Explainable AI: Transparent Pedagogical Agents that help the Learner to Reflect

Deborah Richards, PhD
Macquarie University, Australia, deborah.richards@mq.edu.edu

I. Introduction

Pedagogical Agents (PAs) emerged from the research field of Intelligent Tutoring Systems as an AI-based approach to deliver support and learning benefits in roles such as peer tutors and peer learners [1]. However, like much of AI, they are not pervasive in today's classrooms. Reasons for their absence are many and complex, ranging from limited applications and insufficient awareness of the specific learner to resistance from teachers to try new technology that can be based on unfamiliarity or even fear of being replaced. This article looks at how adding transparency to PAs can help learners and teachers evaluate the agent's behaviours, decide whether to accept the support offered and, perhaps more importantly, reflect on, and potentially revise, their own understanding.

II. Relevant Theories of Learning

A. Social-Cognition and Socially Constructed Learning

Theories from cognitive science, educational psychology and the learning sciences have driven PA research since its beginning. Pioneers in PA research, Kim, et al. [2] used a social-cognitive lens to design their PAs by looking at the impact of learners' affective experience and self-efficacy on their cognitive engagement and deep learning [3-5].

The social nature of PAs, typically with humanlike appearances or behaviours, recognized that learning is a social activity [6] and supports a social constructivist view of learning, where learning is both an individual and social process involving artefacts and other people [7]. PAs provide a conversational humanlike way of delivering information that overcomes literacy barriers [8].

B. Active Learning and Interactive Pedagogical Drama

The ability to simulate social interactions makes PAs particularly suitable for providing social support and training in social skills. The use of scenarios, storytelling and dramas involving PAs allows the student to explore social situations and consider alternative solutions to handle the dilemmas presented. As described, "The goal of Interactive Pedagogical Drama (IPD) is to exploit the edifying power of story while promoting active learning. An IPD immerses the learner in an engaging, evocative story where she interacts with realistic characters. The learner makes decisions or takes actions on behalf of a character in the story, and sees the consequences of her decisions" [9] p. 1. IPD was used in the FearNot! Project to allow students to explore alternative strategies to cope with bullying scenarios [10] and extended in the Orient Project [11] to provide scenarios to evoke empathy for migrant/refugee children amongst school children in the UK and Germany.

C. Media Equation: Humanlike social responses to machines

Kim and Baylor [12] attribute the social, emotional and often unexpected responses by learners to PAs to the human tendency to treat media socially and anthropomorphize objects. This tendency was recognized in early studies by psychologists who observed that humans treat machines with similar politeness rules that they use with other humans [13].

III. Enabling Technological Advances

A. Technologies behind PAs

PAs utilise a range of artificial intelligence technologies including intelligent tutoring systems, cognitive agent architectures that drive the reasoning of the agent, natural language processing and affective technologies involving multimodal inputs to persuade learners to engage in computer-based training and to detect epistemic emotions, the emotions evoked when learning, such as engaged, bored, frustrated and confused [14]. Game engine technology is used to develop avatars and build the virtual worlds that the PAs inhabit and gaming elements are used to improve motivation.

IV. Evidence of Potential Impacts

A number of articles providing the state of the art of Pedagogical Agents in education have been published in recent years (see, Johnson and Lester [15], Kim and Baylor [12], Schroeder, et al. [16] and Sottilare and Hart [17]). The meta-analysis undertaken by Schroeder, et al. [16] included 43 studies covering a range of learner populations (i.e. K-12 students, university students, workplace trainees); domains (e.g. business, defence and medical education) and PA representations (i.e. stick figures to full-bodied virtual humans). Overall, the studies found statistically significant findings in support of interventions involving PAs with variations depending on subject domain, age, gender and ethnicity of the learners.

Nevertheless, despite demonstrated improved learning outcomes, their presence in classrooms is limited and often restricted to one-off research projects. Perhaps it can be argued that the technology is not mature enough for general and common use. However, I contend that current PA capabilities already have the potential to advance student learning if they are recognised as a 1) tool for reflection rather than an oracle of knowledge and 2) learning aid rather than a replacement for teacher or student endeavour and persistence. Transparent PAs that are able to explain their behaviour, including their suggestions and questions, allow the learner or teacher to take control of their learning and decide when the PA is useful for their learning and how they will utilise them.

Since the 80s, the use of explanation in AI, particularly in Expert Systems, has been a strategy to gain user acceptance, including adoption by domain experts. In Expert Systems, also known as Knowledge Based Systems, the user had the ability to ask "why" [did you ask me that?] and "how" [did you come to that conclusion]. Decades on, we again see growing interest in explainable AI (where the AI explains itself) in response to current underutilization of AI technology [18], despite the powerful insights machine learning promises. This underutilization hinges on the lack of transparency in current neutral network-based algorithms that produce classifications but are unable to explain how those outputs were reached. PAs driven by such methods also suffer from inability to explain themselves. However, the majority of PAs are logic or rule-based and thus the user is able to inspect the agent's reasoning processes. Through inspection and explanation, the student and/or teacher is able to decide whether to follow the advice offered. But, more importantly transparent reasoning allows the learner to reflect on their own knowledge as they seek to make sense of the reasoning provided by the PA and evaluate it against their own understanding. This process becomes a valuable opportunity to reinforce concepts or to challenge current (mis)conceptions.

Reflection is an essential skill for teachers and learners. However, it is a skill that requires motivation and takes time and practice to master. We're all familiar with either giving or receiving the response "nothing" to the general question "what did you learn at school today?" However, as with so many questions we ask, you need to know the answer to ask the right question. Asking the right question and asking it the right way can address the problem of motivation. This is where the value of a personalized and context-aware PA can be realized.

The learning approach afforded by PAs, seeks to encourage reflective learning and the development of tacit, rather than declarative or procedural knowledge. In line with this approach, PAs have been used to significantly improve episodic memory by using a reminiscing agent to debrief the student following exploration of a virtual world [19]. Similar to the use of virtual agents to assist the elderly to reminisce [20], PAs could help students with remembering and act as a form of external memory aid that does more than store or retrieve information. Another direction to encourage reflection is the use of teachable agents where the human (co)-learner teaches the agent what to do or provides the agent with answers to its questions [21].

V. SUMMARY

PAs can play many different roles in the context of learning. Some of these roles will be in classroom and formal learning contexts, but many will be in situations in the workplace, home or community acting as personalised companions and coaches to life-long learners. To achieve this vision, advances are needed in developing agents that exhibit empathy and personality, improved user state recognition, affective user modeling and personalization and natural language processing. Increasing transparency of the PAs reasoning and educating humans to seek explanations and reflect on those are essential for ensuring appropriate social and ethical uses of PA in our society.

REFERENCES

[1] W. L. Johnson, J. W. Rickel, and J. C. Lester, "Animated pedagogical agents: Face-to-face interaction in interactive learning environments," *International Journal of Artificial intelligence in education,* vol. 11, pp. 47-78, 2000.

[2] Y. Kim, A. L. Baylor, and P. Group, "Pedagogical agents as learning companions: The role of agent competency and type of interaction," *Educational Technology Research and Development,* vol. 54, pp. 223-243, 2006.

[3] A. Bandura, "Social foundations of thought and action," *Englewood Cliffs, NJ,* vol. 1986, 1986.

[4] R. W. Picard, *Affective computing*. Cambridge, Massachsusetts: MIT Press, 1997.

[5] S. Spence, "Descartes' error: Emotion, reason and the human brain," *BMJ,* vol. 310, p. 1213, 1995.

[6] L. S. Vygotsky, *Mind in society: The development of higher psychological processes*: Harvard university press, 1980.

[7] J. G. Greeno, A. M. Collins, and L. B. Resnick, "Cognition and learning," *Handbook of educational psychology,* vol. 77, pp. 15-46, 1996.

[8] T. W. Bickmore, L. M. Pfeifer, D. Byron, S. Forsythe, L. E. Henault, B. W. Jack*, et al.*, "Usability of conversational agents by patients with inadequate health literacy: Evidence from two clinical trials," *Journal of health communication,* vol. 15, pp. 197-210, 2010.

[9] S. Marsella, W. L. Johnson, and C. LaBore, "Interactive pedagogical drama for health interventions," in *11th International Conference on Artificial Intelligence in Education, Sydney, Australia*, 2003, pp. 341-348.

[10] J. Dias, S. Mascarenhas, and A. Paiva, "Fatima modular: Towards an agent architecture with a generic appraisal framework," in *Emotion modeling*, ed: Springer, 2014, pp. 44-56.

[11] R. Aylett, N. Vannini, E. Andre, A. Paiva, S. Enz, and L. Hall, "But that was in another country: agents and intercultural empathy," in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 2009, pp. 329-336.

[12] Y. Kim and A. L. Baylor, "based design of pedagogical agent roles: A review, progress, and recommendations," *International Journal of Artificial Intelligence in Education,* vol. 26, pp. 160-169, 2016.

[13] B. Reeves and C. I. Nass, *The media equation: How people treat computers, television, and new media like real people and places*: Cambridge university press, 1996.

[14] O. M. Nezami, D. Richards, and L. Hamey, "Semi-Supervised Detection of Student Engagement," 2017.

[15] W. L. Johnson and J. C. Lester, "Pedagogical Agents: Back to the Future," *AI Magazine,* vol. 39, 2018.

[16] N. L. Schroeder, O. O. Adesope, and R. B. Gilbert, "How effective are pedagogical agents for learning? A meta-analytic review," *Journal of Educational Computing Research,* vol. 49, pp. 1-39, 2013.

[17] R. Sottilare and J. Hart, "Cognitive and affective modeling in intelligent virtual humans for training and tutoring applications," *Advances in Applied Human Modeling and Simulation,* p. 113, 2012.

[18] T. Miller, "Explanation in artificial intelligence: insights from the social sciences," *arXiv preprint arXiv:1706.07269,* 2017.

[19] M. Nicholas, P. Van Bergen, and D. Richards, "Enhancing learning in a virtual world using highly elaborative reminiscing as a reflective tool," *Learning and Instruction,* vol. 36, pp. 66-75, 2015.

[20] S. Nikitina, S. Callaioli, and M. Baez, "Smart conversational agents for reminiscence," *arXiv preprint arXiv:1804.06550,* 2018.

[21] L. Pareto, "A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning," *International Journal of Artificial Intelligence in Education,* vol. 24, pp. 251-283, 2014.