Privacy considerations of using social robots in education: Policy recommendations for learning environments

Dr. Samantha-Kaye Johnston

Department of Computer Science, The University of Oxford, United Kingdom

Abstract

In learning environments, there is widespread adoption of technologies that depend on artificial intelligence. Social robots – the focus of this brief – provide an interesting example because they have both an indirect (e.g., online software and direct (e.g., physicalised devices) presence in classrooms. Alongside their implementation are well-justified concerns about access, well-being, AI literacy and student privacy. The aim of this brief is to present the findings from a series of roundtables on the topic of social robots in classrooms with 105 educators, primary-school students, human-centred researchers, policymakers, and education technology industry representatives across four regions: the US, the Caribbean, Africa and Australia. Given that privacy was identified as the most thematically important concern around social robots usage within learning environments, it is the focus of this brief, with a deep dive into Lumilo (a potential feature of social robots). As this brief presents a consensus view of industry and public sectors across diverse learning contexts, it offers a missing piece of the puzzle by integrating these voices from the Global North and Global South, suitable for an international research and policy audience.

The age of AI: Social robots in education

Artificial Intelligence (AI) has been incorporated across several areas of education (Miao et al., 2021). One example is the use of educational social robots (chatbots), which function as a tutor for students and teachers in their learning and teaching journey, respectively (Belpaeme et al., 2018; Smakman et al., 2021). In a typical learning scenario, users input textual or audio requests. Then, cloud-based services and AI techniques allow chatbots to deploy natural language processing and machine learning to provide appropriate responses and action various tasks (du Boulay et al., 2018; Gena et al., 2020; Verbert et al., 2013). Several examples of social robots exist, including Ada[1] and Deakin Genie[2], which are underpinned by intelligent tutoring systems [ITSs] (Burns & Capps, 2013).

Despite their value in learning environments (Alemi et al., 2017; Belpaeme et al., 2018; Johnson & Lester, 2016; Jones et al., 2014), social robots pose several risks (see Mechelen et al., 2020 for an 18-year ethical overview in child-computer interaction research). Key to this is the large, representative dataset requirement for social robots, which relies on continuous data collection – when, how and what to collect from end users is not always clear. Moreover, psychological concerns abound (Lutz et al., 2019) in relation to social robots replacing humans and their role in socially isolating students (Kennedy et al., 2016). Together, these instances represent risks to privacy (Serholt & Barendregt, 2014), safety (Serholt et al., 2017) and well-being (Kennedy et al., 2016). Moreover, managing these risks are often situated in environments where technology develops more rapidly than their governance frameworks. Although, it is worth mentioning that emerging policy discussions such the Online Safety Bill in the UK (Trengrove et al., 2022; cf Show, 2023) and, in more Global South contexts, the Jamaica Data Protection Act (Jamaica DPA, 2020), offer promising avenues for how to better manage risks related to current and emerging technologies. In the current study, social chatbot usage is qualitatively examined through the lens of an overarching question: *What are the affordances and challenges of social robots within learning environments?*

Methodology

Participants. 105 participants (educators, human-centred researchers, primary-school students, policymakers and education technology representatives) across the US (n = 22), Australia (n = 25), Africa (n = 24) and the Caribbean (n = 34) agreed to participate. Consent/assent was individually gained from each stakeholder. Participants were purposively selected not for their experience with social robots, but rather to ascertain a more diverse understanding of perceived impacts across learning environments (Friedman et al., 2017; Miller et al., 2007; Winkler & Spiekermann, 2021).

[1] https://www.jisc.ac.uk/news/chatbot-talks-up-a-storm-for-bolton-college-26-mar-2019

[2] https://futured.deakin.edu.au/concepts/internet-of-things/genie-smart-speaker-integration/

Procedure and Analysis. Roundtable sessions were held online; in the children's session, an adult (parent or teacher) was present. Given participants' diverse exposure to social robots, explanations about this technology were provided using real-life case studies. Participants also watched a video with students engaging with a NAO[3] in a learning environment: this familiarisation approach is commonplace in child-robot interaction research (Ahmad et al., 2016; Belpaeme, 2020; Rosanda & Istenič Starčič, 2019).

All recordings were transcribed and inductively coded and analysed following Braun and Clarke's (2021) framework. Two researchers iteratively identified codes, themes and sub-themes from the data. Given its thematic importance, privacy and its sub-themes[4] will be discussed in the remainder of this article. Moreover, the example of Lumilo emerged quite often in the data. Consequently, this potential feature of social robots is used to contextualise the discussion, before offering policy recommendations.

Privacy challenges of social robots: The case of Lumilo

Lumilo is "a pair of mixed-reality smart glasses designed to support K-12 teachers in orchestrating personalized class sessions" (Holstein et al., 2018, p.80).[5] For example, during a problem-solving task, Lumilo's features permits students to communicate with the software to request hints. Its developers provide evidence of Lumilo (1) helping teachers to better allocate time to students with lower prior knowledge and (2) predicting which students would benefit from the combined help from Lumilo and their teacher (Holstein et al., 2018).

Despite these promising findings, given Lumilo's intended purpose, there are multiple instances where data is gathered, processed and stored. That is, once a teacher engages with Lumilo, it provides live feedback of student work patterns (via floating text appearing above students' head), it shows student (include deep dive screens) and class-level analytics, and it alerts teachers of students' current states (e.g., emotions). Therefore, Lumilo requires several functions: It must comprise the ability to store data on student performance, states, and teachers' actions. Together, this suggests that, on any given occasion, Lumilo collects multiple data representations (e.g., physical, audio and visual).

Lumilo's data streams are the bedrock of enabling teachers to provide personalised tutelage. For example, these streams enable educators to access granular insights about students' strengths and areas for improvement. But, inherent in this scenario is the need to illuminate AI's black box (Rai, 2020). Interestingly, participants viewed issues around social robots' black box, which is closely linked to **AI explainability (XAI)**, to be a privacy-related challenge since: "*Not knowing how [student] outcomes are derived is problematic. What data is being collected, or not collected, to arrive at these conclusions about students and our teaching practice? It's still a matter of privacy, isn't it?*" (Teacher participant). Recent shifts in the XAI field have seen more socio-technical approaches to explain AI decision making processes, where explanations are both product and process-oriented and adopt a human-centred perspective (Ehsan et al., 2021; Miller, 2019; Srinivasan & Chander, 2020). Key to this is the reduction of explanations that conceal process-oriented information from end-users (Solove, 2002). For example, the recently proposed XAI in education (XAI-ED) framework adopts principles of fairness, accountability, transparency and ethics, all aimed at increasing trust in AI among educational stakeholders (Khosravi et al., 2022). Moreover, emerging XAI notions, involving novels ways of conceptualising explainability, beyond heatmaps and neural network classifiers and exploring explanation quality, have been proposed (Holzinger et al., 2022).

In the current and previous studies, much of what entices educators about social robots is task automation, alongside the agency to activate or deactivate specific features to achieve a desired learning outcome (Holstein et al., 2017; 2018). But it was recognised that there are little privacy design considerations for **student autonomy**. In Lumilo, teachers can opt-out of sharing class comparison details with students and override the system if necessary. Moreover, teachers know their Lumilo interactions are deposited into DataShop, a large educational repository (Holstein et al., 2018; Koedinger et al., 2017). However, there is virtually surface-level conversations about student autonomy over their data. Therefore, despite the importance of collecting multiple data points to understand student performance, this raises issues around agency and control (or a lack thereof) and the ability for students to prevent access to personal cognitive states. Two components of Solove's (2002) privacy framework are applicable in this scenario: (1)

[3] Nao is a small humanoid robot designed to interact with people.

[4] Privacy sub-themes are bolded.

[5] Lumilo is designed using Unity3D for the Microsoft (2017) HoloLens and has the ability to be integrated across a variety of ITSs.

the right to be left alone and (2) the right to shield oneself from access. In the context of social robots, although these two components suggest that consent must be obtained for lawful collection of student data, there remains "unknowns" in relation to when their data is collected, the type of data collected and the extent of the functionalities of social robots (Ozmen Garibay et al., 2023).

This emerging requirement (need for explainability) and unintended consequence (insufficient student autonomy) raises the question around the **balance between privacy, collection and autonomy.** At least two technological solutions offer a starting point to strike this balance. Firstly, advances in multi-modal learning analytics could limit AI's dependence on large, unregulated data flows within learning software, especially if paired with advances in modelling student data (Desmarais & Baker, 2012). Secondly, human-centred solutions, such as SOLID (Sullivan, 2022), aim to ensure that data ownership remains in the hands of users. However, both solutions require a more refined understanding of user needs for autonomy, given the previously observed privacy paradox: users claim they want privacy but act in opposite ways when "protecting" personal details (Kokolakis, 2017).

Policy recommendations

The emergence of new education technologies requires a new social contract (UNESCO, 2022) that (1) prioritises user autonomy, (2) elevates privacy by design, (3) is subject to contextualised privacy-enhancing techniques, (4) strengthens AI literacy, (5) is subject to algorithmic impact assessments, and (6) ensures governance across the entire cycle of data collection, processing and storage. More specific recommendations are offered below:

- Consider introducing configurable, context-specific governance models that address when, how and what data is collected and under what circumstances. These structures should also explicitly address what information requires consent;
- Elevate student voices through long-term co-design sessions or study-beds (see Cortesi et al., 2021 for models of youth participation model);
- Collaborate with education assessment specialists to develop psychometrically sound algorithmic impact assessments, specifically for learning environments;
- School curricula should prioritise privacy literacy: Students should be taught to critically evaluate data processing approaches, including data storage approaches (centrally or de-centrally), who can alter their personal data and in what context. Educators and students should ask questions about how privacy by design is considered in social robot development.

Beyond these recommendations, there is scope to provide more robust evidence around the longstanding impact of privacy intrusions on educational stakeholders, especially young children. More importantly, although privacy is the starting question (and most thematically important in the current article), data collection, processing and storage processes, ultimately rests of the nature of security protocols and the governance structures that underpin them.

Acknowledgments

Gratitude is extended to all the participants in the roundtable sessions. One Research Associate assisted with coding and agreement and three additional research assistants who engaged in reviewing this process. This was aimed at the chance of reducing bias in coding and interpretation.

References

Ahmad, M. I., Mubin, O., & Orlando, J. (2016). Understanding behaviours and roles for social and adaptive robots in education: Teacher's perspective. *Proceedings of the Fourth International Conference on Human Agent Interaction*, 297–304. https://doi.org/10.1145/2974804.2974829

Alemi, M., Meghdari, A., Basiri, N. M., & Taheri, A. (2015). The effect of applying humanoid robots as teacher assistants to help Iranian autistic pupils learn English as a foreign language. In A. Tapus, E. André, J.-C. Martin, F. Ferland, & M. Ammi (Eds.), *Social robotics* (Vol. 9388, pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-25554-5_1

Belpaeme, T. (2020). Advice to new human-robot interaction researchers. In C. Jost, B. Le P´ev´edic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge, & N. Mirnig (Eds.), *Human-robot interaction: Evaluation methods and their standardization* (pp. 355–369). Springer International Publishing. https://doi.org/10.1007/978-3-030-42307-0_14

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21). https://doi.org/10.1126/scirobotics.aat5954

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, *21*(1), 37-47. https://doi.org/10.1002/capr.12360

Burns, H. L., & Capps, C. G. (2013). of Intelligent Tutoring Systems: An Introduction. *Foundations of Intelligent Tutoring Systems*, 1-19.

Cortesi, S., Hasse, A., & Gasser, U. (2021). Youth participation in a digital world: designing and implementing spaces, programs, and methodologies. *Berkman Center Research Publication*, (2021-5).

Desmarais, M. C., & Baker, R. S. D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, *22*, 9-38. https://doi.org/10.1007/s11257-011-9106-8

du Boulay, B., Poulovassilis, A., Holmes, W., & Mavrikis, M. (2018). What does the research say about how Artificial Intelligence and Big Data can close the achievement gap. *Enhancing learning and teaching with technology*, 316-327.

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445188

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends in Human-Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Gena, C., Mattutino, C., Botta, M., Daniele, C., Di Sario, F., Ignone, G., & Cena, F. (2020). Cloud-based user modeling for social robots: a first attempt. In *CEUR WORKSHOP PROCEEDINGS* (Vol. 2724, pp. 1-6). CEUR.

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). 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* (pp. 79-88).

Holstein, K., McLaren, B.M. & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. *LAK*, 257–266 https://doi.org/10.1145/3027385.3027451

Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K. R., & Samek, W. (2022, April). xxAI-beyond explainable artificial intelligence. In *xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers* (pp. 3-10). Springer International Publishing.

Jamaica DPA. (2020). *Jamaica Data Protection Act 2020*. Government of Jamaica

Johnson, W. L., & Lester, J. C. (2016). Face-to-Face interaction with pedagogical agents, twenty years later. International *Journal of Artificial Intelligence in Education*, *26* (1), 25–36. https://doi.org/10.1007/s40593-015-0065-9

Jones, A., Castellano, G., & Bull, S. (2014). Investigating the effect of a robotic tutor on learner perception of skill-based feedback. In M. Beetz, B. Johnston, & M.- A. Williams (Eds.), *Social robotics* (Vol. 8755, pp. 186–195). Springer International Publishing. https://doi.org/10.1007/978-3-319-11973-1_19.

Kennedy, J., Lemaignan, S., & Belpaeme, T. (2016). *The cautious attitude of teachers towards social robots in schools*. Robots 4 Learning Workshop at IEEE RO-MAN 2016.

Khosravi, H., Shabaninejad, S., Bakharia, A., Sadiq, S., Indulska, M., & Gašević, D. (2021). Intelligent Learning Analytics Dashboards: Automated Drill-Down Recommendations to Support Teacher Data Exploration. *Journal of Learning Analytics*, *8*(3), 133-154. https://doi.org/10.18608/jla.2021.7279

Koedinger, K., Liu, R., Stamper, J., Thille, C., & Pavlik, P. (2017). Community based educational data repositories and analysis tools. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 524-525).

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media & Communication*, *7*(3), 412-434. https://doi.org/10.1177/2050157919843961

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: A guidance for policymakers*. UNESCO Publishing.

Microsoft (2017). https://www.microsoft.com/en-us/hololens

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Miller, J. K., Friedman, B., Jancke, G., & Gill, B. (2007). Value tensions in design: The value sensitive design, development, and appropriation of a corporation's groupware system. In *Proceedings of the 2007 international ACM conference on supporting group work* (pp. 281–290). https://doi.org/10.1145/1316624.1316668

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... & Xu, W. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction*, 1-47. https://doi.org/10.1080/10447318.2022.2153320

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*, 137-141. https://doi.org/10.1007/s11747-019-00710-5

Rosanda, V., & Istenič Starčič, A. (2019). A review of social robots in classrooms: Emerging educational technology and teacher education | Education and Self Development. *Education and Self Development*, *14*(3), 93–106. https://doi.org/10.26907/esd14.3.09

Serholt, S., & Barendregt, W. (2014). Students' attitudes towards the possible future of social robots in education. *Workshop Proceedings of Ro-Man*, *6*.

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., & Paiva, A. (2017). The case of classroom robots: Teachers' deliberations on the ethical tensions. *AI & Society*, *32*(4), 613–631. https://doi.org/10.1007/s00146-016-0667-2

Show, B. (2023). Just 1 in 10 teachers think the Online Safety Bill will protect children online https://www.fenews.co.uk/education/just-1-in-10-teachers-think-the-online-safety-bill-will-protect-children-online/

Smakman, M., Vogt, P., & Konijn, E. A. (2021). Moral considerations on social robots in education: A multi-stakeholder perspective. *Computers & Education*, *174*, 104317.https://doi.org/10.1016/j.compedu.2021.104317

Solove, D. J. (2002). Conceptualizing privacy. *California Law Review*, *90*(4), 1087. https://doi.org/10.2307/3481326

Srinivasan, R., & Chander, A. (2020). Explanation perspectives from the cognitive sciences—a survey. In *Proceedings of the twenty-ninth international joint conference on artificial intelligence, international joint conferences on artificial intelligence organization* (pp. 4812–4818)

Sullivan, M.J. (2022). *An introduction to Solid*. https://www.ateam-oracle.com/post/an-introduction-to-solid

Trengove, M., Kazim, E., Almeida, D., Hilliard, A., Zannone, S., & Lomas, E. (2022). A critical review of the Online Safety Bill. *Patterns*, *3*(8), 1-10. https://doi.org/10.1016/j.patter.2022.100544

UNESCO. (2022). *Reimagining our futures together: A new social contract for education*. UN.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500-1509. https://doi.org/10.1177/0002764213479363

Winkler, T., & Spiekermann, S. (2021). Twenty years of value sensitive design: A review of methodological practices in VSD projects. *Ethics and Information Technology*, *23*(1), 17–21. https://doi.org/10.1007/s10676-018-9476-2