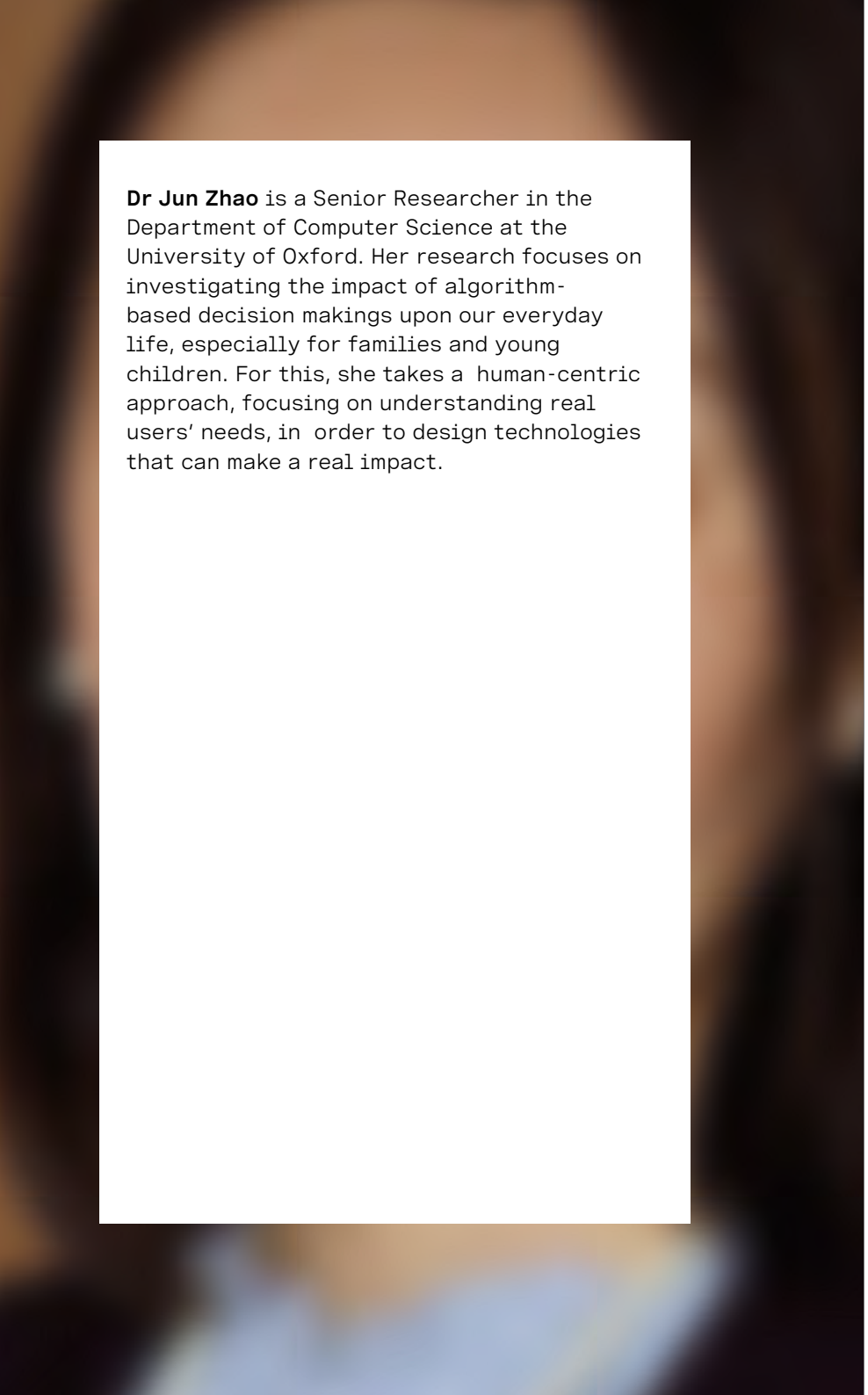

Dr Jun Zhao is a Senior Researcher in the Department of Computer Science at the University of Oxford. Her research focuses on investigating the impact of algorithm-based decision makings upon our everyday life, especially for families and young children. For this, she takes a human-centric approach, focusing on understanding real users' needs, in order to design technologies that can make a real impact.

Call for a new data governance structure for datafied childhood

Jun Zhao, University of Oxford

Why datafied childhood needs data trusts

This essay discusses a new data governance model in UK state schools so that they can regain control of education data and be better supported to ensure sufficient data stewardship. A wave of new decentralised paradigms for data sharing and ownership is being explored to expand individual data subjects' ability to access data and establish data autonomy. A data trust provides a promising response to schools' need for an independent and trustworthy body of experts, who can make critical decisions about who has access to data and under what conditions. We use a case study to demonstrate what a data trust model may provide. However, creating a new data governance structure is not without challenges. We conclude the essay by discussing open social, legal and technological challenges to be considered, calling for a pilot model of data trusts in the educational technology (EdTech) sector.

These challenges must be met because children today are spending more time with digital technologies, which provides a wide range of unprecedented opportunities for their education, socialisation or entertainment. However, this also contributes to the rise of a datafied childhood (Mascheroni & Siibak, 2021), during which children's actions are pervasively recorded,

tracked, aggregated, analysed and exploited by digital technologies and platforms in unpredictable ways. Like many other types of digital technologies, EdTech is increasingly included in UK schools to enhance children's learning opportunities, and the COVID-19 pandemic has greatly accelerated this adoption. The EdTech sector is reported to have grown by 72% in 2020 (Walters, 2021), and Google reported in May 2021 that its user numbers for Google Classroom rose to 150 million from 40 million the previous year (Williamson, 2021). This growth of EdTech usage at schools is raising concern about risks to children's data privacy, and the actual benefits of these technologies.

Reports have shown that the amount and range of education data being routinely collected in UK state schools have grown exponentially in the last few years (Persson, 2020). These data can be collected and processed at schools for a variety of purposes (see Table 1). Sometimes, schools are collecting these data under their obligation to the Department for Education (DfE); other times, they need to process children's data as part of teaching, assessment, administrative and safeguarding (DFC, 2021b). Finally, schools are increasingly contracting external EdTech companies to process children's data to enhance their learning and education opportunities. We name this last type of data 'learning education data', which is the focus of this essay, because it poses special challenges to schools' ability to safeguard children's data rights given the current UK education data regulatory frameworks.

Types of data	Purposes	Examples
National school data (e.g., from the central pupils' record	Under the obligation of the DfE	• Names • Date of birth • Gender • Ethnicity • First language • Special educational needs and disability • Home address • Unique Pupil Number (UPN) 0+ • Unique Learner Number (ULN) 14+ • Any form of UPN
Data generated for safeguarding children (e.g., Education and Health Care Plan)	For safeguarding and child protection	• Health data (from health & safety management) • Demographic data (for looked-after or vulnerable children) • Online activity monitoring
Data generated by learning tech for management	Helping schools with administrative tasks	• Lesson and homework delivery • Sometimes biometrics data for accessing facilities such as libraries or cashpoints
Data generated by learning tech for learning and assessment (e.g., Google Classroom, Show My Homework, HegartyMaths or other AI-based personal learning assistants)	Helping schools to enhance children's learning and education opportunities	• Lesson and homework delivery • Online learning, including attendance and absence and resulting metadata (e.g., IP address, device information) • Assessment and testing results • Behaviour traits data, for measuring engagement and usage

Table 1: An outline of different types of data being generated in a UK school setting

Source: Adapted from DFC (2021b)

While schools are expected to be the primary duty bearers of children's best interests, their duties are compounded by the complexity of legislation in the education sector and the extreme challenge of carrying out compliance validation. The diverse range of education data being collected at UK state schools for different purposes is subject to a variety of regulatory frameworks developed for different purposes and at different times, including the Data Protection Act 2018 that sets up the UK-specific data protection framework and sits alongside and supplements the UK General Data Protection Regulation (GDPR), the Human Rights Act 1998 and the Equality Act 2010, which ensure the protection of children's rights, as well as the Digital Economy Act 2017, for the protection of public sector data.

As a result, UK schools are equipped with no specific legislation concerning EdTech or education data, and there is no overarching education data governance framework in the UK (DFC, 2021b, p. 22). While not arguing for sector-specific legislation, we emphasise the need for schools to have more coherent guidance to simplify the navigation of data protection regimes. Furthermore, schools are given no oversight concerning EdTech companies' compliance with data protection or cybersecurity laws and standards, leaving them with a market of EdTech companies that are not subject to systematic audits (DFC, 2021b, p. 22).

Validating data protection compliance is particularly challenging for schools to manage as they are often collected and processed via external EdTech companies that schools contract with. Identifying whether schools or EdTech companies should be the data controller (who determines the purpose and means of data processing, and is thus responsible for compliance with the GDPR) or the data processor (who acts on behalf of, and only on the instructions of, the relevant controller) in these scenarios is not always easy. Data controllers and processors have different responsibilities for the type of data that can be collected and how they are used. Schools are mostly expected to be the data controller, or a joint data controller, to ensure that children's data is not misused, or its processing is compliant with all regulations (DFC, 2021b, p. 27). However, recent cases show that schools can struggle to exercise all the data audits that are needed when they have allowed children's data to be accessed and used for third-party commercial purposes (Persson, 2020).

Determining an EdTech company's role is a complex task as it requires a degree of legal analysis and a sufficient understanding of the EdTech company's data-processing practices (DFC, 2021b). When the data collected and processed by an EdTech company is used for the sole purpose of education and learning of the child, the company is most likely to be the data processor and the school the data controller. However, when the same data is used by the same EdTech company for their own product development or marketing to children, then it would also become an independent data controller. Both data controllers and data processors are accountable for data processing, but controllers are more so because they decide how the data will be used. The granular data-processing purposes could also affect whether the EdTech companies will be subject to the risk-based age assurance statutory code produced by the Information Commissioner's Office - the Age Appropriate Design Code (AADC), otherwise known as the Children's Code.[1]

This essay discusses a technical alternative to learning data management and governance for schools. While experts have not yet reached a consensus regarding the effectiveness of digital technologies for improving children's learning (DFC, 2021a), children's rights and best interests are in jeopardy. Schools need more transparency and better control concerning the data-processing practices of third parties, and they need to be better supported by independent entities to navigate the complex legal frameworks, who also have children's best interests at heart.

Data trusts

A wave of new decentralised paradigms for data sharing and ownership is being explored to expand individual data subjects' ability to access data, by enabling collective access requests through representative intermediaries such as non-governmental organisations (NGOs) and trade unions, therefore increasing the agency of individual data subjects. A range of data governance structures has emerged, such as

data commons, data trusts or data cooperatives, in response to different social and legal needs from individuals and organisations.

EdTech companies hold much of the education data, and schools and families have limited visibility and control of what is being collected and used. A new paradigm, with increased transparency and autonomy, must be investigated. Data trusts provide a good starting point for the challenges that UK schools are facing for the following reasons. First, we need to ensure we avoid overburdening the commitment of individual families to manage the stewardship of their datasets, as would be commonly expected in data commons or cooperatives. This would demand a level of data literacy that may leave some families in a more disadvantaged position and overburden busy families. Second, we recognise the complexity and sensitivity of the range of pupils' data involved in the education settings, which requires a trusted body with sufficient understanding of children's best interests and legal obligations to carry out the scrutiny. This body of trustees should include not only conventional data protection officers, but also various other stakeholders, such as parents or caregivers, who should be better informed of their children's data rights and involved in the process of consent, and educators, who are at the forefront of data protection obligations.

The Open Data Institute defines a data trust as 'a legal structure that provides independent stewardship of data', including deciding who has access to data, under what conditions and to whose benefit (Open Data Institute, 2019). It is different from other data governance structures because it represents 'a legal relationship where a trustee stewards data rights in the sole interests of a beneficiary or a group of beneficiaries' (van Geuns & Brandusescu, 2020). Instead of taking a grassroots governance model (such as a data commons), the trustees can be the decision-makers regarding who has access, under what conditions and to whose benefit, and they take on a legally binding responsibility for data stewardship (Open Data Institute, 2019).

To date, different forms of data trusts have been developed for different purposes. Data trusts have been established in support of democratic purposes, such as the civic data trust established in Toronto to free citizens' access to urban data (Dawson, 2018) and govern which companies have the right to operate and collect data in a particular urban public space. We also see data trusts providing 'bottom-up' support for a group of individuals to help regain control over their personal data and provide a legal mechanism to exercise their data stewardship that reflects their needs and preferences (Delacroix & Lawrence, 2018). This format is expected to 'enhance protection for individual privacy and autonomy, address existing power asymmetries between technology companies, government and the public, and empower the latter to share in the value that data and artificial intelligence promise' (Open Data Institute, 2019).

The third type of data trust responds to the needs of one or more organisational data holders - which may or may not include personal data. Here, trusts are expected to make decisions for the organisations regarding when to allow access to their data for broader, or more strategic, purposes. The trust is often set up with a board of trustees to reflect the different interests and priorities of its users and to safeguard the organisation's vision for the public good. This type of data trust provides a strong fit to the needs of schools wishing to safeguard access to children's learning education data without having to make regular, granular decisions about data access.

Data trusts for education data: a case study

It has been exciting to see some practical developments of data trusts recently, built on extensive theoretical landscaping. However, developing data trusts is a complex task and requires a strong commitment from data holders and the users' community. Furthermore, existing legal frameworks are not necessarily ready to support all the data stewardship and legal binding responsibilities designed for a data trust. Here, we use a case study to illustrate how a data trust can provide an alternative data governance structure for the learning education data collected and processed at schools.

The case study is developed based on existing research about education data in UK schools (Persson, 2020) and the use of artificial intelligence (AI) systems in children's lives (Wang et al., 2022). Research has shown that personalised

learning or intervention applications are one of the most dominant areas in which machine learning techniques are applied to children (Wang et al., 2022). These include AI systems created for purely educational purposes, such as generating personalised learning content for children or assessing children's learning outcomes as well as systems that support the physical wellbeing of children or cognitive development, such as scheduling personalised strategies to promote children's physical or cognitive development. The types of data processed by these systems can include: (a) demographic data, such as age, gender and ethnicity; (b) health data, such as medical history or treatment records; (c) biometrics data, such as video, voice or fingerprints; or (d) behaviour data, such as search history, watching history, chat records, location data, or users' preferences (Wang et al., 2022).

Figure 1(a) shows how three types of actors are currently involved in a scenario where learning EdTech is deployed for improving students' learning (by tracking their performance and interactions with the programme) or development (by accessing their previous health histories and tracking their behaviour traits). *EdTech platforms* often claim that this learning data is used by their algorithms to improve the accuracy of the learning support provided by the students. However, *schools* do not always have direct access to these data or control who may access them or the ability to control what the companies are doing with the data. In addition to carrying out product improvements, *EdTech companies* may use *students'* data for system performance monitoring, marketing or other commercial purposes. In many cases, students' data are simply tracked and accessed for non-core purposes without any explicit users' consent.

Figure 1(b) shows that a fourth stakeholder, a data trust, can act as an intermediary for schools, making decisions about what education data can be accessed by a third party, investigating the purposes of data access, and assessing how they may be aligned with students' best interests. A data trust can be established by the needs of several schools and constitute a group of trustees that represent the diverse interests of schools and data subjects (such as parents and

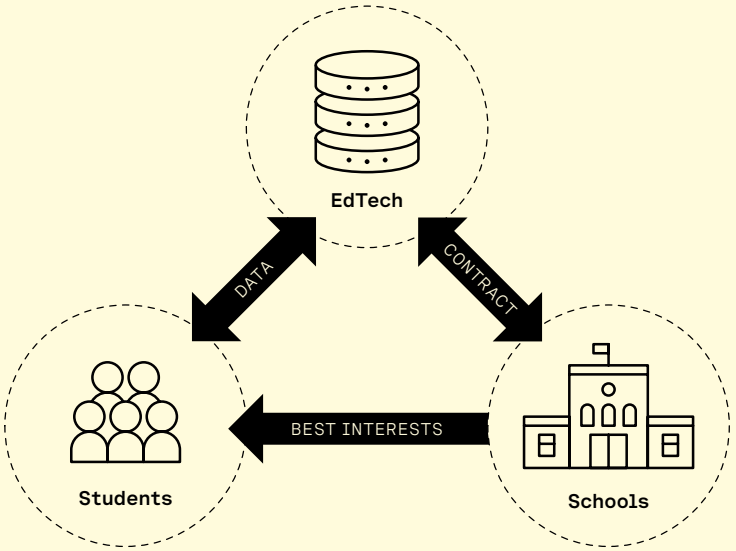

Figure 1(a): A hypothetical scenario of a school contracting an EdTech, with students' data held and processed by the EdTech company

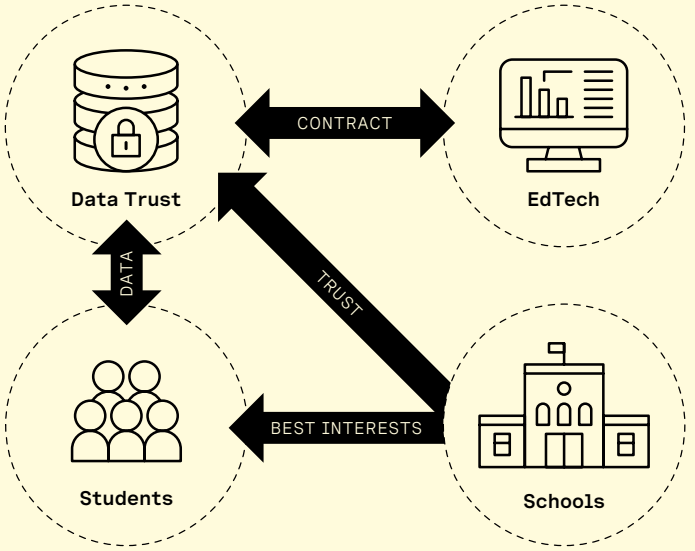

Figure 1(b): A hypothetical design of a school delegating the data protection and compliance responsibility to a trusted education data trust

caregivers or children). All the trustees should be involved in the requirements gathering and design of the data trust model from the onset as well as the evaluation of the data trust for fulfilling its objectives. It provides a promising direction for mitigating the challenges that schools face regarding data safeguarding and compliance checking. It also provides a great opportunity to enhance data sharing, reuse and the development of new education technologies, with improved access to a diverse range of data that is currently being held privately by third parties.

Despite all its promise, there aren't actually many data trusts. Existing pilot studies have shown that it is critical that users are engaged from the start of a data trust so that its development is guided by users' needs, and supports all the critical decision-making points that a user must operate with. The Open Data Institute proposed a six-phase methodology for the development of a data trust (Open Data Institute, 2019), and the engagement with data holders, users and beneficiaries should be involved from the onset of the six phases of scoping, co-designing to launching, operating, evaluation and retirement.

It is critical that the development of a data trust starts with a clear scope, by identifying a specific issue to be addressed (such as better control of students' health data for personalised cognitive support) and researching existing efforts that may address the issue. This will involve engagement with all relevant data holders (aka students and schools) and users (aka EdTech), to understand incentives for their engagement and associated risks.

Furthermore, the operation of a data trust needs to be underpinned by technical infrastructures, and there are few specialist data trust tools, technologies or platforms. A data trust intermediary needs to be able to process requests from data holders and users, carry out audits and verification to ensure compliance with the agreement, and detect and manage breaches of rules. In this use case, students' data is predominantly held by the EdTech company. The data trust intermediary is expected to negotiate ownership of this data with EdTech to ensure that EdTech accesses the data in a way that is compliant with the current data protection rules.

Furthermore, the intermediary should be able to provide information to establish the accountability of the EdTech company by tracking what data is accessed and for what purpose. Many of these technical solutions would need to be developed, and at the same time work with existing data management infrastructures of schools and their technical skills. Emerging technical solutions for enabling decentralised data governance, such as Databox (Mortier et al., 2016) or Solid,[2] are expected to support this new data governance model by enabling users' data autonomy and control of use. However, these solutions' support for complex data requests and accountability has not yet been validated.

Finally, there is still a vast range of legal considerations to be undertaken regarding data trusts (HCC, 2021). Under this new and decentralised paradigm, one must carefully think about: who will be responsible for data sharing, data control and data curation? How can responsibilities be attributed if something goes wrong, which can range from the compromised quality of the data provided by a data subject to misuse of shared data by the data processor (such as data re-sharing or re-identification)? How can users be helped to adapt to this new data governance structure? And in the case of data trusts, how should the legal responsibility of a data trust be defined, and under which legal framework?

Conclusion

UK state schools are facing unprecedented challenges concerning the safeguarding of children's data, given the complex legal landscape and lack of consistent guidance. A data trust offers a new data governance structure that may serve as a starting point to re-landscape the different parties responsible for the sharing, control and curation of education data. Data trusts encourage schools and other data holders to rethink how to establish a legal body to represent their best interests and carry out data stewardship. However, the implementation and operation of data trusts requires the involvement of all relevant stakeholders, and new technologies and possibly legal frameworks to be developed for specific needs.

Schools need more consistent support for data protection

to ensure compliance and enforcement when they may not always have the knowledge. This new data governance structure could bring new opportunities as well as challenges. It will be exciting to see how piloting of the data trust model may provide more insights regarding possibilities as well as the social, technical and legal challenges.

Dawson, A. H. (2018). An update on data governance for Sidewalk Toronto. *Sidewalk Talk*, 15 October

Delacroix, S., & Lawrence, N. (2018). *Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance*. Birmingham Law School. doi:10.2139/ssrn.3265315

DFC (Digital Futures Commission). (2021a). *Addressing the problems and realising the benefits of processing children's education data*

DFC. (2021b). *Governance of data for children's learning in UK state schools*

HCC (Human Centred Computing Research Group). (2021). *On the planned reform of the UK Data Protection Law*

Mascheroni, G., & Siibak, A. (2021). *Datafied childhood: Data practices and imaginaries in children's lives*. Peter Lang

Mortier, R., Zhao, J., Crowcroft, J., Wang, L., Li, Q., Haddadi, H., Amar, Y., Crabtree, A., Colley, J., Lodge, T., Brown, T., McAuley, D., & Greenhalgh, C. (2016). Personal data management with the databox: What's inside the box? *Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking*, December

Open Data Institute. (2019). Data trusts: Lessons from three pilots

Persson, J. (2020). *The state of data 2020*

van Geuns, J., & Brandusescu, A. (2020). Shifting power through data governance. mozi://a

Walters, R. (2021). UK EdTech sector grows to £3.5bn as demand surges for digital classrooms and AR. *FE News*, 14 January

Wang, G., Zhao, J., van Kleek, M., & Shadbolt, N. (2022). Informing age-appropriate AI: Examining principles and practices of AI for children. *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '22)*, April

Williamson, B. (2021). Google's plans to bring AI to education make its dominance in classrooms more alarming. *Fast Company & Inc.*, 28 May

1 https://ico.org.uk/for-organisations/childrens-code-hub/additional-resources/faqs-for-education-technologies-edtech-and-schools

2 http://solidproject.org