska-Pomsta & Chryssafidou, 2018).

Our next example is that by (Long & Aleven, 2017) where they propose a system that uses a scrutable OLM to help students improve their ability to self-assess their knowledge and share the responsibility for selecting the next problem to work on. The system relies on a student model that is built using the technique known as example-tracing (Aleven et al., 2016): the system evaluates students' problem-solving steps against typical examples of correct problem-solving steps, which are represented in terms of behaviour graphs that encode the different problem-solving strategies applicable for a given problem. Each problem-solving step is related to a piece of domain knowledge (knowledge component, or KC) that needs to be known in order to generate the step. Thus, the evaluation of student's problem-solving steps against the example solutions are used by the system as evidence to generate a probabilistic assessment of student's knowledge (or lack thereof) of the corresponding KC. This process is known as Bayesian Knowledge Tracing (BKT), which has been employed by many ITS to date (Aleven & Koedinger, 2013). The probabilities over KCs generated by BKT are visualised to students in terms of 'skill bars' or skillometer (Fig.1 2). To support the students' learning and to foster reflection skills, in this OLM the students are asked to self-assess their knowledge of the specific KCs before they can see the system's assessment. The visualisation of such assessment is designed to draw the student's attention to how the assessment changes in response to student's problem-solving actions. That is, once the student asks to see the system's assessment in the skillometer, the relevant bars grow/shrink to new places, based on students' updated skill mastery after finishing the current problem, and as calculated by BKT (Koedinger & Corbett, 2006). The dynamic updating of the bars is a form of feedback on students' self-assessment. Here, student self-reflection constitutes an explicit learning goal, which has been shown through a formal user study to significantly improve learning outcomes for the students who used this OLM (Long & Aleven, 2017) .

Our final example is of fully editable OLM, where users have the greatest control over their model. In an early evaluation of user preferences with respect to different forms of OLMs conducted by (Mabbott & Bull, 2006), editable models have been shown to lead to a decreased user trust,
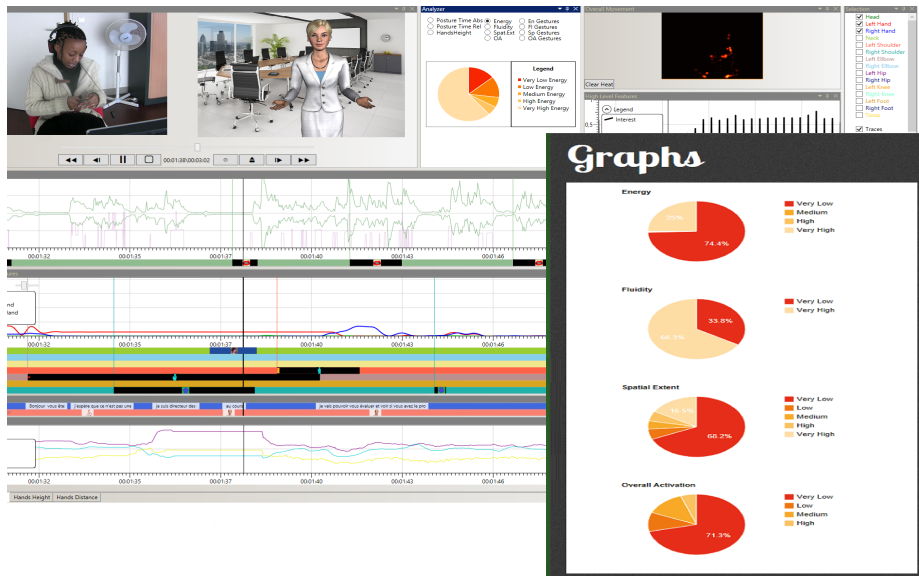
Figure 1. TARDIS scrutable OLM showing synchronised recordings of the learners interacting with the AI agents along with the interpretation of the learner's low level social signals such as gaze patterns, gestures, voice activation in terms of higher level judgements about the quality of those behaviours, e.g. energy in voice.

especially if the users were novice learners who lacked confidence in their own judgments. More recent examples, however, show that such models can provide effective, engaging and trusted learning tools if they are accompanied by fine-grained support from the system. This in turn necessitates access to detailed model representations and inferences. In a radical approach, (Basu et al., 2017) implemented an editable OLM which allows students to build models of their knowledge by exploring concepts, properties and relations between them in open ended exploratory learning environments (OELE). To achieve this, they employed a hierarchical representation of tasks and strategies (implemented as a directed acyclic graph) that may need to be undertaken to solve a problem. this representation was that it allowed for the expression of a particular construct or strategy in multiple variations that related to each other, which in turn gave the system access to a set of desired and suboptimal implementations of a task or strategy employed by the user. Based on this, the system can analyse the users behaviour by comparing their action sequences and their associated contexts against desired and suboptimal strategy variants defined in the strategy model and, in turn, to offer targeted support when the users seem to flounder. This representation allows for a conceptual support to be given to the user at a fine-grained level of detail, e.g. low-level objects description in terms of their properties, relations between them and temporal ordering of actions that could be performed on them. This allows the system to guide the user in model building through relatively simple step-by-step interfaces for the different modelling tasks, gradually
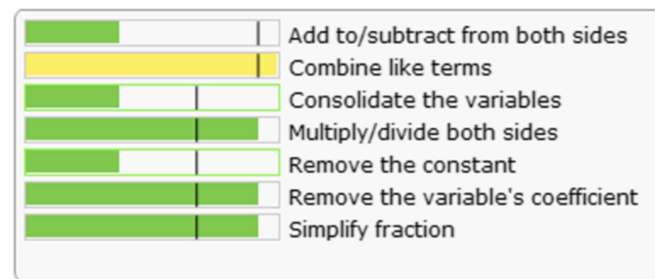

Figure 2. Long and Aleven's (2017) skills meter bars indicating the level of student skill mastery

building users' confidence in their abilities, their buy-in to the system's advice and prompts, ultimately significantly increasing the learning outcomes for the users (Basu et al., 2017).

4. Discussion and Future Work

Section 2 established the need for models based on machine learning. As educational technology is deployed at scale, and computational power no longer presents a barrier to adoption, machine learning is used increasingly for cognitive and non-cognitive student modelling. However, to make ML-based models that can be meaningfully employed in the context of supporting human learning and teaching exposes them to demands related to their interpretability, explainability and ultimately trustworthiness (Weller, 2017). Despite the fact that the vast majority of the

OLMs developed over the past thirty years are knowledge-based, the insights offered by the substantial ITS work we introduced here provide an important conceptual and practical framework for developing ML-based models in education, with potential application in other high-stakes contexts concerned with modelling of human behaviour and decision-making.

Initial evaluation studies of the different types of OLMs have started to shed some light on key considerations that need to be taken into account when deciding on both what information to reveal to the user, how and why, as well as to what extent this information needs to approximate the underlying representations of the AI models. The way in which those questions are addressed will have implications for how effective the learning support delivered by an ITS will be. Comparison studies such as those conducted by (Mabbott & Bull, 2006; Kerly, 2009) provide initial indications on user preferences for the particular types of OLMs, along with the implications on user trust and improvements in model accuracy that such models engender. For instance, (Mabbott & Bull, 2006) present anecdotal evidence that the maximum level of control facilitated by the editable OLM they tested was the least favoured by the users in their study, compared to non-editable variations of this OLM, because learners did not trust their own judgments and expected targeted support from the system (Mabbott & Bull, 2006). When such support is not available their trust in the system tends to dwindle along with their motivation to follow the systems instruction. Negotiable or co-operative open learner models that maintain an interaction symmetry where both the system and the learner have to justify and explain their actions, represent the preferred and trusted by users mode for their engagement with such models. Thus, finding the balance between the level of control to be given to the user over the content of their model and the level of system?s support in changing the model that can be offered to them seems critical to deciding the level of algorithmic interpretability needed.

As a summary of the key considerations, we cite four dimensions (expressed as questions) as proposed by (Bull & Kay, 2016):

1. *Why* the OLM is being built, e.g. to improve model accuracy, to engender user right of access and trust, to nurture self-regulation and reflection, etc.;

2. *Which* aspects of the model are made available to the user. Examples include the extent of the learner model that is open; closeness of the externalisation of the learner model to the underlying model representations; extent of access to (un)certainty in the model's assessment; access to different temporal views, e.g. current, past, anticipated future models; access to sources of input to the model; access to explanation of the relationship between the learner model and personalisation of the interventions based on such a model;

3. *How* is the model accessed and the degree to which it can be manipulated by the user, such as visualisation used in the OLM; type of interactivity with the model; flexibility of access to the model;

4. *Who* has access to the model, e.g. intended users such as students, teachers, parents.

The four dimensions allow OLM architects to calibrate, at least in principle, pedagogically optimal designs of those tools. Much research is still needed to deliver a universal framework for interpretable AI and to understand better how to make ML-based OLMs viable in supporting human learning and development at scale. Nevertheless, we propose that the examples and the preliminary empirical findings gleaned from the work on OLMs in the context of Artificial Intelligence in Education have implications for how we may address the need for interpretability of AI models, be it knowledge- or ML-based, and offer additional, and thus far, largely ignored conceptual starting point from ITS research for consideration by a wider AI and machine learning community.

References

Aleven, V. and Koedinger, K.R. Knowledge component approaches to learner modeling. In *Design Recommendations for Adaptive Intelligent Tutoring Systems*, volume Volume 1 of *Learner Modeling*, pp. 165–182. US Army Research Laboratory, Orlando, Florida, r.sottilare, a. graesser, x. hu, & h. holden edition, 2013. ISBN 978-0-9893923-0-3.

Aleven, V., McLaren, B M., Sewall, J., van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., and Koedinger, K R. Example-Tracing Tutors: Intelligent Tutor Development for Non-programmers. *International Journal of Artificial Intelligence in Education*, 26(1):224–269, March 2016. ISSN 1560-4306. doi: 10.1007/s40593-015-0088-2. URL https://doi.org/10.1007/s40593-015-0088-2.

Baker, R.S., Clarke-Midura, J., and Ocumpaugh, J. Towards general models of effective science inquiry in virtual performance assessments. *J. Comp. Assist. Learn.*, 32(3):267–280, June 2016. ISSN 0266-4909. doi: 10.1111/jcal.12128. URL https://doi.org/10.1111/jcal.12128.

Basu, S., Biswas, G., and Kinnebrew, J. S. Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. *User*

Modeling and User-Adapted Interaction, 27(1):5–53, March 2017. ISSN 0924-1868. doi: 10.1007/s11257-017-9187-0. URL https://doi.org/10.1007/s11257-017-9187-0.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., and at Vanderbilt, The Teachable Agents Group. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4):363–392, 2005. doi: 10.1080/08839510590910200. URL https://doi.org/10.1080/08839510590910200.

Bosch, N., D'Mello, S., and Baker, R. Emotions in computer-enabled classrooms. In *2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT)*, pp. 4125–4129, 2016.

Bull, S. Did I say what i think i said, and do you agree with me?: Inspecting and Questioning the Student Model. 1995.

Bull, S. and Kay, J. SMILI?: A Framework for Interfaces to Learning Data in Open Learner Models, Learning Analytics and Related Fields. *International Journal of Artificial Intelligence in Education*, 26(1):293–331, Mar 2016. ISSN 1560-4306. doi: 10.1007/s40593-015-0090-8. URL https://doi.org/10.1007/s40593-015-0090-8.

Conati, C. and Kardan, S. Student Modeling: Supporting Personalized Instruction, from Problem Solving to Exploratory Open Ended Activities. *AI Magazine*, 34(3):13–26, September 2013a. ISSN 0738-4602. doi: 10.1609/aimag.v34i3.2483. URL https://aaai.org/ojs/index.php/aimagazine/article/view/2483.

Conati, C. and Kardan, S. Student Modeling: Supporting Personalized Instruction, from Problem Solving to Exploratory Open Ended Activities. *AI Magazine*, 34(3):13–26, September 2013b. ISSN 0738-4602. doi: 10.1609/aimag.v34i3.2483. URL https://aaai.org/ojs/index.php/aimagazine/article/view/2483.

Conati, C. and Maclaren, H. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3):267?303, 2009. doi: 10.1007/s11257-009-9062-8.

du Boulay, B. Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems*, 31(6):76–81, Nov 2016. ISSN 1541-1672. doi: 10.1109/MIS.2016.93.

Fratamico, L., Conati, C., Kardan, S., and Roll, I. 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 (2):320–352, Jun 2017. ISSN 1560-4306. doi: 10.1007/s40593-016-0131-y. URL https://doi.org/10.1007/s40593-016-0131-y.

Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., and Rummel, N. Affective learning: Improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction*, 27(1):119–158, March 2017. ISSN 0924-1868. doi: 10.1007/s11257-017-9188-z. URL https://doi.org/10.1007/s11257-017-9188-z.

Kerly, A. *Negotiated Learner Modelling with a Conversational Agent*. PhD thesis, University of Birmingham, 2009.

Koedinger, K R. and Aleven, Vincent. An interview reflection on "intelligent tutoring goes to school in the big city". *International Journal of Artificial Intelligence in Education*, 26(1):13–24, Mar 2016. ISSN 1560-4306. doi: 10.1007/s40593-015-0082-8. URL https://doi.org/10.1007/s40593-015-0082-8.

Koedinger, K. R. and Corbett, A. T. Cognitive tutors: Technology bringing learning science to the classroom. In *The Cambridge Handbook of the Learning Sciences*, pp. 61–78. Cambridge University Press, Orlando, Florida, k. sawyer edition, 2006.

Kostakos, V. and Musolesi, M. Avoiding pitfalls when using machine learning in hci studies. *interactions*, 24 (4):34–37, June 2017. ISSN 1072-5520. doi: 10.1145/3085556. URL http://doi.acm.org/10.1145/3085556.

Long, Y. and Aleven, V. Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Model. User-Adapt. Interact.*, 27(1):55–88, 2017. doi: 10.1007/s11257-016-9186-6. URL https://doi.org/10.1007/s11257-016-9186-6.

Ma, W., Adesope, O. O, Nesbit, J C., and Liu, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, Volume 106(issue 4):901–918, November 2014. doi: http://dx.doi.org/10.1037/a0037123. URL http://dx.doi.org/10.1037/a0037123.

Mabbott, A. and Bull, S. Student preferences for editing, persuading, and negotiating the open learner model. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, ITS'06, pp. 481–490, Berlin,

Heidelberg, 2006. Springer-Verlag. ISBN 3-540-35159-0, 978-3-540-35159-7. doi: 10.1007/11774303_48. URL http://dx.doi.org/10.1007/11774303_48.

Mavrikis, M. Modelling student interactions in intelligent learning environments: constructing bayesian networks from data. *International Journal on Artificial Intelligence Tools*, 19(06):733–753, December 2010. ISSN 0218-2130. doi: 10.1142/S0218213010000406. URL https://www.worldscientific.com/doi/abs/10.1142/S0218213010000406.

Mitrovic, A., Martin, B., and Suraweera, P. Intelligent tutors for all: The constraint-based approach. *IEEE Intelligent Systems*, 22(4):38–45, July 2007. ISSN 1541-1672. doi: 10.1109/MIS.2007.74.

Monkaresi, H., Bosch, N., Calvo, R., and D'Mello, S. Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Trans. Affective Computing*, 8(1):15–28, 2016.

Nesbit, J. C., Adesope, O. O., Liu, Q., and Ma, W. How effective are intelligent tutoring systems in computer science education? In *2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT)*, volume 00, pp. 99–103, July 2014. doi: 10.1109/ICALT.2014.38. URL doi.ieeecomputersociety.org/10.1109/ICALT.2014.38.

Porayska-Pomsta, K. and Chryssafidou, E. Adolescents' Self-regulation during Job Interviews through an AI Coaching Environment. In *19th International Conference on Artificial Intelligence in Education*, volume 00, 2018.

Porayska-Pomsta, K., Anderson, K., Bernardini, S., Guldberg, K., Smith, T., Kossivaki, L., Hodgins, S., and Lowe, I. Building an intelligent, authorable serious game for autistic children and their carers. In Reidsma, Dennis, Katayose, Haruhiro, and Nijholt, Anton (eds.), *Advances in Computer Entertainment*, pp. 456–475, Cham, 2013. Springer International Publishing.

Porayska-Pomsta, K., Rizzo, P., Damian, I., Baur, T., André, E., Sabouret, N., Jones, H., Anderson, K., and Chryssafidou, E. Who's Afraid of Job Interviews? Definitely a Question for User Modelling. In Dimitrova, Vania, Kuflik, Tsvi, Chin, David, Ricci, Francesco, Dolog, Peter, and Houben, Geert-Jan (eds.), *User Modeling, Adaptation, and Personalization*, pp. 411–422, Cham, 2014. Springer International Publishing. ISBN 978-3-319-08786-3.

Schroeder, N L., Adesope, O. O., and Barouch Gilbert, R. How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review. *Journal of Educational Computing Research*, 49(1):1–39, 2013. doi: 10.2190/EC.49.1.a. URL https://doi.org/10.2190/EC.49.1.a.

Stamper, J C., Eagle, M., Barnes, T., and Croy, M. Experimental evaluation of automatic hint generation for a logic tutor. In *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, AIED'11, pp. 345–352, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-21868-2. URL http://dl.acm.org/citation.cfm?id=2026506.2026553.

Weller, A. Challenges for transparency. In *Proceedings of the ICML Workshop on Human Interpretability in Machine Learning*, pp. 55–62, 2017.

Woolf, B. *Building Intelligent Interactive Tutors*. Morgan Kaufmann, 2009. doi: 10.1016/B978-0-12-373594-2.00016-2.