

Raspberry Pi
Foundation

Understanding computing education

Volume 3

Theme: AI, data science, and young people

Understanding computing education

Volume 3

Theme: AI, data science, and young people

Raspberry Pi Foundation Research Seminars

Published in December 2022
by the Raspberry Pi Foundation

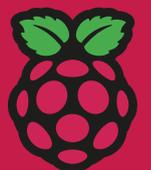
www.raspberrypi.org

ISSN 2514-586X (18)

Table of contents

Understanding computing education, Volume 3 - Theme: AI, data science, and young people

Foreword	5
Information about the authors	6
<i>Perspectives on AI and data science education</i>	10
<i>Sue Sentance and Jane Waite (University of Cambridge and Raspberry Pi Foundation)</i>	
<i>Engaging children with AI ethics</i>	19
<i>Mhairi Aitken and Morgan Briggs (The Alan Turing Institute)</i>	
<i>How can we make AI education a priority without using scare tactics?</i>	31
<i>Rose Luckin (University College London)</i>	
<i>Exploring the data-driven world: teaching AI and ML from a data-centric perspective</i>	40
<i>Carsten Schulte, Yannik Fleischer, Lukas Höper, Rolf Biehler, Daniel Frischemeier, Sven Hüsing, and Susanne Podworny (Paderborn University)</i>	
<i>Cross-boundary co-design for learning machine learning</i>	48
<i>Henriikka Vartiainen (University of Eastern Finland)</i>	



Foreword

Artificial Intelligence (AI) and data science are two overlapping fields that are advancing rapidly and having huge impacts on our lives; for example, machine learning (ML) systems are now used to make decisions in areas such as healthcare, finance, education, social care, and employment. Here at the Raspberry Pi Foundation, AI, machine learning, and data science are important topics both in our learning resources for young people and educators, and in our programme of research. We're particularly interested in the way young people learn AI and data science, rather than the use of AI systems as tools for education.

Between September 2021 and March 2022, we were delighted to offer, in partnership with the Alan Turing Institute, seven online seminars presenting different perspectives on AI and data science education for young people, as well as a panel discussion. The series proved very popular with attendees from education, research, industry, and other sectors, and we enjoyed animated breakout sessions following the presentations and insightful discussions with the speakers. How you teach AI and data science to young people is a nascent field, and our seminar speakers all approached the challenges it brings from different perspectives. Taken together, I feel that the seminars gave a realistic view of where we are in AI and data science education at the moment.

I'm now even more excited to share with you chapters from four of the research groups

who presented at the seminar series. Mhairi Aitken and Morgan Briggs give a fascinating introduction to AI ethics and how we might introduce it to young children. I really recommend Rose Luckin's reflections on why AI is important for everyone to understand and how we might educate the educators. From Paderborn University, Germany, we hear from Carsten Schulte and his colleagues about how we need to rethink the way we teach AI and ML to young people, and finally, Henriikka Vartiainen of the University of Eastern Finland describes a project that used co-design to develop ML learning projects for Finnish basic education. Jane Waite and I have also included a synthesis of all the seminars and how, together, they can help us frame future work in this area. We've also introduced our SEAME model, which we are using at the Raspberry Pi Foundation to frame the plethora of resources that are being developed to teach AI. One thing is clear from all these chapters: there is still more research needed to understand the teaching and learning of AI more fully. We aren't even touching the surface of knowing what it means to educate young people about AI.

Our seminar series continues monthly on the first Tuesday of the month and you can find the up-to-date schedule at rpf.io/research-seminars. We've hosted over 30 seminars since March 2020. We hope that you enjoy these chapters from previous seminars and that you come and join us for future ones!

Dr Sue Sentance
Chief Learning Officer
Raspberry Pi Foundation
December 2022

Information about the authors (in alphabetical order)



Mhairi Aitken
(The Alan Turing Institute, UK)

Mhairi Aitken
(The Alan Turing Institute, UK)

Dr Mhairi Aitken is an Ethics Fellow in the Public Policy Programme at The Alan Turing Institute, and an Honorary Senior Fellow at ACHEEV, University of Wollongong in Australia. She is a Sociologist whose research examines the social and ethical dimensions of digital innovation, particularly relating to uses of data and AI. She has also held roles as a Public Engagement Research Fellow in both the Farr Institute of Health Informatics Research and the Scottish Health Informatics Programme (SHIP).



Morgan Briggs
(The Alan Turing Institute, UK)

Morgan Briggs
(The Alan Turing Institute, UK)

Morgan Briggs currently serves as the Research Associate for Data Science and Ethics within the Public Policy Programme at The Alan Turing Institute. Morgan works on a variety of projects at The Alan Turing Institute, including continued, ongoing work on AI explainability. In addition to her role at The Alan Turing Institute, Morgan serves as an Associate Teaching Fellow for the Global Leadership Initiative at the University of Oxford.



Yannik Fleischer
(Paderborn University, Germany)

Yannik Fleischer
(Paderborn University, Germany)

Yannik Fleischer is a PhD student in mathematics education research at Paderborn University, Germany. His main research interest is to develop a concept to teach machine learning methods in school with a focus on decision trees, and to evaluate this by developing and examining teaching materials in practice. Since 2019, he has been supervising year-long project courses on data science in upper secondary and developing, implementing, and evaluating teaching modules for different levels in secondary school, mainly about machine learning with decision trees.



Lukas Höper
(Paderborn University, Germany)

Lukas Höper
(Paderborn University, Germany)

Lukas Höper is a PhD student in computing education research at Paderborn University, Germany. His main research interest is to develop the concept of data awareness for computing education and evaluate this by developing and examining teaching materials in practice. Since 2020, he has been working on data awareness in the ProDaBi project, among other topics on AI and data science in schools.



Rose Luckin
(University College London, UK)

Rose Luckin
(University College London, UK)

Dr Rosemary (Rose) Luckin is Professor of Learner Centred Design at UCL Knowledge Lab. She was named one of the 20 most influential people in education in the Seldon List, 2017. Rose is Founder of EDUCATE Ventures Research Ltd., a London hub for start-ups, researchers, and educators. She is treasurer of the International Society for AI in Education and co-founder of the Institute for Ethical AI in Education. Prior to joining Knowledge Lab in 2006, Rose was Pro-Vice Chancellor for Teaching and Learning at the University of Sussex.



Carsten Schulte
(Paderborn University, Germany)

Carsten Schulte
(Paderborn University, Germany)

Dr Carsten Schulte is a professor of computing education research at Paderborn University, Germany. His work and research interests are the philosophy of computing education and empirical research into teaching-learning processes (including eye movement research). Since 2017, he has been working together with Didactics of Mathematics (Paderborn University) in the ProDaBi project, in which data science and artificial intelligence are prepared as teaching topics. He is also PI in the collaborative research centre 'Constructing Explainability' on explainable AI.



Sue Sentance
(University of Cambridge and Raspberry Pi Foundation, UK)

Sue Sentance
(University of Cambridge and Raspberry Pi Foundation, UK)

Dr Sue Sentance is Chief Learning Officer at the Raspberry Pi Foundation and Director of the Raspberry Pi Computing Education Research Centre. She researches the teaching of programming in schools, teacher professional development, and physical computing. Her academic background is in computer science, artificial intelligence, and education, and she is a qualified teacher and teacher educator. She has created and researched the PRIMM methodology for structuring programming lessons in school.



Henriikka Vartiainen
(University of Eastern Finland, Finland)

Henriikka Vartiainen
(University of Eastern Finland, Finland)

Dr Henriikka Vartiainen is a senior researcher and university lecturer at the University of Eastern Finland, School of Applied Educational Science and Teacher Education. She has also worked as responsible researcher in several multidisciplinary projects focusing on, for example, technology education, co-design in school context, and design-oriented pedagogy. Currently, her research focuses especially on learning machine learning through co-design as well as on the ways to support children's data agency.



Jane Waite
(University of Cambridge and Raspberry Pi Foundation, UK)

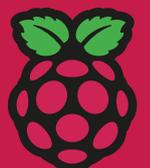
Jane Waite
(University of Cambridge and Raspberry Pi Foundation, UK)

Jane Waite is a computing education researcher who has worked in industry and as a classroom teacher for many years. She currently works as the senior research scientist at the Raspberry Pi Foundation in their research team. Jane is currently working on a wide range of research projects, including investigating culturally relevant pedagogy for teaching computing and looking at the underpinning concepts for teaching and learning AI in schools. She has published on a wide range of topics, such as on computational thinking with Tim Bell and Paul Curzon, pedagogy for teaching programming for the Royal Society, and primary program design, which is her main interest and passion.

Perspectives on AI and data science education

Sue Sentance and Jane Waite

(University of Cambridge and Raspberry Pi Foundation)



Perspectives on AI and data science education

Sue Sentance and Jane Waite (University of Cambridge and Raspberry Pi Foundation)

Introduction

Artificial intelligence (AI) and data science education have a far-reaching and growing impact on our lives, and it is important for young people to understand them both from a technical and a societal perspective, and for educators to learn how to best support them to gain this understanding.

From September 2021 to March 2022, the Raspberry Pi Foundation ran a series of seven seminars on the topic of AI and data science education for young people. The objectives of the seminar series were:

1. To learn from experts in the field about their perspectives on the future of AI and data science education for young people;
2. To develop a community of interested researchers, teachers, and industry experts around this topic.

The invited speakers brought a range of different perspectives to the topic, in terms of their approaches to theory, resources, and their ambitions for AI and data science education. In this short article, we summarise our understanding of their presentations and how their work may contribute to a research agenda for this new and emerging field.

Four of these seminars have been supplemented with chapters in these proceedings. We also held a special panel session including young people and a UK Minister, which looked at policy and perspectives of AI and data science education as a school subject. You can view all the seminars at this link: rpf.io/ai-research-seminars.

Why teach AI?

In the UK, like many countries, we have a very crowded school curriculum with many different subjects jostling for curriculum time. We have made progress globally in introducing computer science into some school-aged contexts, mostly at the secondary school level (Vegas et al., 2021). However, AI is a subject typically taught at the Master's level, although some undergraduate degrees in the topic have been available in recent years. Although AI is spoken about in many contexts, we may not even have a shared definition of what it really is or covers. So why would we even consider adding the teaching of AI either in school or in non-formal settings? We still have much to learn about how and what to teach in terms of AI, machine learning, and data science. During the seminar series, our speakers provided a range of different perspectives to the question of why we should teach AI. From these, it is possible to extract a number of different reasons for teaching AI.

Children are already growing up with AI: This is probably the most obvious reason that people cite when thinking about AI. Young children are already surrounded by devices and apps that use AI, so the argument goes that they should learn about how they work to become discerning consumers. Stefania Druga discussed her research on working with families who were developing an understanding of the potential of smart devices (Druga et al., 2021).

AI is impacting children's lives: AI may have far-reaching consequences in children's lives, where it's being used for decision-making around access to resources and support. From an ethical perspective, Mhairi Aitken holds the view that AI systems are already having a significant

Table 1: Seminars hosted by the Raspberry Pi Foundation.

Title	Speaker(s)
AI ethics and engagement with children and young people	Dr Mhairi Aitken, The Alan Turing Institute
Exploring the data-driven world: Teaching AI and ML from a data-centric perspective	Professor Carsten Schulte, Yannik Fleischer, and Lukas Höper, University of Paderborn
ML education for K–12: Emerging trajectories	Professor Matti Tedre and Dr Henriikka Vartiainen, University of Eastern Finland
What is it about AI that makes it useful for teachers and learners?	Professor Rose Luckin, University College London
Teaching artificial intelligence in K–12	Professor Dave Touretzky, Carnegie Mellon University, and Professor Fred Martin, University of Massachusetts Lowell
Teaching youth to use AI to tackle the Sustainable Development Goals	Dr Tara Chklovski, Technovation
Democratising AI education with and for families	Stefania Druga, University of Washington

impact on young people’s lives through systems deployed in children’s education, in apps that children use, and in children’s lives as consumers. Children’s data is being collected, and decisions are being made about them using AI; therefore, awareness of the impact of AI should be raised (Aitken & Briggs, 2022).

AI requires a new way of thinking: Two of our seminars covered the ways in which our understanding of computational thinking changes when we move away from traditional programming to more data-driven approaches. Matti Tedre and Henriikka Vartiainen propose a new version of computational thinking called CT 2.0. In contrast to CT 1.0, which is rule-driven, CT 2.0 is data-driven, so requires skills such as being able to experiment with data (Vartiainen et al., 2021). Dave Touretzky and Fred Martin proposed a broad version of AI thinking, which includes perception, reasoning, representation, machine learning, and language understanding (Touretzky et al., 2019).

People need to use AI safely and effectively: In order to build a citizenry of people who use AI safely and effectively, we need to educate them in the subject. Rose Luckin shared a very broad view of AI – in education, for education, and as part of our education. Luckin emphasised the importance of being able to customise AI tools to your context (Luckin et al., 2016).

We want to empower children to effect change: Understanding AI and data science will be very empowering in the years to come. It’s important that AI education is inclusive and that opportunities to learn about AI are for everyone. To do this, we need to make AI education available and accessible.

Humans are starting to interact with machines in new ways: In our most theoretically-focused seminar, Carsten Schulte argued for a new discipline around machine behaviour and hybrid human interaction, focusing on the ways in which society and individuals interact with data-centric systems (Rohlfing et al., 2021).

We need a skilled AI workforce: This is another reason for teaching AI, one that was not named in any of the talks, but which is put forward by policymakers describing their country's development of AI: in order for a country to lead in developments in AI, it needs a trained workforce with the appropriate technical skills (Galindo et al., 2021). In the UK, we've seen the publication of the National AI Strategy¹ and the AI Roadmap², both highlighting the need for AI

education, and this year's AI action plan³ focuses on supporting the development of a diverse workforce in AI, and other countries in the world have similar policy drives.

Just as there are many views on why we should teach AI, experts and academics hold different views on what we should actually teach within the vast area of AI. It is clear that the motivation for teaching AI to young people impacts the

Table 2: Seven reasons for teaching AI in schools.

Why teach AI?	Implications for teaching content
Children are already growing up with AI	Young people need to learn that there are both drawbacks and advantages of innovative technologies, particularly where they use AI.
AI is impacting children's lives	Creators of systems that children will use should understand that AI may impact their privacy and that systems are being used to make decisions that affect them.
AI requires a new way of thinking	We should teach skills and knowledge around data-driven programming and how AI works in addition to traditional programming techniques.
People need to use it safely and effectively	Young people should be taught to use AI tools and applications.
We want to empower children to effect change	We should ensure that there are opportunities to learn all aspects of AI, both technical and socio-ethical, for all children at an early age.
Humans are starting to communicate with machines in new ways	We need to teach students about the ways that machines, including AI systems, impact individuals and society, and to be curious about the way machines behave.
We need a skilled AI workforce	We need to provide a progression of learning opportunities that lead towards highly technical courses in AI later on in school or ensure that facilitating subjects such as mathematics, physics, and computer science are taught effectively to all.

¹ <https://www.gov.uk/government/publications/national-ai-strategy>

² <https://www.gov.uk/government/publications/ai-roadmap>

³ <https://www.gov.uk/government/publications/national-ai-strategy-ai-action-plan/national-ai-strategy-ai-action-plan>

actual content that might be taught, and at what age and stage it could be introduced. Table 2 summarises, at a very high level, the implications of certain motivations for teaching AI on the type of content that might be needed.

It is clear that we need to have specific goals in AI education, and that curriculum developers and educational resource developers may have different views on what we need to teach. In one of the seminars, we heard about the five big ideas of AI from Dave Touretzky and Fred Martin (Touretzky et al., 2019). The five big ideas from the AI4K12 project⁴ are perception, representation and reasoning, learning, natural interaction, and societal impact. These have been really useful in both mapping to school standards in the US in computer science, and also in giving a framework for resource developers. In this way, the big ideas used by the AI4K12 framework help to show the breadth of AI content that we could cover. Our own research (in progress) has shown that many current resources focus on machine learning, so the AI4K12 framework highlights other areas of AI that could be studied.

However, there is another dimension: the degree to which we abstract from the technical aspects of AI. Do we teach children how to actually create AI, or do we teach them how it impacts them and how to be informed users of AI? And there is much in between those two aspects of the subject.

In Appendix 1, we have included a simple framework that we are using to categorise different levels of AI as SE (socio-ethical), A (applications), M (models), and E (the engine – or how AI works). This gives us a way of understanding different resources and their learning goals. It provides levels of abstraction for the subject, with the SE level most abstracted from the technical aspects. We are calling this the SEAME Framework.

Our seminar speakers had different perspectives on which of these elements were important

to be understood. While Mhairi Aitken gave an excellent exposition of ethical issues and why we should engage children in them (focusing on the SE level), Dave Touretzky and Fred Martin talked about the fact that while young children might be using applications of AI (the A level), older children should engage more with the models of AI (the M level) and argued for transparent AI demonstrations that made the E level visible. Rose Luckin focused on teachers' knowledge of AI, which she argued could be developed by actually using data to create a model (the M level).

How and when would we teach it?

Computing is increasingly being introduced into curricula around the world, and in England has been mandatory for students aged 5–16 since 2014. Students in many countries can opt to take computer science as an elective in upper secondary or high school.

AI at the university level is likely to be included as part of a computer science department's course, but it may not necessarily follow that AI education will fit into the computing school curriculum. Some of the socio-ethical components could be addressed across a range of school subjects, for example.

Some of our seminar speakers gave examples of AI education in **non-formal** settings. For example, Stefania Druga shared the findings of studies carried out with families working in an informal way with their children (Drugá et al., 2021), and Tara Chklovski gave examples of an annual challenge that children could sign up to as an extracurricular activity. Some of the work described in Matti Tedre and Henriikka Vartiainen's talk took place in the homes of young children aged from 3–9 (Vartiainen et al., 2021).

In terms of **formal** education, the Finnish speakers have also conducted some research with 11–12-year-olds in schools. The AI4K12

⁴ <https://ai4k12.org/>

project is intended for formal education, through its mapping to the CSTA (Computer Science Teachers Association) standards (Touretzky et al., 2019), although in reality many US states may not have the curriculum in place to deliver this content. ProDaBi⁵, the data science and AI education project described by the researchers from Paderborn, is being designed for students in school at the lower secondary and upper secondary levels.

The age range of children in the studies we heard about in the seminars ranged from 3–18 years old, so it was clear that the discussion about AI education spans from kindergarten through primary and secondary education. One question that came up often from our seminar audiences was the extent to which teachers can be involved in research projects such as those described, and the level of training being developed for teachers to enable them to teach and understand AI. Some research projects we heard about were conducted in a participatory way, and certainly the AI4K12 project has developed sufficiently to have courses for teachers embedded into it. Apart from Rose Luckin, who described an adult-facing programme that could be used to support teachers wishing to understand AI, the seminar speakers did not particularly focus on the needs of the teacher in this context.

Our speakers had different views on how AI should be taught to young people. It was clear that AI is relevant to young people's lives, and Tara Chklovski highlighted how young people could be engaged in building solutions to problems that they could see in their own lives by accessing technology (Chklovski et al., 2021). Other speakers discussed how AI might require a change to the way we think. For example, Matti Tedre, of the University of Eastern Finland, proposed CT 2.0, which he's written about elsewhere, explaining that a data-driven approach to solving a problem is fundamentally different to writing an algorithm to solve it (Tedre et al., 2021). Carsten Schulte and colleagues, of the University of Paderborn in Germany, also highlighted issues

around the role of code and the approach to accuracy, and how these are both different in machine learning in comparison to traditional programming (Rohlfing et al., 2021). Both research groups are developing resources that reflect these differences, and studying the way that learners interact with them. This is interesting work, and we will be following the updates of these two research groups with much interest.

Where is research needed?

What our speakers said they wanted to do next gives us an interesting range of ideas for future research. Tara Chklovski wants to continue to broaden participation by ensuring that more girls and underrepresented groups in computing can access the opportunities to develop AI skills through team challenges. Stefania Druga calls on us to consider family life as a third space for AI learning and suggests there is much more research to do in this area. The AI4K12 project is concerned with reaching more US states and also proposed further work on tools development around teaching AI thinking. Rose Luckin's work at UCL is much broader than our specific context and extends to the use of AI in education, where there is much to do to ensure that this is implemented ethically. Linked to this, Mhairi Aitken's future work will involve actually engaging with children to support ethical practices in AI: this is crucial, as we so often ignore the young person's voice. Matti Tedre and Henriikka Vartiainen left us with many questions and challenges regarding how we can make the shift from CT 1.0 to CT 2.0. Aligned to this, Carsten Schulte's summary included a call to action for us to conduct research that helps us to understand the data-driven and emergent ecosystem, and to investigate how that might impact a paradigm change in teaching.

Drawing together these ideas for future work, we have suggested four areas which we believe should be included within a research agenda for AI education:

⁵ <https://www.prodabi.de/>

1. **Teaching and learning.** It's clear that our traditional approach to teaching programming, which involves writing an algorithm that can be implemented, will need to change as we introduce young people to more data-driven approaches to solving problems. What does CT 2.0 look like in a teaching and learning environment?
2. **Learner voice.** We need to engage learners in research around their perception of issues that affect them. There are links to culturally responsive computing research as well as opportunities to develop learners' thinking around social justice and equity in our teaching about AI.
3. **Teachers/educators.** At the Raspberry Pi Foundation, we've recently conducted a

literature review of empirical research in the area, which demonstrates that educators are not often included in studies or not seen as a stakeholder in AI education work. There is much to do here.

4. **Tools and resources.** We've conducted a mapping exercise looking at AI resources written for children, which surfaces a complex picture, beyond simply what and who is taught. For example, the choice of software can limit the transparency of what is being taught. A framework for resource development would be a useful addition to the field.

We're going to start working on some of the four areas listed and would encourage others to do the same. There is much research to be done!

References

- Aitken, M. & Briggs, M. (2022). In AI, data science, and young people. Understanding computing education (Vol 3). *Proceedings of the Raspberry Pi Foundation Research Seminars*. [rpf.io/seminar-proceedings-vol-3-aitken-briggs](https://doi.org/10.1162/ba67f642.646d0673)
- Chklovski, T., Jung, R., Anderson, R., & Young, K. (2021). Comparing 2 Years of Empowering Families to Solve Real-World Problems with AI. *KI-Künstliche Intelligenz*, 35(2), 207–219. <https://doi.org/10.1007/s13218-021-00738-2>
- Druga, S., Yip, J., Preston, M., & Dillon, D. (2021). The 4As: Ask, Adapt, Author, Analyze - AI Literacy Framework for Families. In *Algorithmic Rights and Protections for Children*. <https://doi.org/10.1162/ba67f642.646d0673>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. UCL Knowledge Lab. <http://oro.open.ac.uk/50104/>
- Galindo, L., Perset, K., & Sheeka, F. (2021). An overview of national AI strategies and policies. *Going Digital Toolkit Note*, 14. <https://doi.org/10.1787/c05140d9-en>
- Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., Buschmeier, H., Esposito, E., Grimminger, A., Hammer, B., Hab-Umbach, R., Horwath, I., Hüllermeier, E., Kern, F., Kopp, S., Thommes, K., Ngonga Ngomo, A.-C., Schulte, C., Wachsmuth, H., Wagner, P., & Wrede, B. (2021). Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3), 717–728. <https://doi.org/10.1109/TCDS.2020.3044366>
- Tedre, M., Denning, P., & Toivonen, T. (2021). CT 2.0. *21st Koli Calling International Conference on Computing Education Research*, Article 3. <https://doi.org/10.1145/3488042.3488053>
- Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K–12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
- Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., & Valtonen, T. (2021). Machine learning for middle schoolers: Learning through data-driven design. *International Journal of Child-Computer Interaction*, 29, Article 100281. <https://doi.org/10.1016/j.ijcci.2021.100281>
- Vegas, E., Hansen, M., & Fowler, B. (2021). Building skills for life: How to expand and improve computer science education around the world. The Brookings Institution. https://www.brookings.edu/wp-content/uploads/2021/10/Building_skills_for_life.pdf

Appendix 1: The SEAME framework

A simple learning levels framework can be used to categorise research and resources. This is derived from a framework developed at Queen Mary University of London by Jane Waite and Paul Curzon⁶.

The AI learning levels framework (SEAME)

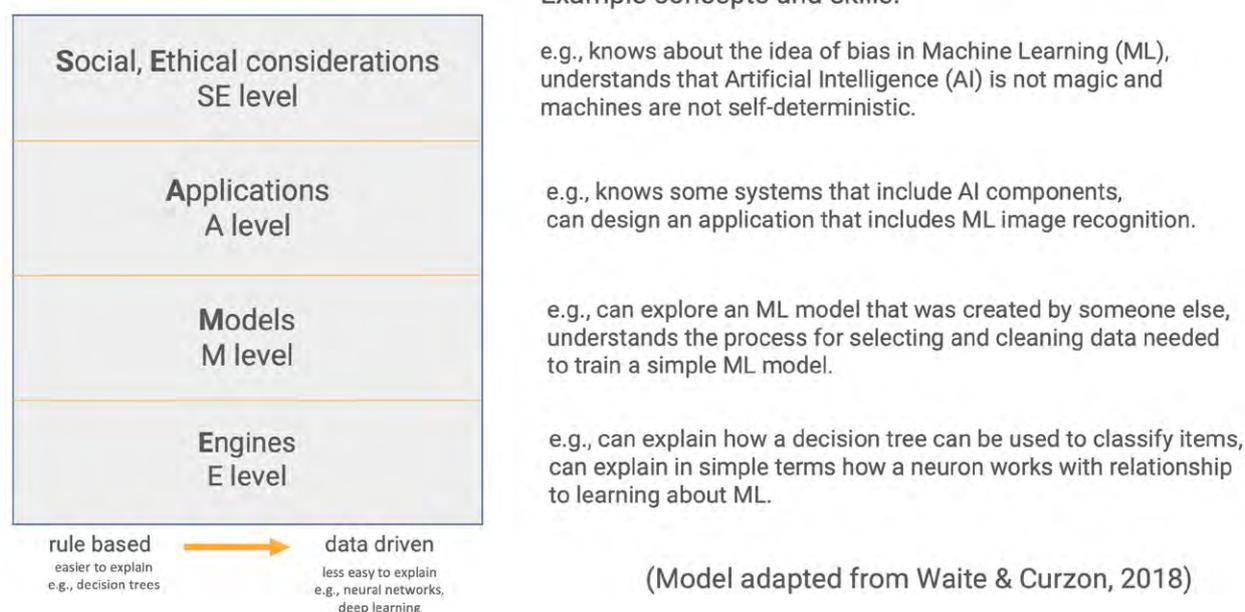


Figure 1. The AI learning levels framework (a revised version of Waite & Curzon, 2018⁶).

The SEAME AI learning levels framework used in our studies is shown in Figure 1. This framework has four levels and provides a simple way to reflect upon the content included in AI resources and activities.

SE: This is the level of **social** and **ethical** considerations.

A: This is the **applications** level, where we might use, modify, or create applications that have some AI or ML component.

M: This is the **models** level, where we train the model with data. Models output recommendations and predictions for use in applications.

⁶<https://teachinglondoncomputing.org/machine-learning/>

E: This is the engines level, including neural networks, generative algorithms, data structures, etc. This is the most hidden level, which we are not aware of when we use an application with an ML component.

The framework will require full evaluation, but is currently providing a valuable way for the Raspberry Pi Foundation to review and reflect on available research and resources. It is not intended to cover data science resources and research, as there are aspects of data science that are more statistics related, but it covers aspects of early data literacy.

How to use the framework

The framework can be used to categorise resources developed using AI. Some examples are given below.

The ethical dilemmas of self-driving vehicles as discussed with students can be described as level SE, the level relating to ethical and societal considerations.

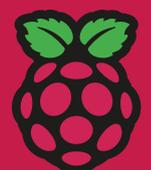
Some activities might span two levels. For example, an activity where students use an existing 'rock-paper-scissors' application that uses an ML model to recognise hand shapes works at the Applications level. If students then move on to train the model to improve accuracy by adding more image data, they work at the Model level.

Other resources drill down through the layers for a single concept. For example, if studying bias, an activity might start with an example of the societal impact of bias. Students might then discuss the applications they use personally to reflect on bias, and the activity might finish with students exploring data in a simple ML model. This involves students working through layers SE, A, and M.

Another approach to using the framework is to see whether some age groups might have more learning activities available at one level than another and whether this changes over time. For example, younger learners might work mostly at levels SE and A, and older learners might move between the levels with increasing clarity as they develop their knowledge.

Engaging children with AI ethics

Mhairi Aitken and Morgan Briggs
(The Alan Turing Institute)



Engaging children with AI ethics

Mhairi Aitken and Morgan Briggs (The Alan Turing Institute)

Abstract

Internationally, there is growing interest in engaging children with artificial intelligence (AI) and data science. In this paper, we argue that rather than focusing solely on equipping children with skills to be the future AI workforce, we must also aim to equip children with skills to be the future – and current – critical public, which is needed to hold AI systems and their developers to account. As AI is impacting children's lives in ever more ways and increasingly shaping the future societies in which children will live and work, it is vital that children and young people are equipped to interrogate and understand the role of AI systems. This paper makes the case that education relating to AI must go beyond traditional STEM approaches to encompass ethical and social considerations relating to AI. This is important to ensure that children understand the role of AI in their lives (now and in the future) and are able to critically engage with AI to make informed choices about the ways in which they interact with AI. There are also substantial benefits for development and deployment of AI, since children's views and values need to be included in order to inform ethical practice.

Introduction

Internationally, there is growing interest in engaging children and young people with artificial intelligence (AI) and data science. This is considered important to build skills and

capacities, and to equip the next generation to pursue careers in these fields. However, comparatively, little attention is directed at engaging children and young people with discussions of the ethical and social considerations around the ways that AI is designed, developed, and deployed. As AI is impacting children's lives in ever more ways and increasingly shaping the future societies in which children will live and work, it is vital that children and young people are equipped with the skills not just to develop the AI systems of the future but also to interrogate and understand the role of AI systems today. This paper therefore seeks to make the case that education relating to AI must go beyond traditional STEM approaches to encompass ethical and social considerations relating to AI.

Children interact with AI systems in myriad ways on a daily basis. Some of these interactions are intentional (e.g., playing with interactive toys or speaking with voice assistants), whereas others may be much less visible (e.g., in accessing tailored or personalised services, such as in education). AI is present in smart toys that "learn" and develop new skills when children play with them, and in smart home devices such as smart speakers and voice assistants, with which children increasingly interact. AI is also used to sort, filter, and target content online and may have a significant role in shaping children's views of the world, the information they receive, and the friendships they develop (e.g., through social media). AI is also used in ways that impact and shape children's lives through the provision or prioritisation of services in the public sector,

for example, through identifying which children, or families, are considered at risk and require interventions by social services. The significant impacts of such systems both for individual children and families as well as for wider society cannot be overstated.

There are a host of risks AI applications create for children. Not least among these are the potential transformative effects these technologies have on their development and their participation in the communities they belong to. Other big challenges that pose concerns are managing the privacy of children and their families in online settings in which data is constantly being collected about them.

There has been significant research conducted across many disciplines including psychology, education, healthcare/social services, etc. While there are different points emphasised across these literatures, there are some key points of overlap throughout. In the field of psychology, research has found that AI devices can alter young children's perceptions of their own intelligence (Druga et al., 2017; Howley, 2019; Williams et al., 2019). There is also a wider discussion surrounding the balance between protecting children and empowering them to learn and explore (Macenaite, 2017; Data Protection Working Party, 2009; Montgomery & Chester, 2015). Data is being collected about children and young people through what is called a 'data footprint' – all the data that is collected about an individual when they use online services (Kadho Inc., 2018; Lieber, 2018; Gibbs, 2015; Lupton & Williamson, 2017; Taylor & Michael, 2017; Harris, 2017). This footprint can be used to profile children and young people as well as to personalise ads and products, among other harmful uses. Another topic that is being widely discussed is the potential insufficiency of traditional forms of informed consent (Berman & Albright, 2017). The frequency with which parents and guardians have to sign consent

forms has given way to 'consent fatigue', in which details outlined in the consent form may be overlooked due to the high volume of consent forms present, along with the fact that parents and guardians may not be in the position to fully understand the best interests of the child (Macenaite, 2017). Furthermore, there is the overarching question of individualised notions of consent versus the average child dilemma. Should the age of consent to access certain online services be generalised, as it currently stands in GDPR, or should it be individualised to cater to different levels of maturity, development, and the unique needs of individual children?

One of the largest challenges in this field is the fact that often services are not designed with children in mind, but they are accessed by children (Barassi, 2018; Howley, 2019). For example, when a 10-year-old child asked Amazon Alexa for a challenge to do, Alexa responded with a challenge that placed the child's well-being and safety at risk (Segal, 2021). In fact, it was later found that Alexa pulled this so-called 'challenge' from a website in which parents were warning other parents about letting their children do an activity such as this (Segal, 2021). This instance exemplifies the possible harms that can occur when services are not designed with children in mind but are accessible to them.

The ongoing dialogue on children's rights as they relate to AI should be much more than an analysis of privacy concerns. While privacy is an important consideration, the best interests of the child must be considered. This is precisely why children and young people should be engaged on topics of the design, development, and deployment of AI systems that use their data. Children and young people have unique needs and considerations, and these should be not only taken seriously but incorporated into ongoing and future dialogues on this topic.

Introducing AI ethics

To explore how children can and should be involved in these processes, we must first better understand the landscape that enables an analysis of the ethical and social implications that AI technologies may have on society. This field is called Artificial Intelligence (AI) ethics. AI ethics is a growing field of research, which aims to mitigate the possible negative impacts of the uses of these technologies while maximising the value and benefits that AI can bring. It also aims to engage community members, policymakers, and AI developers to consider the effects that AI technologies may have.

There are a wide variety of ethical concerns expressed regarding AI and its role in society. Some of these concerns relate to the ways in which AI works, or how it has been developed, for example, whether AI has been trained on biased or incomplete data, which might lead to it reproducing or exacerbating inequalities in its outcomes. Other concerns relate to the impacts that AI has on society, for example, through producing unfair outcomes or changing the ways that services are delivered and accessed, leading to transformative impacts on society. Mittelstadt et al. (2016) labelled these broad categories as “epistemic” and “normative” concerns. While the two are interlinked, epistemic concerns draw attention to potential shortcomings in how AI is designed and developed, while normative concerns focus on the impacts AI has on society. AI ethics engages with both sets of concerns and notes that the ethical challenges associated with AI are interwoven with broader, long-standing social, political, and cultural factors (Aitken et al., 2021). AI ethics requires a combination of technical and social approaches that take account of the social, cultural, political, and economic dimensions of data and AI, and the ways in which these dimensions have shaped how AI is designed, developed, and deployed as well as the impacts it has. This entails broader consideration of the role these technologies play

in society and the conditions under which they may be appropriate and acceptable (Aitken et al., 2021).

Importantly, ethics is not the same as legal compliance and there may be significant differences between what is legally permissible and what is ethically acceptable. Indeed, in many instances, ethics requires going substantially beyond legal requirements. While laws and regulation set out what we must or must not do (e.g., in terms of data protection, fair processing of data, or safeguarding of children), ethics grapples with the tricky questions of what we should or should not do (e.g., in what contexts or for what purposes should an AI system be deployed? How should the benefits of technologies be equitably distributed? What are the reasonable expectations users should have around privacy and consent?). Ethical questions typically do not have straightforward answers or clear-cut solutions, rather they require nuanced consideration and engagement with diverse perspectives to ensure that approaches taken align with societal values and expectations (Aitken et al., 2021).

An overview of AI ethics: principles and concerns

Given the tricky nature of ethical considerations, ethical approaches are typically guided by principles rather than fixed rules, and, as the field of AI ethics has grown, a proliferation of principles and guidance have emerged to attempt to address these tricky questions and guide ethical practice (Aitken et al., 2020). These principles have been developed and adopted by a range of organisations including research institutes, policy bodies, and tech companies of all sizes. While this can be taken as an illustration of the significant interest – and investment – in this field, it has equally been criticised as enabling organisations to engage in ‘ethics shopping’ – selecting the set of principles that most closely aligns with their current practices, or which do not require them

to make substantial changes (Floridi, 2021). This is closely related to criticisms of ‘ethics washing’, which are often levied at organisations that make statements about their ethical commitments without taking meaningful actions or enforcing ethical practices (Floridi, 2021).

While there is a proliferation of sets of principles and guidance relating to AI ethics, there are generally common themes within these. Fjeld et al. (2020) carried out a review of existing sets of principles relating to AI, from a wide range of international organisations, and identified eight main themes that consistently emerged within these:

- Privacy
- Accountability
- Safety and security
- Transparency and explainability
- Fairness and non-discrimination
- Human control of technology
- Professional responsibility
- Promotion of human values

These principles highlight the relevance of both technical and social methods to underpin ethical approaches to AI.

In combination with principles, ethics also requires reflection on the values that underpin the innovation and deployment of technologies. At the Alan Turing Institute, we have produced a guidance document entitled *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector* (Leslie, 2019). This guidance helps to lay the foundation for key principles related to AI ethics. There are four values that support, underwrite, and motivate responsible innovation, referred to as the SUM values; these were created to help researchers think about the possible impacts that using AI could have on society. This is also referred to as determining whether the use of AI is ‘ethically permissible’. The four SUM values are: respect, connect, care, and protect.

Respect the dignity of individual persons
Connect with each other sincerely, openly, and inclusively
Care for the well-being of each and all
Protect the priorities of social values, justice, and the public interest

In order for an AI system to be ‘ethically permissible’, it is important that we consider how each of these four values are met, so that our uses of AI do not produce negative and harmful effects. Some additional considerations that fall under these four SUM values are things like ensuring everyone is free to make their own decisions about their own lives, making certain that diversity, participation, and inclusion are prioritised throughout the entire project, and thinking critically about how the use of AI could empower and advance the well-being of as many people as possible.

While principles are helpful to guide ethical practice, ascertaining how to maximise the benefits of AI and identifying the varied and unequal potential negative impacts of the technology requires engaging with diverse views and experiences to fully understand and anticipate the impacts of AI on society and to ensure that the ways in which it is developed and deployed reflect societal values. In particular, given the well-documented potential for AI to have inequitable impacts across society, it is important to engage and incorporate the views and interests of the most vulnerable groups. Children and young people are one such group who have so far been underrepresented in discussions of AI ethics.

As the field of AI ethics continues to expand, it is necessary and critical that the voices of children and young people are encouraged and heard. These voices are a critical piece of AI ethics work going forward. Next, we will explain why.

Why does this matter for children and young people?

While there are positive examples of AI being used to help better deliver public services and advance the well-being of individuals, there is unfortunately no shortage of examples of where the use of AI technologies has caused harm to people – including children and young people.

In 2020, due to the COVID-19 pandemic, all secondary education examinations were cancelled in the UK. In the absence of exam results, a solution was needed. While there were predicted grades available from teachers, there was concern that these may lead to inflated results due to over-optimistic or unrealistic estimations. Therefore Ofqual, the UK's Office of Qualifications and Examinations Regulation, decided to produce an algorithm that would determine the qualification grades for each student for that year. It was intended to moderate and standardise teachers' predictions of students' grades (Tennison, 2020). However, the algorithm resulted in skewed exam scores, and a pattern was detected where students from less-privileged schools were more likely to have their exam results downgraded, while those from private schools were more likely to receive the estimated grades given by their teachers (Bedingfield, 2020). This was in part because exam results for each school were moderated to reflect previous attainment levels for each school, and also because the size of the cohort played a significant role in the model's output – in schools with small class sizes (predominantly private schools), the algorithm could not be relied upon to moderate the results to the same extent as for larger cohorts (Bedingfield, 2020). There was a public outcry as it was discovered that higher exam scores were highly related to privately funded independent schools; thus, students from state schools were penalised. Ultimately, while the algorithm was intended

to address potential unfairness of relying on estimated grades, it, in fact, exacerbated existing inequalities in society leading to unfair outcomes. Following backlash from students and legal action on behalf of advocacy organisations, exam scores were reissued based on unmoderated teacher predictions. This example illustrates how algorithms that do not fully consider ethical and social implications can cause significant harm and discriminatory outcomes.

Another example in which an AI technology has caused harm is through AI-assisted chatbots. In 2018, the BBC conducted a study to test the effectiveness of chatbots in a mental health setting (White, 2018). After testing two chatbots, the researchers concluded that the applications failed to "properly handle children's reports of sexual abuse" even though this chatbot technology was designed with children in mind.

Unfortunately, there are countless examples across many sectors of ways in which AI technologies have caused harm, especially to children and young people. Children and young people have a unique set of needs, and it is important to note that if developed ethically and responsibly and with children's voices included and listened to, AI technologies could provide beneficial outcomes. For example, AI technologies have immense potential to improve the provision of public services in a variety of settings, such as education.

Within the education sector, there are several examples of how AI could be used to better support children, parents, guardians, and teachers. The use of translation tools to expand access to education for students across the globe is one way in which AI systems could provide benefits. AI systems can provide real-time translations into different languages as well as provide increased accessibility to the services for those with visual or hearing impairments, so that universal access to education is expanded.

There is also the potential for AI to be used to provide curated tutoring lessons for students' specific learning styles and ensure that they are not struggling with lessons. These are a few examples of how AI could be harnessed to improve the quality of education. However, in order to realise benefits such as these, AI technologies must account for children's needs and interests, which are informed by not only the potential risks but ethical principles – such as those outlined above. One way in which AI ethics principles have been framed with the unique needs of children placed at the forefront is a developing area of research called 'child-centred AI'.

Child-centred AI

Child-centred AI ensures that children are involved throughout all stages of the AI lifecycle in a meaningful and worthwhile way. A summary of the main components of child-centred AI can be found below:

- Helping children to make informed choices about their interactions with and uses of AI
- Enabling children and young people to play a role in discussions shaping future AI practices
- Ensuring the next generation of developers and policymakers are equipped with an understanding of the ethical considerations relating to AI and its uses
- Ensuring ethical mindsets of future developers and members of the tech industry

The United Nations Children's Fund (UNICEF) has been working on the topic of child-centred AI. In 2020, UNICEF and the Government of Finland co-authored a draft policy guidance entitled **Policy Guidance on AI for Children**¹ (UNICEF, 2020). The draft contains an introduction to what is meant by the term AI and includes descriptions of the key opportunities and risks AI poses in the context of children's rights. UNICEF's nine requirements for child-centric AI are at the basis

of this developing field of research. These are:

1. Support children's development and well-being
2. Ensure inclusion of and for children
3. Prioritize fairness and non-discrimination for children
4. Protect children's data and privacy
5. Ensure safety for children
6. Provide transparency, explainability, and accountability for children
7. Empower governments and businesses with knowledge of AI and children's rights
8. Prepare children for present and future developments in AI
9. Create an enabling environment

Putting child-centred AI into practice

Our team at the Alan Turing Institute was invited to test UNICEF's draft policy guidance and share our findings with the public about what worked and what did not. The goal of organisations participating in this programme was to improve child-centred AI moving forward. We interviewed 14 public sector organisations across the UK to gain perspectives on how they think about developing child-centred AI applications, their opinions on the UNICEF guidance and other data protection regulations, and how they wish to see children, young people, and parents and guardians involved in the design, development, and deployment of AI technologies that use their data.

Our main findings are summarised here and are discussed further in our full-length case study (Pauwels et al., 2021):

- Public sector organisations believe there are low rates of data literacy amongst the public.
- There is an overall lack of understanding and clarity surrounding the implementation of GDPR principles.
- There are many guidance documents being drafted on the topic of children's rights and

¹ <https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf>

AI, and organisations are unsure which to use moving forward. Organisations wished to see synergies formed between existing and upcoming guidance documents.

- There is a desire to make the UNICEF *Policy Guidance on AI for Children* more actionable, to include more specific recommendations by sector, and to ensure the guidance is delivered in an age-accessible manner.
- Public sector organisations want to engage children and young people, but they stated that they did not know the best way to do this.

The findings from these interviews revealed public sector stakeholders' commitments to protecting children's rights and their enthusiasm to engage children in discussions relating to AI, but they also revealed many challenges associated with doing so. To illustrate this point, one interviewee stated:

There are lots and lots of ways and metrics that can be used to prove 'Ensure the inclusion of and for children' has happened without those children in the room actually being informed of what's going on. I'm not just talking about informed consent. I mean being fully appraised [sic] of the process and fully understanding.

It is clear that to address these challenges, children and young people must be involved in decision-making about the ways that AI is used in the public sector now and in the future. Our findings demonstrated that while public sector organisations wish to engage with children on these topics, they are not sure how to go about this in a meaningful way. In the next section, we will explore potential approaches to engaging children around AI ethics.

Approaches to engaging children with AI ethics

There are a number of reasons why an organisation developing or deploying AI might be motivated to engage with children or young people. These reasons in turn reflect different underpinning rationales, which can be normative, instrumental, and/or substantive (Fiorino, 1990; Wilsdon & Willis, 2004). First, a normative rationale leads to moral positions that suggest that if an organisation is developing or deploying an AI system that might impact on children they should engage with children as 'it's the right thing to do' (Wilsdon & Willis, 2004). Second, more practically minded approaches follow instrumental rationales, which view efforts to engage children as a means to achieve an organisation's own objectives (Wilsdon & Willis, 2004). Instrumental rationales might lead to a variety of potential approaches, including: efforts to build and maintain public trust in order to attract and retain customers; adopting ethical and transparent approaches to business to anticipate and respond to regulatory and policy developments; or efforts to demonstrate an ethical brand. However, following a purely instrumental rationale can lead to approaches that pay 'lip service' to public concerns through enacting purely cosmetic forms of engagement without genuine intentions to address concerns or reflect public values in an organisation's operation.

A final set of motivations are underpinned by substantive rationales that regard engagement as being aimed at creating wider positive outcomes across society.

From this point of view, citizens are seen as subjects, not objects, of the process. They work actively to shape decisions, rather than having their views canvassed by other actors to inform decisions that are then taken.
(Wilsdon & Willis, 2004, p. 39).

Following this approach, engagement with children offers opportunities to 'do things better' and maximise benefits not only for the organisation concerned but also for children and wider society. This might lead to AI being developed in ways that not only reduce risks or potential harms of technologies, but that are also more appropriate and beneficial for children and young people. Here it is important to emphasise that engagement is not simply about avoiding or mitigating potential negative impacts but equally about maximising the benefits of AI.

When we think about these different rationales for engagement within educational contexts, they also lead to different approaches and priorities. An instrumental rationale would tend to emphasise more direct goals (e.g., increasing skills and knowledge relating to AI in order to prepare children for future careers in AI), while substantive rationales are more likely to emphasise indirect and less quantifiable outcomes. For example, a substantive rationale might underpin approaches to engagement that seek to engage children with discussions of AI ethics in order to enable them to understand the role and impact AI plays in their lives and their society and to equip them with the skills and understandings to be able to critique the ways that AI is designed, developed, and deployed. The main difference here is that rather than focusing on teaching children to equip them with skills to be the future AI workforce, we are also aiming to equip them with skills to be the future – and current – critical public, which is needed to hold AI systems and their developers to account.

The benefits of these substantive approaches are manifold. Firstly, children benefit from better understanding the role of AI in their lives (now and in the future) and by being able to critically engage with AI and make informed choices about this. Secondly, there are substantial benefits for the development and deployment of AI, since children's views and values need to be included and reflected in order to inform

ethical practice. Quite simply, AI systems which impact – or have the potential to impact – children cannot be said to be developed or deployed ethically if children's experiences and perspectives have not been reflected in the design and development processes. Thirdly, there are wider benefits for society through engaging with diverse stakeholders in relation to AI policy as well as in the design, development, and deployment of AI. Reflecting public values and interests in all these processes is essential to establish a social licence for AI (Aitken et al., 2020), which ensures that uses of AI reflect societal values and expectations.

Importantly, these approaches to engaging children with AI require going beyond one-way forms of communication and instead require engaging in dialogue with children. Previous studies in public engagement with science and technology have demonstrated the limitations of approaches aimed at gaining public trust through improving public understanding. Such approaches treat members of the public as "passive recipients of scientific knowledge" (Cunningham-Burley, 2006, p. 206), overlooking how members of the public critically assess, deconstruct, and evaluate claims to scientific knowledge in line with their own ideologies, experiences, and the contexts in which the information is received (Hagendijk & Irwin, 2006). Demonstrating technical competence or communicating the robustness of technical responses to ethical challenges will not automatically lead to public trust and support. Rather, technical approaches need to be combined with social responses that build relationships of trust through which claims to technical competence will be evaluated (Aitken et al., 2020). While scientific and technical expertise is important, "such expertise cannot resolve the moral and political aspects of policy-making" (Elstub et al., 2021) or ethical considerations relating to AI. As such, engagement and deliberation can play a role in establishing the trustworthiness of science and technology

through efforts to address and reflect public values (Aitken et al., 2016; Wynne, 2006). Such deliberations do not require a detailed technical understanding of AI technologies; an understanding of the contexts in which technologies will be applied and the lived experience of communities that may be impacted are valuable forms of expertise and knowledge within these deliberative processes (Aitken et al., 2021).

Recognising the importance of dialogue and deliberation, approaches to engaging children with AI should start by asking questions around what children currently know, and what they want to know. What are their concerns, interests, or priorities? What are the important issues in their lives to which AI may relate, or for which AI might have a positive or negative impact? A child-led approach to engagement, which does not begin with assumptions of what children already know, or what they *should* know about AI, is likely to be more fruitful in leading to discussions that can underpin ongoing engagement with AI ethics.

Conclusions and looking ahead

This paper has introduced the context of children's rights and AI, AI ethics principles, and the why and how of engagements with children surrounding the development of AI technologies that use their data. We have introduced principles that underpin the field of AI ethics and explained how these principles must be combined with meaningful participatory engagement to ensure the unique needs of children are met. UNICEF's nine child-centric principles have laid the foundation for child-centred AI, which places children's voices at the centre and ensures that the voices of children are taken on board throughout the entire AI lifecycle. Child-centred AI does not exist solely for harm reduction, but also to create new and beneficial approaches to the design, development, and deployment of AI

technologies that involve children's data.

Child-centred AI and participatory engagement with children provide benefits for children by equipping them with new understandings of the impacts that these technologies can have on their lives as well as the ability to critically engage with AI on a day-to-day basis. There are also possible benefits to the field of AI by providing a landscape that considers children's unique set of needs and circumstances, leading to better, more appropriate, sustainable, and beneficial use of technology.

While we advocate for expanding the focus of children's engagement with AI beyond the development of skills for the future AI workforce, it is also vital that those children who do later choose to follow careers in, or with, AI enter those careers with a sound understanding of the importance of ethics in AI and an appreciation of both the tremendous potential benefits of AI but also the risks. The future AI workforce needs a diverse mix of skills and expertise encompassing technical, social, ethical, legal, and policy dimensions. Engagement with children relating to AI must aim at addressing this interdisciplinarity and broad relevance of AI, recognising that AI is not purely a technical or scientific subject but one that touches on all aspects of our lives, and about which we should all have a voice.

Ultimately, realising the benefits of AI will require an engaged and critical public whose voices and experiences are taken on board in design, development, and deployment processes. Children and young people have so far been underrepresented in discussions of AI and AI ethics, but it is vital that their views and interests are taken on board to inform future approaches. This will be an important area of research and practice in the coming years.

In the Ethics Theme at the Alan Turing Institute, we are embarking on an exciting project to

explore this further. Working in collaboration with the Scottish Children's Parliament and the Scottish AI Alliance, we are engaging children across Scotland in discussions about AI. This research will explore what children currently know about AI and what they want to know about AI and how it is used; how children feel about the ways in which AI may be used to inform decisions about their lives (e.g., access

to services); how they would like AI to be used in the future; what they think are the limits to how AI should be used in the future; and how children want to be involved in decision-making about future uses of AI.

We are excited about this next chapter in our research and to place children's voices at the heart of AI ethics.

References

- Aitken, M., Ng, M., Horsfall, D., Coopamootoo, K.P., van Moorsel, A., & Elliott, K. (2021). In pursuit of socially-minded data-intensive innovation in banking: A focus group study of public expectations of digital innovation in banking. *Technology in Society*, 66(7), Article 101666. <https://doi.org/10.1016/j.techsoc.2021.101666>
- Aitken, M., Toreini, E., Carmichael, P., Coopamootoo, K., Elliott, K., & van Moorsel, A. (2020). Establishing a social licence for Financial Technology: Reflections on the role of the private sector in pursuing ethical data practices. *Big Data & Society*, 7(1), Article 2053951720908892. <https://doi.org/10.1177/2053951720908892>
- Aitken, M., Cunningham-Burley, S., & Pagliari, C. (2016). Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. *Science and Public Policy*, 43(5), 713–723. <https://doi.org/10.1093/scipol/scv075>
- Barassi, V. (2018, September 18). *Home life data and children's privacy*. Child Data Citizen. Retrieved February 27, 2022, from <http://childdatacitizen.com/cdc/wp-content/uploads/2018/09/'HOME-LIFE-DATA'-AND-CHILDREN'S-PRIVACY-1.pdf>
- Bedingfield, W. (2020, August 19). *Everything that went wrong with the botched A-Levels algorithm*. Wired. <https://www.wired.co.uk/article/alevel-exam-algorithm>
- Berman, G., & Albright, K. (2017). Children and the data cycle. *Innocenti Working Papers*. <https://doi.org/10.18356/7d33bc21-en>
- Cunningham-Burley, S. (2006). Public knowledge and public trust. *Public Health Genomics*, 9(3), 204–210. <https://doi.org/10.1159/000092658>
- Data Protection Working Party (2009, February 11). *Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools)*. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2009/wp160_en.pdf
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it ok if I eat you?": Initial explorations in child-agent interaction. In *Proceedings of the 2017 conference on interaction design and children* (pp. 595–600). Association for Computing Machinery. <https://doi.org/10.1145/3078072.3084330>
- Elstub, S., Thompson, R., Escobar, O., Hollinghurst, J., Grimes, D., Aitken, M., McKeon, A., Jones, K., Waud, A., & Sethi, N. (2021). The resilience of pandemic digital deliberation: An analysis of online synchronous forums. *Javnost-The Public*, 28(3), 237–255. <https://doi.org/10.1080/13183222.2021.1969616>
- Fiorino, D.J. (1990). Citizen participation and environmental risk: A survey of institutional mechanisms. *Science, Technology, & Human Values*, 15(2), 226–243. <https://doi.org/10.1177/016224399001500204>
- Fjeld, J., Hilligoss, H., Achten, N., Daniel, M.L., Feldman, J., & Kagay, S. (2020). *Principled artificial intelligence: A map of ethical and rights-based approaches*. Berkman Klein Centre for Internet & Society. https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y
- Floridi, L. (2021). Translating principles into practices of digital ethics: Five risks of being unethical. In: Floridi, L. (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence* (pp. 81–90). Springer. https://doi.org/10.1007/978-3-030-81907-1_6
- Gibbs, S. (2015, March 13). *Privacy fears over 'smart' Barbie that can listen to your kids*. The Guardian. <https://www.theguardian.com/technology/2015/mar/13/smart-barbie-that-can-listen-to-your-kids-privacy-fears-mattel>
- Hagendijk, R., & Irwin, A. (2006). Public deliberation and governance: Engaging with science and technology in contemporary Europe. *Minerva*, 44(2), 167–184. <https://doi.org/10.1007/s11024-006-0012-x>
- Harris, R. (2017, December 27). *72M data points collected on children in spite of COPPA*. App Developer Magazine. <https://appdevelopermagazine.com/72m-data-points-collected-on-children-in-spite-of-coppa/>
- Howley, R. (2019). *The effects of artificial intelligence on the youth*. (Publication No. 22622916) [Master's thesis, Utica College]. ProQuest Dissertations Publishing. <https://www.proquest.com/openview/d84a86cf41d62229ab84860328e20c92/1?pq-origsite=gscholar&cbl=18750&diss=y>
- Kadho Inc. (2018, May 18). *Kid's privacy in the age of artificial intelligence*. Medium. <https://medium.com/@KadhoInc/kids-privacy-in-the-age-of-artificial-intelligence-4c113e642124>
- Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>
- Lieber, C. (2018, December 5). *Big tech has your kid's data – and you probably gave it to them*. Vox. <https://www.vox.com/the-goods/2018/12/5/18128066/children-data-surveillance-amazon-facebook-google-apple>
- Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, 19(5), 780–794. <http://journals.sagepub.com/doi/full/10.1177/1461444816686328>
- Macenaite, M. (2017). From universal towards child-specific protection of the right to privacy online: Dilemmas in the EU General Data Protection Regulation. *New Media & Society*, 19(5), 765–779. <https://doi.org/10.1177/1461444816686327>
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), Article 2053951716679679. <https://doi.org/10.1177/2053951716679679>
- Montgomery, K.C., & Chester, J. (2015). Data Protection for youth in the digital age: Developing a rights-based global framework. *European Data*

Protection Law Review, 1(4), 277–291.
<https://doi.org/10.21552/edpl/2015/4/6>

Pauwels, E., Briggs, M., Aitken, M., & Leslie, D. (2021). *Case study: Understanding AI ethics and safety for children: The Alan Turing Institute*. UNICEF. <https://www.unicef.org/globalinsight/media/2321/file>

Segal, E. (2021). *A crisis is averted after Amazon's Alexa tells child about lethal viral challenge*. Forbes. <https://www.forbes.com/sites/edwardsegal/2021/12/28/a-crisis-is-averted-after-amazons-alexa-tells-child-about-lethal-viral-challenge/?sh=4c32eca139b6>

Taylor, E., & Michael, K. (2017, June 29). *Smart toys that are the stuff of nightmares*. IEEE Technology and

Society. Retrieved February 27, 2022, from <https://technologyandsociety.org/smart-toys-that-are-the-stuff-of-nightmares/>

Tennison, J. (2020, August 16). *How does Ofqual's grading algorithm work?* RPubS. <https://rpubs.com/JeniT/ofqual-algorithm>

UNICEF (2020, September). *Policy guidance on AI for children*. UNICEF and the Ministry of Foreign Affairs of Finland. <https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf>

White, G. (2018, December 11). *Child advice chatbots fail to spot sexual abuse*. BBC News. <https://www.bbc.co.uk/news/technology-46507900>

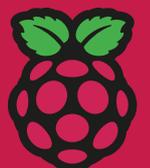
Williams, R., Park, H.W., & Breazeal, C. (2019). *A is for Artificial Intelligence: The impact of Artificial Intelligence activities on young children's perceptions of robots*. In *CHI Conference on Human Factors in Computing Systems Proceedings* (pp. 1–11). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300677>

Wilsdon, J., & Willis, R. (2004). *See-through science: Why public engagement needs to move upstream*. Demos.

Wynne, B. (2006). *Public engagement as a means of restoring public trust in science – Hitting the notes, but missing the music?* *Community Genetics*, 9(3), 211–220. <https://doi.org/10.1159/000092>

How can we make AI education a priority without using scare tactics?

Rose Luckin (University College London)



How can we make AI education a priority without using scare tactics?

Rose Luckin (University College London)

Introduction

I was listening to a Financial Times Tech Tonic podcast (Kynge, 2022) the other day, entitled *US–China Tech Race: brave new world* and it reminded me of an experience that I had during one of the last face-to-face events that I attended before the pandemic.

I was at AI Everything in Dubai: a conference and an exhibition. I had only just arrived and decided to start in the main conference hall to try and get a sense of the event. I listened to some of the keynote talks and found myself mesmerised with horror as I listened to the head of security for Huawei technology. He was telling a captivated audience how his company had developed amazing face recognition technology that was able to take the ageing process and its impact on the face into account. This meant that even if the AI technology had been trained with a picture of my 30-year-old face, when I hit my 50th birthday and beyond, the technology would still be able to analyse my face and recognise my identity. I had to admit that this was impressive technology. However, I have developed the habit of finding highly impressive technology developments that use AI disconcerting as well as exciting.

On this occasion, my unsettled feelings grew. The Huawei executive couched his narrative about the development of this technology in a story about the heartbreak of missing children and how this face recognition technology had been developed with the purpose of reuniting

parents with children who had gone missing. The talk was accompanied by high-definition video footage of tearful parents being reunited with children whom they had not seen for many years. Their reunion was all thanks to Huawei's amazing face recognition technology. The technology had brought together some 169 missing children with their parents and was going on to help reunite more and more families. The speaker himself was dewy-eyed, as was the video footage. I was open mouthed, my jaw having dropped in disbelief that such a flimsy fig leaf of social responsibility could possibly be used to make a potentially deeply worrying technology seem like the best thing to have been invented this decade.

I turned around to look at the rest of the audience and their faces, assuming that they, like me, would be 'gobsmacked', but people just seemed to be taking it in, believing it, and seeing the good. Had I just become a cynical academic all too ready to criticise? I really didn't think I had, but I was left with a bad feeling.

So why was I reminded of this incident from 2019 while I was listening to the Tech Tonic podcast in 2022? And what does this have to do with the talk that I gave as part of the Raspberry Pi Foundation seminar series?

Let me answer the first question to start with. The Tech Tonic podcast was recounting a sad tale that has come to be known as The Countryman Case for reasons that will become apparent in just a moment. To cut a long (and fascinating) story short: in 2014, a young man, called Luka, was killed in a hit and run accident

on Serbia's Branko Bridge in Belgrade in the middle of the night. There were no witnesses. The police were not actively trying to solve the case, but Luka's father would not give up; he spent days standing on the bridge protesting and eventually the police started to investigate. Some grainy CCTV footage of the accident showed that the car that had killed Luka was a Mini Countryman, hence the name of the case. The police were unable to find the car, but they did manage to identify the driver. The driver could not be found anywhere in Serbia and they circulated his photograph to other countries and cities, including Beijing in China. Amazingly, less than three days later, the Chinese found the driver and the Chinese authorities immediately deported him back to Serbia.

The police in Beijing were able to find the criminal driver so quickly because of the advanced facial recognition technology used across the city. For several years now, China has been building a huge surveillance system with webs of cameras across the country, all of which have facial recognition technology, and that is why they could find the criminal driver in the Countryman Case so quickly.

Today, Branko Bridge in Belgrade is also monitored by cutting-edge Chinese surveillance cameras purchased from that same technology company whose representative I heard speaking at the AI Everything conference in 2019: Huawei. Of course, Huawei are not the only company who are making advanced facial recognition technologies, but they are one of the leading players in this field. It is also true that Serbia is not the only country that has made large purchases of these technologies from Chinese companies: according to the Tech Tonic podcast, 64 countries, from Africa to the Middle East and Europe have made such investments.

When countries invest in this facial recognition surveillance technology, they are enabling their

police forces to stand a much better chance of capturing criminals, such as the man who killed Luka on the Branko Bridge, and they are probably enabling the speedier recovery of missing children and their abductors if they are still in the same country or in another country that uses this high-spec facial recognition technology. However, these countries are also equipping themselves either intentionally or unintentionally with the tools to enable them to become Big Brother style state surveillance entities.

Returning to the situation in Serbia for just a moment, it is good to hear that whilst the majority of the population believe that China is a beneficial trading partner for their country, there have been significant protests against the use of the facial recognition capacities of the cameras that have been purchased. As a result, the cameras installed across the country, including those overlooking the Branko Bridge, have the capacity to conduct facial recognition, but that capacity has not been turned on. One of the important learnings that we can take away from this story is that when people understand the implications of AI, both good and bad, then they can make more informed decisions about how they want it to be used. As AI becomes increasingly ubiquitous, the need for an educated population becomes even more important if democracy is to be upheld.

People and AI

The reason I started this chapter with these two recollections is that I am increasingly worried by the number of people who tell me that it is not important for teachers, parents, students, and the public in general to understand AI. I am told: "They just need to know how to use it". This is a dangerous situation, and it requires urgent attention. Firstly, I don't believe you can really know how to use AI without understanding something about what AI is and how it operates.

I do not mean that everybody needs to know how to programme and build an AI system, or that they need to understand the complex mathematics within a neural network. What I mean is that people need to understand what AI can do and what AI cannot do, along with the basics about how traditional AI, often referred to as Good Old-Fashioned AI (GOFAI), and modern machine learning AI operate in a non-technical way, as well as how they are different.

The case of the facial recognition technology being portrayed as purely beneficial in the stories at the start of this chapter offers an example of why people need to understand more about AI if they are to protect themselves and to appreciate the genuine risks, so that they are not driven away from beneficial AI by people who are scaremongering. People need to be able to tell truth from fiction and to make informed decisions. To make informed decisions, they need to be informed, or to put it in another way, they need to understand. Teachers need to understand because they can then help their students to understand. In the same way that teachers understand how to teach people to read and to write, they need to be able to teach people to be AI literate. And yet, it is very difficult to engage teachers in learning about AI. It is hard to persuade people who already have 101 things to do that this additional thing should be prioritised, particularly when it is not part of the curriculum or assessment framework.

I firmly believe that AI has a great deal of beneficial potential, way beyond the AI systems currently in use within the classrooms of the western world. I also believe that the vast majority of those companies who are selling AI technology into educational institutions are not posing risks like the facial recognition software I discussed at the start of the chapter. However, there are instances where systems that describe themselves as using AI provide scant information about the AI they use and how it delivers benefits. There are even examples where systems and companies that are described as

using AI do not actually use any AI (for a range of views, see: Narayanan, 2019; Marr, 2018; Hao, 2019; Ram, 2019).

Education, educators, and AI

There are also a significant number of examples of AI being used in invasive and worrying ways in education. For example, students being monitored every minute of their day in China (Xie, 2019), CCTV being used to track down students not wearing masks in the USA (Keierleber, 2022), and classrooms in China using brain-wave trackers to check if a student is concentrating (Wall Street Journal, 2019).

Did the decision makers who brought these AI systems into their establishments know what the consequences, both positive and negative, would be? Were the educators fully involved in the decision? Is this just happening outside Europe, or is it the tip of the iceberg of a worrying global trend?

For those who doubt the sophistication of the AI being developed for education, it is useful to look at research labs to see what studies are being conducted and what ideas are being pursued. Such work will certainly inform the future and will likely foreshadow what will become commercially scaled. For example, the increase in availability and affordability of wearable and remote sensing technologies enables the study of groups of students working together by enabling the capture and analysis of voice, facial expression, speech, and bodily movements. Voice and facial expression data has been used for the analysis and categorisation of discussions using AI in the form of Natural Language Processing (NLP) and machine learning (Stewart et al., 2021). Posture detection and facial recognition has been used to classify the participation states of learners using Bayesian modelling (Kasparova et al., 2020). When AI is combined with other innovative science and technologies, the possibilities

grow profoundly. For example, functional Near-Infrared Spectroscopy (fNIRS) can detect neural signals that indicate when a person is engaging in reflective thinking. Heart rate variability, blood pressure, temperature, and electrodermal activity levels, can all be sensed, collected, and analysed to look for signals of frustration and stress. Gestures and movement patterns can be analysed by using 3D and 2D video to assess student engagement in collaboration; gaze patterns can illustrate student attention, eye-tracking data can be analysed to reveal students' emotions, cognitive load, and focus. The possibilities and potential are unbounded. Sadly, so too are the risks.

When the **Institute for Ethical AI**¹ was created in 2018, it was because we were fearful that the benefits of this technology would be lost because an extremely negative and unethical event would occur due to the use of AI, and this would close down opportunities that could have positively transformed the lives of many people, particularly those who are disadvantaged. We produced a framework to help educational procurement ask the questions of themselves and of AI sellers that would help to ensure that the AI was beneficial. This is useful, but it is just one step in a much greater process through which educators must understand enough to make wise decisions about buying and using AI. There is also an important role for regulation, which must not be overlooked. But regulation will never be able to keep up with the developments in AI, and therefore *education is essential, urgent, and important*. We must educate the educators.

Engaging the educators

The challenge of engaging teachers in learning about AI is a tough one. I have recently published a book called *AI for School Teachers* (Luckin et al., 2022), which I wrote with an academic colleague and a headteacher. We wanted to write with a headteacher in order to try and ensure that we wrote in a manner that was relevant to

teachers. In addition, our headteacher co-author was not someone who was 'a techie.' She is someone who values technology and believes it is important, but also someone who is not particularly proficient with using technology, nor did she initially understand a great deal about AI. We believed that if we could help her to understand AI, then she would be able to help us to understand how to write about it for other teachers. We all enjoyed the writing process and went to great lengths to find convincing, authentic educational examples for all the different ways in which we discussed and explained AI.

For example, we used a very common activity: planning a school trip to explain how Good Old-Fashioned AI (GOF AI) could be used to develop an AI application to help with stepping teachers through all the processes and decision points involved. We included checklists. For example, we suggested that teachers thought about the educational challenges they were facing, because these should lead decisions about how AI should be adopted. We provided a set of questions to prompt teachers about the challenges they might be facing that might be addressed by AI:

EXHIBIT X: QUESTIONS TO HELP IDENTIFY CHALLENGES THAT A SCHOOL TEACHER MIGHT BE FACING

- Are the children in my class attending regularly?
- Have the children developed good relationships amongst their peers?
- Are the children in my class well nourished?
- Have I planned lessons that meet the expected learning outcomes?
- Does my marking and feedback effectively engage students such that they act to improve their learning?
- How do I evaluate parental perceptions of me as a teacher?
- How do I ensure that soft skills are developed by my students?
- How do I know that we are meeting the expected standards for teaching and learning?
- What should I do to close the gender gap in maths?

Figure 1. Questions to help teachers identify challenges (Luckin et al., 2022, p.19).

¹ <https://www.buckingham.ac.uk/research/research-in-applied-computing/the-institute-for-ethical-ai-in-education/>

And similarly, questions for headteachers and school leaders:

EXHIBIT Y: QUESTIONS TO HELP IDENTIFY CHALLENGES THAT HEADTEACHERS MIGHT BE FACING

How do I recruit and retain the right staff for my context?
 How do I source or develop CPD to meet the needs of my staff?
 How do I build ed tech capacity amongst my staff?
 How do I ensure my students with special education needs receive the provision they need?
 How do I ensure we meet our aspirational targets for core academic subjects?
 How do I know that I'm communicating in a timely manner with all stakeholders?
 How can I assess if my school development plan is fit for purpose?
 How can I see if all children across the school are making progress?
 How will I know that the school is getting a good return on the investments we are making?

Figure 2. Questions to help school leaders identify challenges (Luckin et al., 2022, p. 20).

We also set out a self-questioning process to help teachers and their leaders to decide which of their set of challenges is the best one to focus on first for thinking about AI (Luckin et al., 2022, pp. 20–23):

1. "Ask these of yourself, your colleagues, team, peers, managers or stakeholders, and use the answers to narrow your pool. What do you already know about this challenge? [Score 3 if you know a great deal, Score 2 if a modest amount, Score 1 if you don't know much, and Score 0 if you know nothing about this at all].
2. What kind of information is it possible for you to know that you don't know now? For example, if you are wanting to know more about the attainment gaps between different pupil groups, think hard about exactly what you could know about the pupils, their friends, family, context, etc. Or, perhaps you are concerned about bullying - there are different types of bullying in different degrees. For example, cyber, physical, name-calling, etc. It would be possible for you to explore the environmental conditions in the school that allow for these incidents to happen. [Score 3 if you are confident that you could know a great deal more, Score 2 if you believe that you could know a modest amount more, Score 1 if you are not sure that there is a great deal more that you could know, and Score 0 if you believe there is nothing more that you could know].
3. To what extent is the challenge you are facing controllable, and by whom? Are all systems and procedures understood clearly by all staff teaching and support? Are they audited, reported, and monitored? [Score 3 if the challenge is (a) controllable, (b) by someone at the school or within the school group, and (c) you do have all the systems in place to control the challenge; Score 2 if any two of (a), (b), and (c) are true; Score 1 if any one of (a), (b), and (c) is true; and Score 0 if none of (a), (b), or (c) is true]. For example, recruiting, training, and maintaining the best staff team. Any organisation only has limited control over the recruitment challenge, because whilst it can optimise all elements of the recruitment process that it adopts it cannot control how many people apply. Hopefully you have confidence that there are systems in place to help you optimise the elements of the recruitment process that are within your control, and AI can certainly help with that. However, the organisation cannot alter the number of people who are looking for the sort of employment that is on offer. Similarly, the school cannot control the pool of applicants that have the appropriate qualifications, skills, and expertise for the roles that need to be filled. [In this example, the score would be 2, because the whole of the recruitment process is not under your control, but you do have the systems in place to maximise the aspects of the process that are within your control].
4. What level of uncertainty is there? There may well be a level of uncertainty with a challenge based on incomplete reporting procedures

by staff and children, for example, or due to a challenge being surfaced through anecdotal evidence. [Score 3 if the level of uncertainty is negligible, Score 2 if there is a modest amount of uncertainty, Score 1 if there is a great deal of uncertainty, and Score 0 if there is no certainty at all].

5. Do you already have any data to help you understand this challenge or can you access data about this? [Score 3 if you have or can access a large amount of data from different sources, Score 2 if you have or can access a modest amount of data from different sources, Score 1 if you have or can access a very little data from any source, and Score 0 if you neither have, not have access to any data]. For example, you might have data derived from existing surveys, parent comments or complaints, behaviour logs and risk assessments.
6. Can you collect more data if you don't have enough data to help you understand this challenge and work out how best to tackle it? There are always opportunities to collect more data from students. [Score 3 if you can collect a large amount of relevant data, Score 2 if you can collect a modest amount of relevant data, Score 1 if you can only collect a small amount of relevant data, and Score 0 if you are unable to collect any new data at all].
7. How accurate can you be in your assessment of the challenge and your prediction about the best way to tackle it? [Score 3 if you can be very accurate, Score 2 if you can be quite accurate, and Score 1 if you can only be imprecise and therefore not very accurate at all, and Score 0 if you cannot be accurate at all]. For example, cyber bullying is a challenge that can be difficult to assess accurately, because it can occur outside of school grounds and systems.

8. Do you or your organisation have the appetite and capability to change to address this challenge? [Score 6 if the answer is "yes" and Score 0 if the answer is "no"]. If the answer is no, for whatever reason, it may not be a good investment of your time to be looking at ways AI can help you tackle the challenges in new ways.
9. Is the challenge AI compatible? [Score 3 if it is very AI compatible, Score 2 if it is modestly compatible, Score 1 if it is not very compatible, and Score 0 if it is completely incompatible]. This may be a difficult question for you to answer at the moment, but the section of this chapter entitled "Who has got the power, Artificial or Human Intelligence?" will help, as we hope will the rest of the book.
10. Finally, and most importantly, how important is solving this challenge to you or to your organisation? [Score 6 if it is crucial to solve this challenge, Score 4 if it is important to solve this challenge, Score 2 if it is quite important, and Score 0 if it is not important at all]."

And we found novel ways to communicate the intricacies of a machine learning algorithm, for example, through cookery, in this extract from the book (Luckin et al., 2022, pp. 72–73):

"I find it helpful to think about this situation as being a bit like cooking. There are lots of many types of cooking. We can bake, we can fry, we can broil, boil, or braise. We can grill, we can poach, we can smoke, sear, or sous vide. Just for a moment imagine that you are part of one of those TV shows where you are presented with a set of ingredients that are placed on a table and hidden under a cloth. You pull back the cloth to reveal the ingredients from which you must make something wonderful. Your instructions state that you are to make a dessert and that you must use all the ingredients in making your dessert.

Think about the ingredients as being a little bit like the data to which we want to apply AI. Back to the table. You have just whipped off the ingredients in our imaginary cooking show. You have eggs, and you have raspberries, plus there is cream and sugar.

What type of cooking method would be best applied to these ingredients? Frying is not an option, because whilst we could fry an egg, we have to use all the ingredients not just eggs. Grilling also looks unlikely to be suitable. Similarly, broiling or boiling is not really appropriate, but maybe baking could work for this.

The same type of situation exists when it comes to applying AI to our data “ingredients”; we need to decide what sort of AI could and should be applied. This decision is largely driven by the ingredients available and the challenge that we need to address, just as it is with cooking. There may be several options available to us for the same set of ingredients. Experience will help us to know which option to try first. Fortunately, unlike food, data can be subjected to multiple AI techniques that are appropriate to the type of data and the challenge being addressed.

Back to the cooking ingredients. We know we are required to solve the challenge of creating a dessert from the ingredients available. We also know that baking is likely to be the most appropriate cooking method to apply. The options available are now constrained by these parameters, but there are still options. Should we make raspberry pavlova or should we make raspberry soufflé? Which of these is going to best meet the requirements of the cooking show, the challenge? We decide on raspberry soufflé, because we have more experience of making this and therefore believe that a good result is more likely than with pavlova. Now this choice has been made, we know the method that we need to complete in order to produce the solution to our challenge: a dessert using the ingredients available to us.

The situation with data and applying AI is not so dissimilar. We have looked at our data (the ingredients) and the challenge (exploring the quality of teaching and learning when moving some provision online). We decide that the most suitable type of AI (cf. type of cooking) to be applied to these ingredients and this challenge is machine learning (cf. baking). Finally, we make a choice about the type of result we want to produce: finding patterns in the data for the teaching and learning interactions that have happened online and face to face (cf. pavlova or soufflé). We can therefore now also choose the method of machine learning that we are going to apply; we choose unsupervised machine learning.

Returning for a moment to the raspberry soufflé situation. We are now faced with needing to go through a set of preparations to be able to apply the baking process to the ingredients and produce a soufflé. First, we have to wash the raspberries. Then we have to crack the eggs and whisk them. And then we have to add the sugar into the whisked eggs. We also have to whip up the cream and add that to the beaten eggs and the sugar. Finally, we add the washed raspberries. Then we need to mix it all together in the bowl. We now have the soufflé mix and just need to put the mix into a dish, or a set of individual portion dishes, and we will be ready to apply the cooking method of baking to the prepared ingredients. As you can see, there is a lot of preparation. In fact, it may take longer to do all that preparation than it does to bake the soufflé, which is really quick to bake.

For our education data and educational challenge situation, we want to explore the extent to which we have maintained the quality of teaching and learning, as things have moved online. It is important to note that there are many ways in which we could analyse our data, many of which have nothing to do with AI, but the point here is to see what extra insights and understanding the use of AI techniques can bring to the kind of analysis that is normally done with educational data. In our example here, we also want to understand more about AI."

Next steps

However, in conversations with teachers as they read the book, I still feel that we have missed an important connection, a connection that would really motivate and mean that understanding AI would become a priority. But what is that missing connection? Should we try to scare teachers with more lurid examples than those with which I started this chapter? Examples of ways in which the suits of AI already being used in education could be misused, abused, and cause harm? Should we gaze at the next generation of AI that is likely to be appearing in the classroom and illustrate the benefits

and the risks in full and scary detail? I certainly believe we need something dramatic, but I don't feel comfortable with using scare tactics. Nor do I feel comfortable with the thought of teachers, students, parents, and the public being hoodwinked by smart salespeople who know how to tell a positive story and avoid catalysing any concern within their audience.

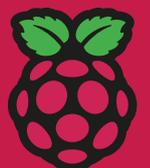
I am still experimenting to find other ways that will capture their attention, to make people sit up and get them to believe that understanding AI is vital. I have no smart answer to conclude the chapter, just a plea for more attention to be paid to finding effective ways to motivate educators to want to understand AI and the right tools for helping them to succeed.

References

- Hao, K. (2019). China has started a grand experiment in AI education. It could reshape how the world learns. *MIT Technology Review*, 123(1).
<https://fully-human.org/wp-content/uploads/2019/09/China-AI-experiment-education.pdf>
- Kasparova, A., Celiktutan, O., & Cukurova, M. (2020, October). Inferring student engagement in collaborative problem solving from visual cues. In *Companion Publication of the 2020 International Conference on Multimodal Interaction* (pp.177–181).
<https://doi.org/10.1145/3395035.3425961>
- Keierleber, M. (2022, March 30). 'Really alarming': the rise of smart cameras used to catch maskless students in US schools. *The Guardian*.
<https://www.theguardian.com/world/2022/mar/30/smart-cameras-us-schools-artificial-intelligence>
- Kynge, J. (Host). (2022, April 25). US-China Tech Race: brave new world [Audio podcast episode]. In *Tech Tonic*. Financial Times.
<https://www.ft.com/content/33ef2363-6f97-43dd-8b59-7fca025684cc>
- Luckin, R., George, K., & Cukurova, M. (2022). *AI for School Teachers*. CRC Press.
<https://doi.org/10.1201/9781003193173>
- Marr, B. (2018, July 25). How is AI used in education. Real world examples of today and a peek into the future. *Forbes*.
<https://www.forbes.com/sites/bernardmarr/2018/07/25/how-is-ai-used-in-education-real-world-examples-of-today-and-a-peek-into-the-future/?sh=70581967586e>
- Narayanan, A. (2019). *How to recognize AI snake oil* [PowerPoint slides]. Princeton University.
<https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf>
- Ram, A. (2019, March 5). Europe's AI start-ups often do not use AI, study finds. *Financial Times*.
<https://www.ft.com/content/21b19010-3e9f-11e9-b896-fe36ec32aece>
- Stewart, A.E.B., Keirn, Z., & D'Mello, S.K. (2021). Multimodal modeling of collaborative problem-solving facets in triads. *User Modeling and User-Adapted Interaction*, 31(4), 713–751.
<https://doi.org/10.1007/s11257-021-09290-y>
- Wall Street Journal. (2019, October 1). *How China is using Artificial Intelligence in classrooms* [Video]. YouTube.
<https://www.youtube.com/watch?v=JMLsHI8aV0g>
- Xie, E. (2019, September 16). Artificial intelligence is watching China's students but how well can it really see? *South China Morning Post*.
<https://www.scmp.com/news/china/politics/article/3027349/artificial-intelligence-watching-chinas-students-how-well-can>

Exploring the data-driven world: teaching AI and ML from a data- centric perspective

**Carsten Schulte, Yannik Fleischer, Lukas Höper,
Rolf Biehler, Daniel Frischemeier, Sven Hüsing,
and Susanne Podworny** (Paderborn University)



Exploring the data-driven world: teaching AI and ML from a data-centric perspective

Carsten Schulte, Yannik Fleischer, Lukas Höper, Rolf Biehler, Daniel Frischemeier, Sven Hüsing, and Susanne Podworny (Paderborn University)

Abstract

In our talk, we discussed the question of whether artificial intelligence (AI) and machine learning (ML) should be taught differently from other themes in the computer science curriculum, and if so, how to teach them. The tentative answer is that these topics require a paradigm shift for some teachers, and that this shift has to do with the changing role of algorithms, data, and the societal context. The talk presented three teaching examples from the beginning of secondary school (11-13 year old students) to illuminate the possible differences in teaching. The first example drew upon the Matchbox Computer and successors like the Sweet Learning Computer to teach the machine learning process, the second was about enactive teaching of decision trees, and the third was about analysing location data.

Introduction

Teaching artificial intelligence (AI) and machine learning (ML) from what we have called a 'data centric perspective' is an idea that originated from our project *ProDaBi* (*Project Data Science and Big Data at school*). We started with a symposium in 2017 to collect some ideas on the topic of data science at school and how to

incorporate it in school curricula. Following that symposium, the project started in 2018 with the aim to develop a curriculum for data science at schools, including AI.

We roughly oriented the work around the curricular spider web (see Figure 1); this helped us identify that questions to answer when designing a curriculum should be related to, for example, the content, suitable learning activities, and so on.

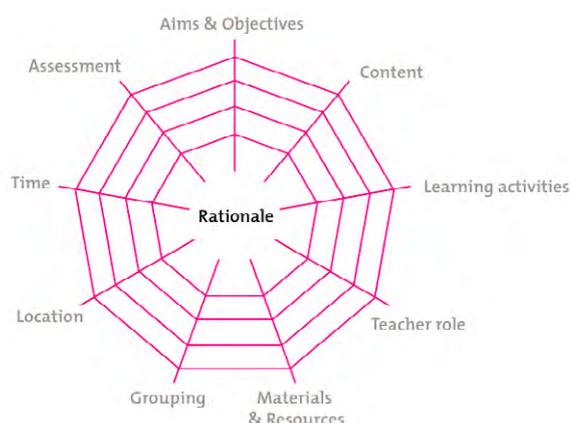


Figure 1. Curricular spider web. From Curriculum in development by Thijs, A., & Van Den Akker, J., 2009, p. 11, Netherlands Institute for Curriculum Development (SLO). Copyright: 2009, Netherlands Institute for Curriculum Development (SLO).

At first, we focused on developing materials and resources, and tried to find a place or time slot at school. We initially started with a special elective course at the end of secondary school, and from there gradually adapted resources and teaching ideas to also be useful in lower secondary grades. However, the focus of this presentation is not on the materials, but on the rationale, that is, the middle of the curricular spider web (see Figure 1). So, the question is why we have done it the way we have, and what is behind it. We will present this framework first, or at least some insights into the framework, and then make it more concrete with three examples.

Paradigm shift needed

We tried to answer the question how new and probably changing and complex topics like data science and AI can be included in a curriculum. The motivation to include these topics was – besides future job opportunities – the increasing number of AI and data-driven applications people can use in their daily lives. This is a typical issue for computer science, and one answer is rather popular, which we do not completely agree with. It's formulated as the slogan: "ideas not artefacts" (e.g., Wing, 2006), and is based on experiences as captured in this paper's title: *Computer science in English high schools: We lost the S, now the C is going* (Clark & Boyle, 2006), referring to a misconceived focus on (using) computers instead of teaching the science behind computers. We think it is important to not only use applications, but to also look under the hood. However, this doesn't just refer to ideas. Modern AI applications are mostly discussed in terms of the increased role and progress of ML, hence data-driven applications. And these can only be understood and evaluated by also taking into account the data they are processing, for example, for training their models – this is discussed and argued for in the paper called *Machine behaviour* by Rahwan et al. (2019). The

authors argue that these data-driven applications can only be understood by including their behaviour 'in the wild', so to speak.

The argument is twofold. First, data-driven applications rely on vast amounts of data; their performance thus can only be understood by knowing about this – and also using such applications unavoidably impacts the role of data and the need to collect data, often including, for example, user or interaction data. Secondly, such applications behave differently from traditional algorithms. A sorting algorithm also relies on data, but its internal mechanics are independent from the used data. It is a mathematical function transforming an unsorted input to a sorted output – and the algorithm itself (e.g., a bubble sort) is not being changed, no matter what data is fed into the sorting application using bubble sort.

In contrast, data-driven applications like ML applications rely on a model that is derived from the data used to train the model, and this model is not independent from the data but a direct result of the data: different data usually leads to different models.

When such an application is used, the input data is processed based on the trained model – hence based on the prior fed data – and if there are any biases or other issues with the data or the training process, the model can produce unexpected or unwanted results. Moreover, the input data can also be stored and used to further train and change the model. Therefore, unlike traditional algorithms, such data-driven applications cannot be examined and evaluated by using some test cases before being employed and confronted with real data. Hence, discussions surrounding research areas like machine behaviour are needed to examine these technologies when in use.

For education, this means we cannot only consider the ideas, algorithms, and training processes when we want to explain data-driven applications, we have to deal with the data too. And this data, or rather its properties and meanings, are bound to a societal context. That context can be situational: for instance, where, when, and under what conditions data for self-driving cars or data for predicting success as employee or learner in some specific domain and institution is collected (a range of examples is discussed by O'Neil, 2017, and Rahwan et al., 2019).

While Rahwan et al. (2019) suggest discussing and analysing machine behaviour rather abstractly on different levels and within a framework from science, namely biology, we suggest for education to differ in two aspects (see Schulte & Budde, 2018). To demonstrate the everyday application and transferability of ideas to artefacts, we think it can be useful to include the notion of interaction between humans and machines. During such interactions, humans in different roles shape and are being shaped by digital systems. For a discussion on shaping and being shaped see, for example, Rushkoff (2010), or the debate on hybrid interaction systems, man-machine or human-in-the-loop ideas.

Bell and Duncan (2018, p. 141) argued that "[a] complaint about older curricula is that they focus primarily on the applications and the data, algorithms, programs and infrastructure are treated as a black box, while the human is expected to conform to the system, rather than viewing the interface critically and considering what is good about it and what might be improved". Instead, "the big picture of an interaction with artefacts should be at the centre of attention: If we can explicitly confront students with all elements of digital systems in a form that makes sense in their world, we can give them a better understanding of how everything works and enable them to be creators, not consumers" (Bell & Duncan, 2018, p. 142).

In these debates, a common theme is to reflect on which roles and responsibilities should be reserved for humans, and what aspects can and/or should be automated to be processed by the digital artefact. Note, this viewpoint makes it important to not only focus on the artefact itself, but to also include the societal context in which it is used – and in which possible different interaction roles it can unfold. This shift can also be seen regarding the role of data: without context, data is just transformed from input to output; but with context, issues like bias, fairness, completeness, or the need to change the data can occur and be included as topics for education.

In the following section, we present three examples in which we explore and develop approaches to balance the role of the ML mechanisms (e.g., the algorithms used for training a model), the role of data and its contexts, and also the role of artefacts in contrast to abstract ideas.

Examples

Man machine computer

This example is based on the idea of the Sweet Learning Computer (Curzon & McOwan, 2015), as referred to in the report on AI and teaching it at school from The Royal Society (2017). Originally, it was part of a set of teaching ideas for demystifying machine learning (Curzon et al., 2008). The origin of this example, however, dates back to the Matchbox Computer.

The example roughly works as follows. It presents a very simplified chess game with only three figures and a 3 × 3 playing field. This simplified game has a limited number of possible moves overall, and the second player can always win when choosing the right ones. Here, the human has to make the first move. The machine has a list of all possible and correct answering moves and randomly chooses one

of them. In the beginning, it is likely that the machine will lose. However, as a kind of machine learning system, every move that leads to the machine losing is removed, so that eventually the only moves that are left as choices are those that let the machine win.

By playing this game repeatedly (we suggest at least ten times), students can experience this gradual and data-based 'learning' process. To make this process more visible, we divided the steps the machine has to take into several sub-roles, each played by a student. This way, the complete mechanism becomes apparent, and students can literally see that the machine is just following an algorithm – there is no human intelligence needed or involved in letting this process of machine learning unfold. The 'intelligence', so to speak, lies in the setup of the machine. The machine does not really understand or learn to play – it just has fewer and fewer possible moves to choose from during the training phase.

It is interesting to take a close look at the data, and the role of the human player or trainer. If the human tries to win and actually wins, a move causing the machine to lose can be removed. But if the human does not make winning moves, the machine cannot remove its own bad moves and does not learn. We can see these moves of the human as input and training data, and students can experience that the result of ML depends on the training data. Regarding the role of the human, if the human chooses to lose or play badly during the training phase, and to later play well, in this way, the human can affect the machine's learning to prevent it from becoming (too) smart.

This way, the man machine game can teach some basic insights into AI and ML. It is, however, an interesting question whether these insights really become conscious to the learners and whether they can relate these insights to real AI applications, e.g., to autonomous cars. This is discussed in the paper by Große-Bölting

and Mühling (2020), where students were asked about their understanding of the inner workings of ML systems after having played the game outlined above (in a somewhat simplified version). Interestingly, the authors conclude that there was no real transfer and interpret the internalisation of this concept as being. The role of a verbalisation and reflection phase in addition to playing the game thus seems important, and such a phase should probably include some explicit transfer to real-world applications. Just teaching ideas without making the relationship to artefacts explicit seems not to guarantee the desired learning outcome.

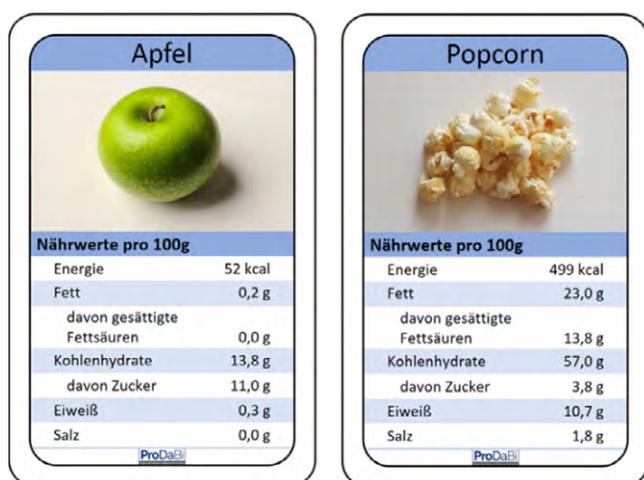
Teaching the systematic creation of decision trees with data cards

This series of lessons aims to give students in grades 5 and 6 an idea of supervised machine learning and artificial intelligence by learning about data-based decision trees. The series is mainly based on unplugged materials that enable action-oriented learning on an enactive level. Additionally, a digital learning environment (for instance, menu-based Jupyter Notebook) can be used flexibly at the end of the series. The selected context of food is relevant for all students and especially suitable for younger students.

Food can be classified as 'rather recommendable' or 'rather not recommendable' based on nutritional information. Several characteristics, such as the amount of fat, sugar, and calories, can be taken into account. Multi-level rule systems that can perform such classifications are called decision trees. Such decision trees can be created based on data. In this case, data means a set of foods for which nutritional information is given and the target attribute (rather recommendable vs. rather not recommendable) is known. Based on this, users can manually create decision trees step by step that classify the food items with a decreasing misclassification rate for every added step. This creation process can also be automated to find

optimal decision rules according to specific criteria. Automation requires representing each food item digitally as a 'data card' – that is, a list of numerical values related to the various nutritional characteristics. A machine learning algorithm then develops a decision tree for this data. In practice, other types of classifiers, e.g. neural networks, are used in addition to decision trees, with machine learning methods adapted to them.

Decision trees have the advantage that they can be understood by students as a system of rules, and the procedures for creating a tree can first be worked out manually with unplugged material and then automated on the computer. In class, food items are represented as physical data cards (see Figure 2) and students can sort and classify the cards to understand the process of creating data-based decision trees on an enactive level. The goal is to gain behind-the-scenes insight into a machine learning algorithm and not just to train classifiers with given systems that remain a complete black box.



Apfel		Popcorn	
Nährwerte pro 100g		Nährwerte pro 100g	
Energie	52 kcal	Energie	499 kcal
Fett	0,2 g	Fett	23,0 g
davon gesättigte Fettsäuren	0,0 g	davon gesättigte Fettsäuren	13,8 g
Kohlenhydrate	13,8 g	Kohlenhydrate	57,0 g
davon Zucker	11,0 g	davon Zucker	3,8 g
Eiweiß	0,3 g	Eiweiß	10,7 g
Salz	0,0 g	Salz	1,8 g

Figure 2. Examples of data cards.

This series of lessons consists of about nine lessons. First, students prepare the data by

labelling data cards as 'rather recommendable' or 'rather not recommendable'. The goal within the lesson series is to create a multi-level rule system for classifying food items. The students first learn to derive decision rules (single-level decision trees) from the data. This is done with the concept of data split, where the data cards are split into two subgroups based on a characteristic and a so-called threshold value (e.g., food with up to 10g of fat or over 10g of fat). In both groups, the majority value is used as the choice of class for food items with similar conditions. The students first learn this concept in a setting of statistics with embodied activities, and then the students use it in small groups with their own set of data cards. The students work out how to systematically search for good decision rules. It becomes obvious that a multi-level rule system is needed. Therefore, based on the first rule, more features are included to create decision rules in the second level of the tree. Depending on how fast the students work, they can create two-level or multi-level decision trees. After different groups of students have created different decision trees, these trees are applied to new food items that the students themselves have created on blank cards. These new food items are classified using all the trees. This makes it obvious that there are also uncertainties in the decision trees, as some food items are classified differently by different trees. In order to systematically investigate the uncertainties in decision trees, each group tests its decision tree with the 15 test cards that are marked as yellow cards. This makes it possible to compare the performance of the decision trees. After carrying out the whole process manually, students can use a prepared menu-based environment in a Jupyter Notebook, for example, to automatically create a decision tree on food data using a computer. They can also change the data in the process and observe the effects on the decision tree. Finally, students reflect on how decision trees are created from data with the help of the computer, what advantages and disadvantages this has, and where the students find such decision models in their everyday lives.

Data awareness in the mobile phone network

This example aims to foster students' data awareness, which means to be aware of the collection of personal data, and its usage and processing for various purposes, during interaction with data-driven digital artefacts. Students should be enabled to be aware of the role of personal data during interactions with data-driven digital artefacts in their everyday lives. This should help students to assess the possibilities, implications, and mutual influences of interaction with a data-driven digital artefact.

This teaching unit for middle school students lasts for about four lessons (45 minutes each), consists of three parts, and addresses the mobile phone network as an example system that collects and processes location data during interaction with it. The unit also connects the students' insights to further examples from students' everyday lives.

In the first part, the context of the mobile phone network is introduced and its composition and inner workings are examined using the example of making a mobile phone call. Thereby, the students also identify which personal data is collected, and what it is primarily used for in this context. For example, the location data of the base station of the mobile phone network to which the user is connected. This location data is necessary to ensure the efficient establishment of a connection between mobile phones – we call this the primary purpose for using and processing the collected data.

In the second part, the students are given some data that was collected by the mobile phone network and published by a German politician (he tried to draw attention to the role of such data). These real-world data include location data collected by using the mobile phone network, e.g., during calls or when texting a mobile phone,

while browsing the internet with the phone, or just from the fact that the phone is logged into the mobile phone network. We call this collection of location data by the mobile phone network an implicit data collection. We developed a web application with which the students can explore these location data. They are set the task of finding out as much information about a person as they can. So, they create a profile or characterisation of a person that the students did not know before, simply by exploring the person's location data. By doing so, students gain some interesting insights about the person, for example, about their leisure activities, or finding out where the person lives or works (Höper et al., 2021). While discussing the profiles created by the students, it becomes apparent why such profiling is regulated by laws in many countries (especially in Europe). The students can then argue for such reasons in a more meaningful way because they have experienced an example of what one can conclude from such data.

In the third part, the insights about the collection and processing of location data during interaction with the mobile phone network are transferred and applied to other data-driven digital artefacts in students' everyday lives. Consequently, the students generalise the insights and examine other data-driven digital artefacts that also collect location data, such as various apps on their smartphones, including those that collect GPS data. During an evaluation and assessment of the collection and processing of location data in various contexts, the advantages and disadvantages of the collection and processing of location data can be discussed. This will foster students' skills for reflective decisions regarding the release of their personal data during everyday interactions with data-driven digital artefacts.

Conclusion

AI education requires developing an adequate picture of the hybrid interaction system – a kind of data-driven, emergent ecosystem that needs to be made explicit to understand its transformative role as well as the technological basics of these AI tools and how they are related to data science.

Interacting with digital artefacts, especially data-driven applications, is often done within a social context, with aims or tasks a human has

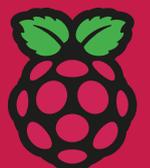
to or wants to reach and complete. While the technical system, by automated processing, helps the human to do so, the question also arises as to which aspects are automated, and what range of possibilities to act and to decide are transferred to the machine, and which are still within the direct control and responsibility of the human. One can use the terms shaping or being shaped, or program or be programmed, to refer to this fundamental issue. It can also be related to the role of the machine. Is it to form a human-machine symbiosis? Is the machinery's purpose to replace humans, or to augment?

References

- Bell, T., & Duncan, C. (2018). Teaching computing in primary schools. In S. Sentance, E. Barendsen, & C. Schulte (eds.). *Computer Science Education: Perspectives on teaching and learning in school* (pp. 131–150). Bloomsbury Academic. <http://dx.doi.org/10.5040/9781350057142.ch-010>
- Clark, M. A. C., & Boyle, R. D. (2006, November). Computer science in English high schools: We lost the S, now the C is going. In *International Conference on Informatics in Secondary Schools-Evolution and Perspectives* (pp. 83–93). Springer. https://doi.org/10.1007/11915355_8
- Curzon, P., & McOwan, P. W. (2015). *The Sweet Learning Computer*. cs4fn. Retrieved March 26, 2019, from <http://www.cs4fn.org/machinelearning/sweetlearningcomputer.php>
- Curzon, P., McOwan, P. W., & Black, J. (2008). *Artificial intelligence ... but where is the intelligence?* cs4fn. www.cs4fn.org/ai/downloads/aiwhereistheintelligence.pdf
- Große-Bolting, G., & Muhling, A. (2020, April). Students perception of the inner workings of learning machines. In *LATICE Seventh International Conference on Learning and Teaching in Computing and Engineering*.
- https://www.researchgate.net/profile/Gregor-Grosse-Boelting/publication/356253786_Students-Perception_of_the_Inner_Workings_of_Learning_Machines/links/61938fca3068c54fa5eebfac/Students-Perception-of-the-Inner-Workings-of-Learning-Machines.pdf
- Höper, L., Podworny, S., Schulte, C., & Frischemeier, D. (2021). Exploration of location data: Real data in the context of interaction with a cellular network. In R. Helenius, & E. Falck (Eds.). *Statistics education in the era of Data Science. Proceedings of the satellite conference of the International Association for Statistical Education*. <https://doi.org/10.52041/iase.nkppy>
- O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., Roberts, M. E., Shariff, A., Tenenbaum, J. B., & Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- The Royal Society. (2017). *Machine learning: the power and promise of computers that learn by example*. <https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf>
- Rushkoff, D. (2010). *Program or be programmed: Ten commands for a digital age*. Or Books. <https://doi.org/10.2307/j.ctt207g7rj>
- Schulte, C., & Budde, L. (2018, November). A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research* (pp. 1–10). <https://doi.org/10.1145/3279720.3279733>
- Thijs, A., & Van Den Akker, J. (2009). *Curriculum in development*. Netherlands Institute for Curriculum Development (SLO). https://www.academia.edu/5132063/Curriculum_in_development_1_?email_work_card=title
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>

Cross-boundary co-design for learning machine learning

Henriikka Vartiainen (University of Eastern Finland)



Cross-boundary co-design for learning machine learning

Henriikka Vartiainen (University of Eastern Finland)

Introduction

Over the years, teaching computer science (CS) has been approached from different theoretical and pedagogical perspectives – both in research and in practice. As Kafai, Proctor, and Lui (2020) argued, some people focus on teaching foundational concepts and practices of the discipline, whereas others focus more on teaching computational design and engineering. Meanwhile, other approaches emphasise strengthening CS experiences that promote critical thinking and social justice, and aim to empower all children to become informed citizens in today's society. Each of these educational traditions has different theoretical underpinnings and priorities, as they typically place emphasis on either the cognitive process, on participation in the practices of communities, or on understanding the impact and influential role of computing in society (Kafai et al., 2020).

Similar questions and debates have also been emerging in terms of teaching artificial intelligence (AI) and machine learning (ML): should we develop pedagogical approaches and tools that support students in understanding the structures and the internal workings of ML? Should we develop pedagogical approaches that promote creativity and broader participation with the help of low-floor ML applications? Or should we focus on AI ethics and data literacy to enable the critical questioning of the practices of our data-driven society?

In the past few years, we have seen a rapidly growing number of initiatives for integrating such AI/ML topics into K–12 education (ages 5–18). However, the challenge is that there are no ready-made teaching practices or clear guidelines on what works when, how, and for whom. Another question regards in which school subject, or combination of subjects, ML should be taught. Moreover, the objectives, tools, and pedagogical approaches can be very different when working in diverse educational settings that range from kindergartens to high schools (Tedre et al., 2021). Research has also shown that what and how subjects are taught in schools is highly dependent on various contextual factors, such as national policies and curricula, as well as local school practices, goals, and values, which shape the everyday activities of teachers (Härkki et al., 2021). The development of new educational practices also requires responsiveness to the learning needs of teachers, as many of them are unaware of the mechanisms, opportunities, and impacts that ML already has on our societies, communities, and individuals (Vartiainen et al., 2022).

On the other hand, over the past decade, efforts by several interdisciplinary teams to conduct educational design-based research (DBR) have shown the significant promise of the strategy of engaging researchers, developers, and practitioners in a model of collaborative, iterative, and systematic research for the development of novel educational practices (Penuel et al., 2011). This approach highlights collaborative endeavours between researchers and practitioners, who work together in designing,

implementing, and evaluating prototypes of learning environments, educational technologies, and pedagogical approaches aimed at addressing concrete educational needs (Penuel et al., 2011). From this perspective, we argue for the potential of cross-boundary collaboration and co-design as a strategy for bringing ML education into existence within school practices.

In this paper, we provide some reflections on our experiences of co-designing and piloting a ML project in Finnish basic education (ages 7–15). We describe how a learning project was co-designed by researchers from different disciplines in collaboration with local school teachers. We also present the pedagogical underpinnings and contextual factors that have informed our cross-boundary work and how these perspectives were transformed into learning activities, infrastructures, and scaffolding provided to school pupils. All this comes together in the closing section, in which we discuss how to support the contextual integration of ML topics into classroom practice.

Co-designing a pedagogical approach for ML: The case of Finland

Mapping local needs and design constraints

In Finland, like in many other countries, ML is a new topic in schools, and there is a significant lack of research and practices on how ML can be made part of our educational practices. In response to these knowledge requirements and challenges, we began our process of co-designing by organising joint discussions between researchers and participating school teachers. In accordance with an educational DBR approach, these joint discussions played

an important role in framing local design constraints, such as how ML projects could be implemented in line with the Finnish national core curriculum, and how our intervention could be customised to serve the local needs and interests of the collaborating school. Although ML is not explicitly included in the Finnish National Core Curriculum for Basic Education (NCCBE), the curriculum's general frames of learning and teaching focus on the development of seven transversal competences:

- T1 – cultural competence, interaction, and self-expression;
- T2 – taking care of oneself and managing daily life;
- T3 – multiliteracy;
- T4 – ICT competences;
- T5 – working life competence and entrepreneurship; and
- T6 – the participation and involvement in and the construction of a sustainable future (T7).

These transversal competencies are to be introduced in local subject-specific curricula as well as through project-based studies that integrate several school subjects. From this perspective, our joint efforts to promote children's agency and ML understanding through collaborative learning and design are well aligned with these national goals.

As a context-specific feature, it is also important to note that Finnish teachers are highly educated professionals who have a high degree of autonomy in their work. While the NCCBE is considered obligatory, the Finnish educational system does not involve standardised testing, auditing, or outside teaching supervision. Instead, the Finnish educational system emphasises trust in teachers' professionalism, and teachers can decide on their teaching and assessment methods. Additionally, research-based approaches for developing educational practices are recognised in both the Finnish national strategies and teacher education (Niemi & Lavonen, 2020).

Making plans for joint action

As yet, there are no ready-made practices for teaching ML, so our joint planning of school projects focused on mapping the key elements of the desired activity system; together, these elements should enable the development of students' ML design skills as well as their understanding of the basics of ML. In practice, this meant negotiating 1) the objectives and learning tasks/problems that pupils face, 2) the tools, technologies, and materials provided, and 3) the forms of social organisation (e.g., individual, small group, and whole class activities) and the division of labour. This also required the creation of a shared understanding of the practical means through which we aimed to orchestrate relations among these elements in diverse stages of the project. These joint agreements were written down in a shared document that served as an externalised plan for the coordination of joint activities.

Implementation of the project through co-teaching

As our previous publications have described this intervention and the educational technologies employed in more detail (Vartiainen et al., 2021; Toivonen et al., 2020), we will now only briefly reflect on how the key elements of the designed activity system emerged during the implementation of the project. As one of the rationales for our pedagogical approach was to provide students with access to expert-like practices by working together with CS experts, we also elaborate on how CS researchers and teachers scaffolded the students' learning of ML.

In short, our pedagogical approach relied on design-oriented pedagogy, which aligns well with the national curriculum in Finland. Design-oriented pedagogy entails students building their conceptual understanding and new ways of thinking by creating digitally or materially

embodied artefacts and projects with the support of technology (Kafai et al., 2020). Instead of scripted, build-a-thing tasks or step-by-step exercises, the students were instructed to work in small groups and were given open-ended learning tasks to generate ML solutions to real-life problems that they considered to be meaningful. In other words, the students had a large degree of freedom in terms of what to co-design within the epistemic, material, and social structures that support the learning of basic ML concepts and practices. Within the project described here, students' could learn to follow the basic epistemic functions related to ML workflow for problem-solving: how to collect data relevant to solving the problem, how to filter and clean the data, how to label the data, how to use those data to train a classifier, and how to link the classifier results with desired behaviours (in a web app, for instance), and evaluate the model (Tedre et al., 2021). While the development of conceptual understanding was deliberately addressed, we emphasised its creative uses with regard to solving everyday problems and gaining new insights into the ML-driven world students are already living in.

In terms of afforded technologies, students worked with Google's Teachable Machine 2 (GTM, see Toivonen et al., 2020), and our own in-house developed educational application for ML (Mariescu-Istodor & Jormanainen, 2019). At the beginning of the project, CS researchers demonstrated the ML workflow using GTM, which also gave the opportunity for the students to observe the processes through which experts make use of basic conceptual and procedural knowledge. As the CS researchers were verbalising their actions and thought processes, the students could also begin to build a mental model of the target processes that were required to accomplish the given learning task. Consequently, the students familiarised themselves with the possibilities of afforded tools. During this hands-on exploration, the CS researchers actively observed the students' activities, and they provided on-demand support

by offering hints, guiding questions, conceptual reminders, and scaffolding (see Figure 1). At the end of the first workshop, the students were assigned an individual homework task, which

asked them to search for and identify everyday problems that could be solved by using ML-based technologies.

EXAMPLES OF SOCIAL SETTINGS AND SUPPORT

AROUND GTM2



Figure 1. Examples of social settings and support.

For the second workshop, we produced a design template with the aim of supporting the further development of ideas through collaboration. This template followed the GTM workflow and asked students to analyse what their own apps would do, what kind of data would be collected and from where (image, sound, poses), how many different categories the model should be able to recognise, and under what conditions the teaching data would be presented. Here, the idea was to provide scaffolding for the students to adapt basic ML concepts to the problems that they themselves were trying to solve. These externalised ideas were further defined in collaborative discussions with computer scientists who, for example, encouraged the students to reflect on their ideas, articulate the reasoning behind their dataset choices, and direct their attention to previously unnoticed aspects, such as the important role of background settings. In this way, the content of the expert scaffolding was immediately related to specific design ideas as well as emerging challenges that arose when the students were creating training datasets for their own ML applications.

Moreover, working in teams provided students with an additional source of scaffolding, in the form of knowledge and processes distributed throughout the groups. Developing their own ML models for applications also provided an important feedback loop, as the students could receive immediate feedback on the implementation of their own design ideas; that is, they could test the quality of their own data. GTM did not always work as students expected; hence, the students also had to analyse the relationship between their own input and the output provided by the responsive tool. Such a feedback loop also made it easier for the students to explore how the agent represents the world and perceives the information it receives and how it modifies its behaviour accordingly (Druga et al., 2019). In addition, each of the student teams worked on its own problem,

and thus, they were also able to observe multiple ways in which this tool and the related conceptual knowledge can be applied.

On the other hand, even if such a low-threshold tool can support novice learners in familiarising themselves with ML, the implementation of such a design project revealed the importance of creating productive social settings and scaffolding around ML-based educational technologies. These technologies were actively used by the students in different phases of the project: when exploring the opportunities presented by these technologies in the first workshop, when creating datasets for their own applications in the second workshop, and when testing their own and other teams' applications in the final workshop. In addition, the collaborative discussions with their peers as well as with the CS teachers were constantly organised around these tools. Evidently, working with computer scientists was important as these experts had a deep understanding of the ML technologies being used and disciplinary ways of thinking, which helped them to be flexible in supporting and responding to the ideas and questions that students developed during the process. Meanwhile, the class teachers played an important role with regard to guiding the student teams in working together productively, engaging in respectful discourses and reflections, taking shared responsibility for the collaborative work, and displaying persistence in developing their understanding of ML. During these workshops, the students could also witness how their teachers were interested in learning about ML, and how they engaged in collaborative discussions to figure out how ML works. In this way, the teachers were also modelling how creative experts deliberately work at the edge of their competences and actively try to go beyond their current understanding (Bereiter & Scardamalia, 1993).

Reflection

Our process of co-design ended with reflections provided by all the actors involved. At the end of the project, each student team was asked to reflect on its own design and learning process. Such reflection deliberately challenges students to explain their own thinking and actions, and thus, these discussions provide information on how the process of co-design was actualised and experienced from the pupils' perspectives. Likewise, verbal explanations of the functionality of the application designs gave additional information about what kind of support pupils might need in their co-design projects. In a similar vein, joint discussions were held with

the school teachers with a focus on how our co-designed activities, tools, and infrastructures supported the design and learning process of the students. In line with the DBR approach, various kinds of empirical data were collected, analysed, and reported in collaboration with the researchers from computer science and education.

Discussion

In this paper, the aim was to provide reflections on our process of co-designing and piloting ML learning projects in Finnish basic education (see Figure 2).

CROSS-BOUNDARY CO-DESIGN

FOR LEARNING MACHINE LEARNING

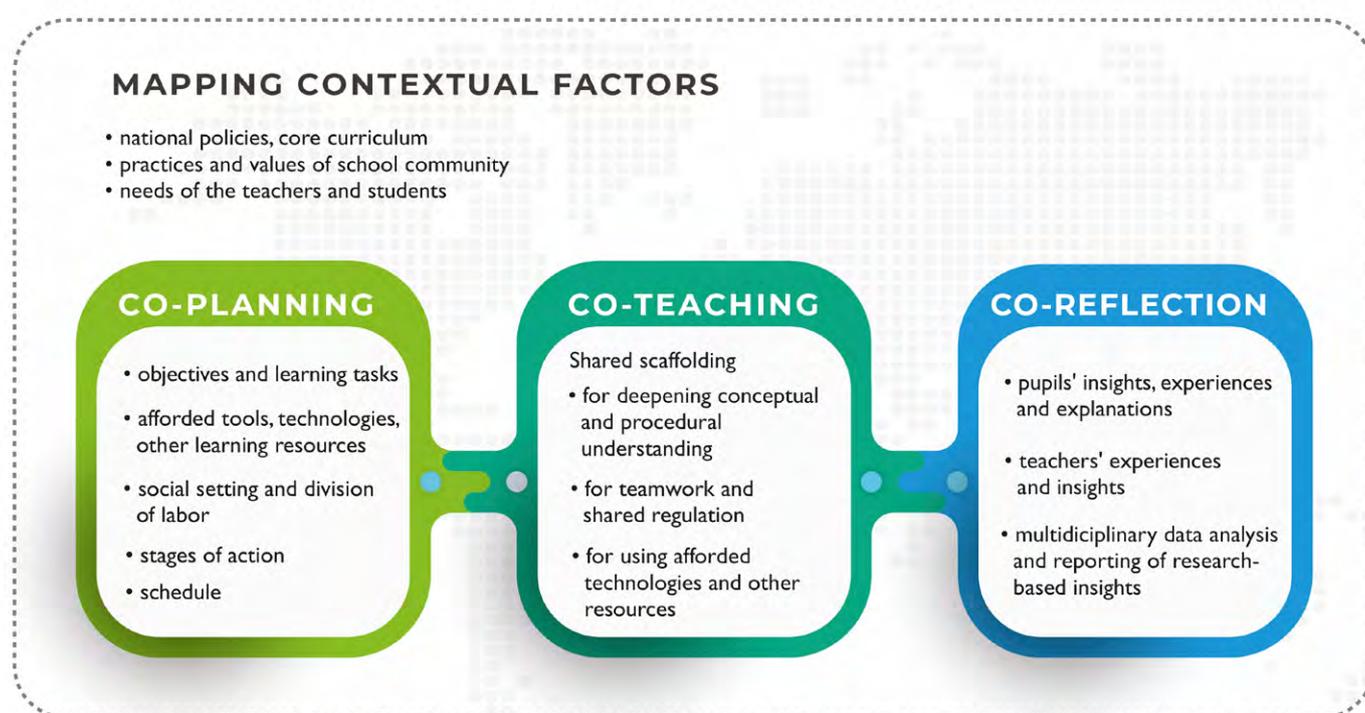


Figure 2. Cross-boundary co-design.

To conclude, our project illustrated how co-design between researchers and teachers was crucial for the success of the project. These joint efforts also provided indirect scaffolding for the emerging learning activities of the students, as the focus of the design was on the specification of the objectives and learning tasks that pupils pursue and the selection of appropriate technologies and forms of social organisation during the different stages of action. Overall, the design of novel educational practices, which were simultaneously relevant, appropriate, and capable of being realised, was crucially dependent on capitalising on complementary expertise. In this design process, the researchers from computer science contributed their extensive experience in computational thinking and ML educational tools, whereas the educational researchers brought theoretical and pedagogical ideas from the learning sciences. The teachers, in turn, provided their unique insights with regard to adapting the practices and educational technologies suggested by the researchers to serve their local contexts and the actual needs of their own students.

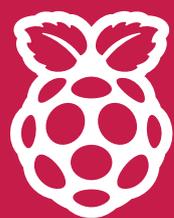
In a similar vein, the potential of cross-boundary collaboration was demonstrated during the implementation of the project as the CS researchers and schoolteachers all brought their unique knowledge, skills, and expertise to help scaffold students' learning and understanding. On the other hand, the arrangement of having researchers and teachers working together in a classroom is a unique setting and far from the reality of most schools. Yet, we believe that co-teaching may provide promising paths for teachers to develop and orchestrate novel practices for integrating ML topics in education. Our example also illustrated how disciplinary knowledge, pedagogical knowledge, and technological skills were all needed in the scaffolding of students' activities (Valtonen et al., 2019). While some studies and practices of design-oriented learning have created the impression that students can deepen their skills

and understanding on their own, it is important to understand how pedagogical infrastructuring and contextual scaffolding provided by teachers and experts are a driving force for learning (Viilo, 2020).

Overall, the process provided promising results in terms of supporting middle schoolers to become co-designers and creators of their own ML applications in a manner that provided valuable learning experiences. Yet, the question remains as to how to scale up such practices and provide concrete resources for teachers to implement ML projects in various educational settings. Challenges also remain in terms of ML-based educational technologies. The use of low-floor ML-based educational technologies, such as those used in this project, may provide a promising entry point for novices looking to learn ML. Yet, students learn about ML workflows but not about the internal ML mechanisms (Tedre et al., 2021). While progress has been made in developing explainable AI for education, we should also develop learning tasks and processes that are responsive to students' evolving skills and understanding by increasing the complexity of skills and concepts. In a similar vein, we should make sure that children are able to extend and refine their creative abilities, critical thinking, and participation with regard to these evolving technologies. Accordingly, we will not only need ML-based educational technologies and learning materials, but also research-based understanding on how to create appropriate social settings and pedagogical infrastructures around these evolving educational technologies.

References

- Bereiter, C. & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Open Court.
- Druga, S., Vu, S.T., Likhith, E., & Qiu, T. (2019). Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019* (pp. 104–111). <https://doi.org/10.1145/3311890.3311904>
- Härkki, T., Vartiainen, H., Seitamaa-Hakkarainen, P., & Hakkarainen, K. (2021). Co-teaching in non-linear projects: A contextualised model of co-teaching to support educational change. *Teaching and Teacher Education*, 97, Article 103188. <https://doi.org/10.1016/j.tate.2020.103188>
- Kafai, Y., Proctor, C., & Lui, D. (2020). From theory bias to theory dialogue: embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education. *ACM Inroads*, 11(1), 44–53. <https://doi.org/10.1145/3381887>
- Mariescu-Istodor, R., & Jormanainen, I. (2019, November). Machine learning for high school students. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research* (pp. 1–9). <https://doi.org/10.1145/3364510.3364520>
- Niemi, H. & Lavonen, J. (2020). Teacher education in Finland: Persistent efforts for high-quality teachers. In L. Lefty & J.W. Fraser (Eds.), *Teaching the World's Teachers* (pp. 153–178). Johns Hopkins University Press. <http://hdl.handle.net/10138/326792>
- Penuel, W.R., Fishman, B.J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40(7), 331–337. <https://doi.org/10.3102/0013189X11421826>
- Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching machine learning in K–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. *IEEE Access*, 9, 110558–110572. <https://doi.org/10.1109/ACCESS.2021.3097962>
- Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., Valtonen, T., & Vartiainen, H. (2020, July). Co-designing machine learning apps in K–12 with primary school children. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* (pp. 308–310). IEEE. <https://doi.org/10.1109/ICALT49669.2020.00099>
- Valtonen, T., Sointu, E., Kukkonen, J., Mäkitalo, K., Hoang, N., Häkkinen, P., Järvelä, S., Näykki, P., Virtanen, A., Pöntinen, S., Kostiaainen, E., & Tondeur, J. (2019). Examining pre-service teachers' Technological Pedagogical Content Knowledge as evolving knowledge domains: A longitudinal approach. *Journal of Computer Assisted Learning*, 35(4), 491–502. <https://doi.org/10.1111/jcal.12353>
- Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., & Valtonen, T. (2021). Machine learning for middle schoolers: Learning through data-driven design. *International Journal of Child–Computer Interaction*, 29, Article 100281. <https://doi.org/10.1016/j.ijcci.2021.100281>
- Vartiainen, H., Pellas, L., Kahila, J., Valtonen, T., & Tedre, M. (2022). Pre-service teachers' insights on data agency. *New Media & Society*. <https://doi.org/10.1177/14614448221079626>
- Viilo, M. (2020). Teacher's orchestration in longitudinal collaborative inquiry process. *Helsinki Studies in Education Kasvatustieteellisiä tutkimuksia*. <http://urn.fi/URN:ISBN:978-951-51-6106-2>



Raspberry Pi
Foundation

