Towards Socio-technical Design for Responsible Voice-enabled AI: Long-term Embedding in Primary School Classrooms

Bilal Ahmad[1]*
Learnovate Centre, Trinity College Dublin, Ireland.
bilal.ahmad@learnovatecentre.org
Helena Ledermann
Learnovate Centre, Trinity College Dublin, Ireland.
helena.ledermann@learnovatecentre.org

CCS CONCEPTS • voice-enabled AI • socio-technical design • primary education

Additional Keywords and Phrases: anthropomorphism, empathy gap, triadic interaction

Abstract

This paper examines voice-enabled AI in primary school classrooms, arguing that current evidence from short-term, small-scale pilots is insufficient to guide responsible long-term integration for educational use. Existing studies highlight promising gains in areas such as engagement, vocabulary, pronunciation, and reading, but focus primarily on functional utility and technical accuracy rather than broader educational value. We contend that voice-enabled AI must be developed and evaluated using a socio-technical system design approach that supports triadic interaction between students, teachers or parents, and AI, and aligns with pedagogical frameworks such as Biesta's three functions of education and Vygotsky's socio-cultural theory. Drawing on existing research on participatory, classroom-rooted design, we show how pedagogically appropriate voice-enabled AI can foster skill development, community belonging, and student agency. We also identify key risks in the literature, including the empathy gap and children's tendency to anthropomorphise AI, and argue that these require explicit design, governance, and classroom-level responses, alongside compliance with regulations such as the General Data Protection Regulation (GDPR) and the Children's Online Privacy Protection Act (COPPA). The paper concludes that, to be integrated responsibly, voice-enabled AI in primary education should be treated not as a standalone solution that adheres to a checklist of guidelines, but as an evolving socio-technical environment that requires the continuous management of ethical trade-offs to support children's rights, wellbeing, and long-term learning.

1 Introduction

Voice-enabled AI in primary schools is a rapidly evolving field [Arinah et al., 2025] and has the potential to positively impact student outcomes [Wen et al., 2024] [Elmaadaway et al., 2025]. Growing evidence from international research suggests that these tools can support outcomes such as student engagement and vocabulary development [Quan et al., 2026], pronunciation [Reddy et al., 2025], academic gains [Butler & Starkey, 2024] and language learning and cognitive engagement [Wen et al., 2024], reading comprehension and oral reading fluency [Elmaadaway et al., 2025]. However, we argue that these gains are typically measured under tightly constrained conditions with relatively small sample sizes and usually over short periods of time. The focus of these studies is also on functional utility like technical accuracy of tools

[1]* Corresponding author : Bilal Ahmad (bilal.ahmad@learnovatecentre.org)

and perceived usefulness, with limited attention paid to the educational value [Butler & Starkey, 2024]. This is significant because technologies that appear effective in short pilot conditions may operate very differently once introduced into the everyday realities of primary classrooms. Short pilots cannot capture how children gradually develop trust in voice-enabled AI or build learning habits around using it; we contend that these patterns only become visible over time in real classroom settings.

2 Position Statement

Recent research and policy developments are a good start, but we argue that responsibly integrating voice-enabled AI into primary school classrooms will require socio-technical design approaches [Bauer et al., 2009] that support triadic interaction [Quan et al., 2026]. This is because voice-enabled AI systems and classroom social dynamics are so intertwined that it is inevitable to optimise both technological and social aspects. These systems must prioritise alignment with pedagogical frameworks, and explicitly address risks such as the empathy gap and anthropomorphism [Kurian, 2025a; 2025b]. Although there are emerging examples of participatory, classroom-rooted voice-enabled AI design in schools [Metatla et al., 2019], most current uses of voice-enabled AI in education repurpose existing commercial voice assistants and applications in predominantly product-led ways, with limited systematic attention to socio-technical design or local classroom dynamics (Terzopoulos & Satratzemi, 2020). We also maintain that adherence to standard regulations such as GDPR, COPPA, and the management of ethical dilemmas are essential to ensure long-term integration within classrooms.

3 Discussion

Triadic interaction design is an approach in which teachers and/or parents are given the agency to act as moderators when students are using such tools for learning. We argue that this calls for a socio-technical design of such tools, not simply because classrooms contain both people and technologies, but because educational outcomes emerge from their social interaction and use of voice-enabled AI over time. These considerations can also lead to the development of voice-enabled AI that is pedagogically appropriate and aligns with established frameworks such as Biesta's three functions of education [Biesta 2015] or Vygotsky's socio-cultural theory [Vygotsky, 1978]. Metatla et al.'s (2019) voice user interfaces co-design project offers an example of pedagogically appropriate voice-enabled AI, where visually impaired and sighted pupils worked alongside teachers to design a shared voice tool for peer revision, fostering the dialogic, collaborative learning that underpins Vygotsky's (1978) socio-cultural theory. Rather than being designed externally and handed to teachers to implement, the tool emerged directly from the lived experiences of the classroom community, with students' own insights shaping its features and purpose. This suggests that voice AI is pedagogically appropriate not simply because it functions technically, but because it can support community belonging, skill development, and student agency, which are the three core purposes of education identified by Biesta (2015). For this reason, responsible design must ask not only whether a tool works, but what kinds of participation, language use, and relationships it normalises in the classroom.

It is also important to address the empathy gap [Kurian 2025b], a paradox where AI mimics human-like care without having genuine emotional comprehension, leading students to unsafe responses. For primary school students in particular, this simulation of care can be remarkably convincing. A child who is struggling academically, feeling isolated, or simply seeking validation may interpret an AI's encouraging responses as authentic concern. This misreading can lead students to place disproportionate trust in AI-generated guidance, follow responses that are emotionally reassuring but factually or ethically unsafe, and fail to seek the human support they actually need. The empathy gap is not a flaw, but rather a structural characteristic of conversational AI, that educators and platform designers must actively account for through systemic and design-level changes.

Another related characteristic of primary school students that needs to be proactively managed is anthropomorphism [Kim et al., 2025], in which a student may perceive the technology as a human confidante, leading to over-trust and the disclosure of excessive personal information. Young children like those in primary schools are naturally inclined to see

human qualities in the things around them [Goldman & Poulin-Dubois, 2024], and conversational AI, which speaks in plain language and responds in a warm, personable way, makes this especially easy to do [Kim et al., 2025]. If this is not actively addressed, anthropomorphism can gradually chip away at both a student's wellbeing and their ability to think critically about the technology they use [Stadler et al., 2024; iRaise, 2026]. This is why schools need clear, age-appropriate guidance on how students understand and relate to AI, not as an afterthought, but as a core part of how these tools are introduced from the start.

For designers of such systems, it is important to navigate ethical dilemmas such as balancing safety versus engagement, personalisation versus privacy and autonomy versus protection [Kurian 2025b]. There is a need to explicitly manage these trade-offs as, by-default, such systems optimisation logics may undermine students' safety through nudging techniques that encourage over-disclosure. Responsible design, therefore, depends on making such trade-offs explicit at the level of interface, governance and classroom use, rather than treating them as once-off issues to be addressed solely at the development stage.

4 Conclusion

Finally, the long-term responsible embedding of voice-enabled AI tools must be supported by the aforementioned considerations along with regulations such as GDPR, COPPA and any other local guidelines. As evaluation shifts from short-term assessments to longitudinal approaches, the focus must also move beyond mere functional utility and technical accuracy towards primary school students' development rights, socio-emotional wellbeing and broader educational experiences. Rather than asking only whether a tool improves immediate attainment, studies should examine how voice-enabled AI shapes classroom discourse, teacher decision-making, trust, participation, inclusion, and wellbeing over time. This would better capture whether such tools become responsibly embedded as educational infrastructures rather than remaining short-term novelties with isolated functional benefits. We conclude that responsible voice-enabled AI in primary education should therefore be understood not as a standalone tool that adheres to a checklist of guidelines, but as a socio-technical system that continually manages ethical tradeoffs to achieve positive educational impact over time.

REFERENCES

[1] Arinah, A., Oya, A. and Suryaningsih, S., 2025. A bibliometric analysis of smart voice assistants for EFL listening and speaking in elementary education. Journal of Language and Literacy Learning, 1(2), pp.62-75.

[2] Bauer, J. M., & Herder, P. M. (2009). Designing socio-technical systems. In Philosophy of technology and engineering sciences (pp. 601-630). North-Holland.

[3] Biesta, G. (2015). What is education for? On good education, teacher judgement, and educational professionalism. European Journal of education, 50(1), 75-87.

[4] Butler, L. and Starkey, L., 2024. OK Google, help me learn: an exploratory study of voice-activated artificial intelligence in the classroom. Technology, Pedagogy and Education, 33(2), pp.135-148.

[5] Elmaadaway, M.A.N., El‐Naggar, M.E. and Abouhashesh, M.R.I., 2025. Improving Primary School Students' Oral Reading Fluency Through Voice Chatbot‐Based AI. Journal of Computer Assisted Learning, 41(2), p.e70019.

[6] Goldman, E. J., & Poulin‐Dubois, D. (2024). Children's anthropomorphism of inanimate agents. Wiley Interdisciplinary Reviews: Cognitive Science, 15(4), e1676.

[7] iRaise & Paris Peace Forum. (2026). Adolescents & anthropomorphic AI: Rethinking design for wellbeing. Paris Peace Forum. https://parispeaceforum.org/publications/adolescents-anthropomorphic-ai-rethinking-design-for-wellbeing/

[8] Kim, P., Chin, J. H., Xie, Y., Brady, N., Yeh, T., & Yang, S. (2025). Young children's anthropomorphism of an AI chatbot: Brain activation and the role of parent co-presence. arXiv preprint arXiv:2512.02179.

[9] Kurian, N., 2025a, July. Designing Child-Safe Conversational AI: Three Dilemmas for Responsible Design. In Proceedings of the 7th ACM Conference on Conversational User Interfaces (pp. 1-5).

[10] Kurian, N., 2025b. 'No, Alexa, no!': Designing child-safe AI and protecting children from the risks of the 'empathy gap'in large

language models. Learning, Media and Technology, 50(4), pp.621-634.

[11] Metatla, O., Ahmed, T., Oldfield, A., Miglani, S., & Vafeas, A. (2019). Voice user interfaces in schools: Co-designing for inclusion with visually-impaired and sighted pupils. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Paper 459, pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300689

[12] Quan, S., Tu-Shea, X., Ding, Y., Du, Y., Zheng, Q. and Gerdich, L.E., 2026. Conversational AI in children's home literacy learning: effectiveness, advantages, challenges, and family perception. Computers and Education, 10(10054), p.9.

[13] Reddy, P.A., Reddy, K.A. and Anand, S., 2025, November. Voice-Enabled Personalized Tutoring System for Rural English Learners. In 2025 5th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS) (pp. 1534-1541). IEEE.

[14] Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Computers in Human Behavior, 160, 108386. https://doi.org/10.1016/j.chb.2024.108386

[15] Terzopoulos, G., & Satratzemi, M. (2020). Voice assistants and smart speakers in everyday life and in education. *Informatics in Education, 19*(3), 473–490. https://doi.org/10.15388/infedu.2020.21

[16] Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.). Cambridge, MA: Harvard University Press.

[17] Wen, Y., Chiu, M., Guo, X. and Wang, Z., 2024. AI‐powered vocabulary learning for lower primary school students. British Journal of Educational Technology, 56(2), pp.734-754.