When Lying, Hiding and Deceiving Promotes Learning - A Case for Augmented Intelligence with Augmented Ethics

Björn Sjödén(✉)

Halmstad University, Halmstad, Sweden
bjorn.sjoden@hh.se

Abstract. If AI systems are to be used for truly human decision-making in education, teachers will need better support for deciding upon educational interventions and strategies on an ethically informed basis. As indicated by a recent call by the AIED Society to focus on the FATE (Fairness, Accountability, Transparency, and Ethics) of AI in education, fundamental issues in this area remain to be explicated, and teachers' perspectives need to be accounted for. The paper offers examples of how AI may serve to promote learning but at the cost of presenting limited or untruthful information to the student. For example, false information about a student's current progress may motivate students to finish a task they would otherwise give up; hiding information from the student that is disclosed to the teacher may decrease students' cognitive load while supporting the teacher's strategic choices, and deceiving the student as to the actual nature of the task or interaction, such as when using virtual agents, can increase students' efforts towards learning. Potential conflicts between such scenarios and basic values of FATE are discussed, and the basis for developing an "augmented ethics" system to support teachers' decision-making is presented.

Keywords: Ethics · Teacher perspectives · FATE · Augmented intelligence · Augmented ethics

1 Introduction

The importance of ethical issues in AIED community motivated a recent call to focus on the FATE (Fairness, Accountability, Transparency, and Ethics) of AI in education. Although FATE makes a nice acronym, it blurs the conceptual relations between these topics (e.g. fairness can be seen as one of several ethical concerns, and accountability as a concept which guides ethical considerations). To guide research and practice, it needs to be situated both in an ethic-theoretical context and in empirical research, and to take into account the perspective of the practitioners – the teachers. Teachers' knowledge of AI and related ethical issues in school needs to increase, and the literature has not clearly addressed the role of the teacher [1, 2]. If AI is to empower

I. I. Bittencourt et al. (Eds.): AIED 2020, LNAI 12164, pp. 291–295, 2020.
https://doi.org/10.1007/978-3-030-52240-7_53

education by augmenting human capabilities, how can ethical standards of human decision-making be ensured? What makes for an ethically informed basis?

This paper aims to address the ethical foundation that can guide empirical research on the teacher's practical knowledge needs, when using presently available AI such as adaptive systems, virtual agents and learning analytics. It argues that the constituents of an augmented ethics system require a broader analysis than that of augmented intelligence in the traditional sense. For instance, there are national curricula, treaties and policy documents, such as the General Data Protection Regulation (GDPR) in Europe and the UN Convention on the Rights of the Child, which must not be neglected to provide useful support for teachers. Hence, ethical theory, teaching practice and policies all need to inform the development of a system that effectively "augments" ethics.

2 Lying, Hiding and Deceiving for the FATE of Learning

There are many opportunities for using AI to enhance student learning at the cost of presenting untruthful, partial or misleading information to the student – in other words, systems that lie, hide or deceive. The message is not to condemn the existence or use of such functions – in fact, teachers have always used deliberate (over-)simplifications and factual misrepresentations in order to help students learn, and so has been done since the early days of AIED [e.g. 3] – but as AIED grows in complexity, and becomes more pervasive in the absence of human reflection and judgment, we need refined conceptual tools to identify and assess potential ethical conflicts with basic human values.

To what extent teachers need support, and of what kind, for taking a position to ethical dilemmas raised by recent AIED, remains an outstanding question. Some cases may appear unethical, such as deliberately inducing confusion in students by staging disagreements between agents [4] or presenting students with erroneous examples [5], but become less problematic for mature learners who are "game", become aware of the manipulations and submit to the pedagogic strategy. Then there are systems which may have personal repercussions far beyond what students and/or their teachers may recognize. A prevalent concern is privacy, relating to learning analytics (LA) [6, 7], for example whether the overall improvement of a learning environment is a valid reason to store and share the exact location of students to facilitate collaboration with peers. Other examples concern the use of Intelligent Tutoring Systems (ITS) that match students with virtual tutors on emotional and cognitive parameters. This raises issues as to when students' interactions with non-human systems are preferred to a human being. As noted in one study, "What is true if the teacher and AI do not agree?" [1].

The message then, in line with other recent work [8], is that ethical use of AI in schools require that teachers' unique human expertise is preserved and promoted. Such expertise is needed for deciding when it is warranted to use misrepresentations or a "deceptive" system for a larger good, in order to secure educational benefits and avoid risks for students' well-being. Next are some examples of such potential conflicts.

Lying refers to deliberately presenting information to the student that is incorrect, with reference to the available data. In principle, this concerns all cases where students are presented with incorrect information and requested to correct it, although the AI “knows” the correct answer. But there are more subtle and specific examples. Studies on learning curves and motivation suggest that students work longer in a problem domain if they make visible progress and are closer to goal (say, 80%) compared to not progressing, further from the goal (say, 40%). Such data can be used for algorithms that – truthfully – match the difficulty of learning tasks to the student’s current performance level in a “personally” adaptive system (e.g. Sana Labs, www.sanalabs.com). Would it then be ethically justifiable to present false information about a student’s current progress, suggesting that one is closer to the goal than performance indicates, in order to motivate students to finish a task they would otherwise give up?

Hiding refers to presenting selective, but not untrue, information to the student, while processing more data that is relevant to the task but may be presented at a later time and/or to another person (a peer or a teacher). AI systems that serve to identify what data are important to students are implemented in Learning Analytics (LA) and motivate the separation between *student-facing* and *teacher-facing* LA [9]. Hiding information from the student that is disclosed to the teacher may decrease students’ cognitive load while supporting the teacher’s strategic choices. Should AI therefore be used to determine what data are ‘better’ communicated to teachers and students, respectively?

Deceiving refers to presenting the student with tasks that are designed to maintain false beliefs or illusions, without making the actual nature of the task or interaction explicit. A form of voluntary deception occurs in all (educational or other) games which involve an “intelligent” opponent that is technically invincible but adapts to the player’s performance. The same can be said about collaborative virtual agents, such as Teachable Agents that increase students’ efforts [10–12]. An interesting example is BELLA [13] which employs a “super-agent” to adapt to students’ knowledge gaps without actual “teaching” by the student. In research, Wizard-of-Oz methodologies exploit student expectations for improving upon existing systems by having human actors simulate AI agents. To what extent are such illusions ethically justifiable to maintain?

3 Towards a System for Augmented Ethics

The wide variety of issues and ethical concerns makes it difficult to define which aspects of FATE to focus on. From consulting ethical-philosophical expertise and standard works [14] four basic values are identified: *privacy*, *safety*, *trust* and *fairness*. These values are fundamental in the sense that there is no obvious way of telling which value trumps another one. As to the FATE dimensions, one can argue that, for instance, “Transparency” is not a fundamental ethical value because it could, at the same time, be a risk and a benefit to safety, and a risk to privacy. “Fairness”, on the other hand, is a fundamental social value (one cannot be “fair” in isolation), theoretically independent of individual privacy and safety.

Addressing the multiplicity of concerns is helped by distinguishing between pedagogies on the screen-level, “how individual systems work with a single student”, and the orchestration-level, “whereby such systems are deployed in the bigger temporal and spatial context of a whole class” [15, p. 6]. The screen/orchestration level distinction thus helps both to direct teachers' attention and to see how accountability is attributed. The results of discussions with teachers can inform ethical guidelines that support decision-making as to what values should be protected, to what costs and benefits.

Figure 1 offers a simplified categorization grid of ethical concerns that emerge from relating teachers' knowledge needs on the screen- and orchestration levels. It is suggested that the teacher take a stand on two questions: *Is the concern a screen-level priority? Is the concern an orchestration-level priority?* It should be emphasized that the yes (✓) or no (✗) to these questions is a deliberate simplification; they are a question of focus rather than exclusion, and they do not definitely tell where concerns belong.

		Screen-level priority	
		✓	✗
Orchestration-level priority	✓	**Privacy** e.g. *Shall teachers be able to see when and for how long a student did her homework?* *Shall students and their parents be allowed to see how much time the teacher spent correcting the students' homework?*	**Safety** e.g. *How does the system support the student while safeguarding students' independence?* *How does the system as a substitute for human interactions affect children's formation of self and personal identity?*
	✗	**Trust** e.g. *Is it possible that students will trust an AI-based system so much as to prefer its company to their teacher and their peers?* *Are AI-based systems more trustworthy than teachers in certain areas of knowledge, e.g. math and computer science?*	**Fairness** e.g. *How should AI resources be distributed to student groups on an equal basis?* *How do we share teacher and teaching resources fairly among students?*

Fig. 1. A grid for determining types of ethical priorities for AIED, with example questions.

On the screen-level, concerns of *privacy* can be viewed with respect to privacy settings available to the individual but also what data the system stores and what personal data is requested at start. The individual's *trust* in the system is dependent on how well it functions, both for protecting personal data and for producing the expected outcomes. The teacher can assist students with available privacy settings and data storage but cannot directly influence students' trust and expectations, which can only develop from personal experience of working with a system in relation to (human or AI) alternatives.

On the orchestration-level, *safety* can be viewed with respect to how the teacher assesses and manages the risks and threats for all students in a class (arguably, students may have different preferences of privacy and what information they are willing to share, but they should all have an equal level of safety). Safety concerns are about the whole group and the orchestration of all systems used in the classroom. *Fairness* is a value of broader ethical concern than can be addressed by either the student on screen or the teacher beyond her own classroom. Issues of fairness must not be ignored, but teachers need to be aware of the complex social, financial, and cultural context in which they are embedded. For example, teachers and policy makers may need to consider gender equality, and whether the use of AI should be mandatory.

In conclusion, this organization of ethical priorities put the theoretical corner stones on which to base ethical positions with respect to the teacher's responsibilities, the system properties and contextual knowledge needs. Each of the four values deserves attention in its own right. For understanding their meaning in practice and further development, it is suggested that teachers are involved at an early stage and work together with researchers, such as in workshops, in an iterative process of identifying, analyzing, evaluating and re-evaluating ethical concerns. Such a project would have great significance both on a societal level and for covering knowledge gaps on the ethics of AIED.

References

1. Hrastinski, S., et al.: Critical imaginaries and reflections on artificial intelligence and robots in postdigital K-12 education. Postdigital Sci. Educ. **1**(2), 427–445 (2019)
2. Humble, N., Mozelius, P.: Teacher-supported AI or AI-supported teachers?. In: European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2019), pp. 157–164. Academic Conferences and Publishing International Limited, Oxford (2019)
3. Gutwin, C., McCalla, G.: The use of pedagogic misrepresentation in tutorial dialogue. In: Frasson, C., Gauthier, G., McCalla, G.I. (eds.) ITS 1992. LNCS, vol. 608, pp. 507–514. Springer, Heidelberg (1992). https://doi.org/10.1007/3-540-55606-0_60
4. Lehman, B., et al.: Inducing and tracking confusion with contradictions during complex learning. Int. J. Artif. Intell. Educ. **22**(1–2), 85–105 (2013)
5. Adams, D.M., et al.: Using erroneous examples to improve mathematics learning with a web-based tutoring system. Comput. Hum. Behav. **36**, 401–411 (2014)
6. Pardo, A., Siemens, G.: Ethical and privacy principles for learning analytics. Br. J. Educ. Technol. **45**(3), 438–450 (2014)
7. Slade, S., Prinsloo, P.: Learning analytics: ethical issues and dilemmas. Am. Behav. Sci. **57** (10), 1509–1528 (2013)
8. Felix, C.: The role of the teacher and AI in education. In: Blessinger, P., Sengupta, E. (eds.) International Perspectives on the Role of Technology in Humanizing Higher Education. Emerald Group Publishing, Bingley (2020)
9. Bodily, R., Verbert, K.: Trends and issues in student-facing learning analytics reporting systems research. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 309–318. ACM, New York (2017)
10. Chase, C.C., Chin, D.B., Oppezzo, M.A., Schwartz, D.L.: Teachable agents and the protégé effect: increasing the effort towards learning. J. Sci. Educ. Technol. **18**(4), 334–352 (2009)
11. Matsuda, N., et al.: Cognitive anatomy of tutor learning: lessons learned with SimStudent. J. Educ. Psychol. **105**(4), 1152–1163 (2013)
12. Pareto, L.: A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. Int. J. Artif. Intell. Educ. **24**(3), 251–283 (2014). https://doi.org/10.1007/s40593-014-0018-8
13. Lenat, D.B., Durlach, P.J.: Reinforcing math knowledge by immersing students in a simulated learning-by-teaching experience. Int. J. Artif. Intell. Educ. **24**(3), 216–250 (2014)
14. Rawls, J.: A Theory of Justice. Belknap Press, Cambridge (1972)
15. du Boulay, B.: Escape from the skinner box: the case for contemporary intelligent learning environments. Br. J. Educ. Technol. **50**(6), 2902–2919 (2019)