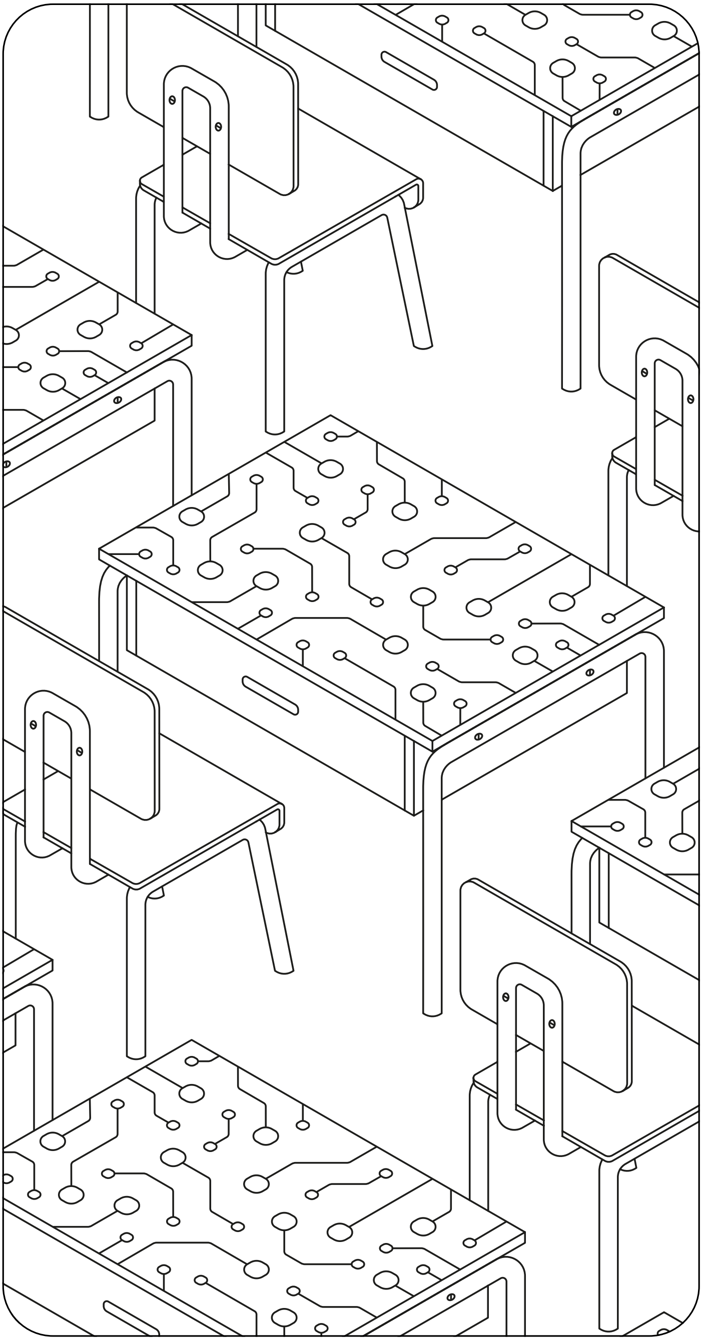
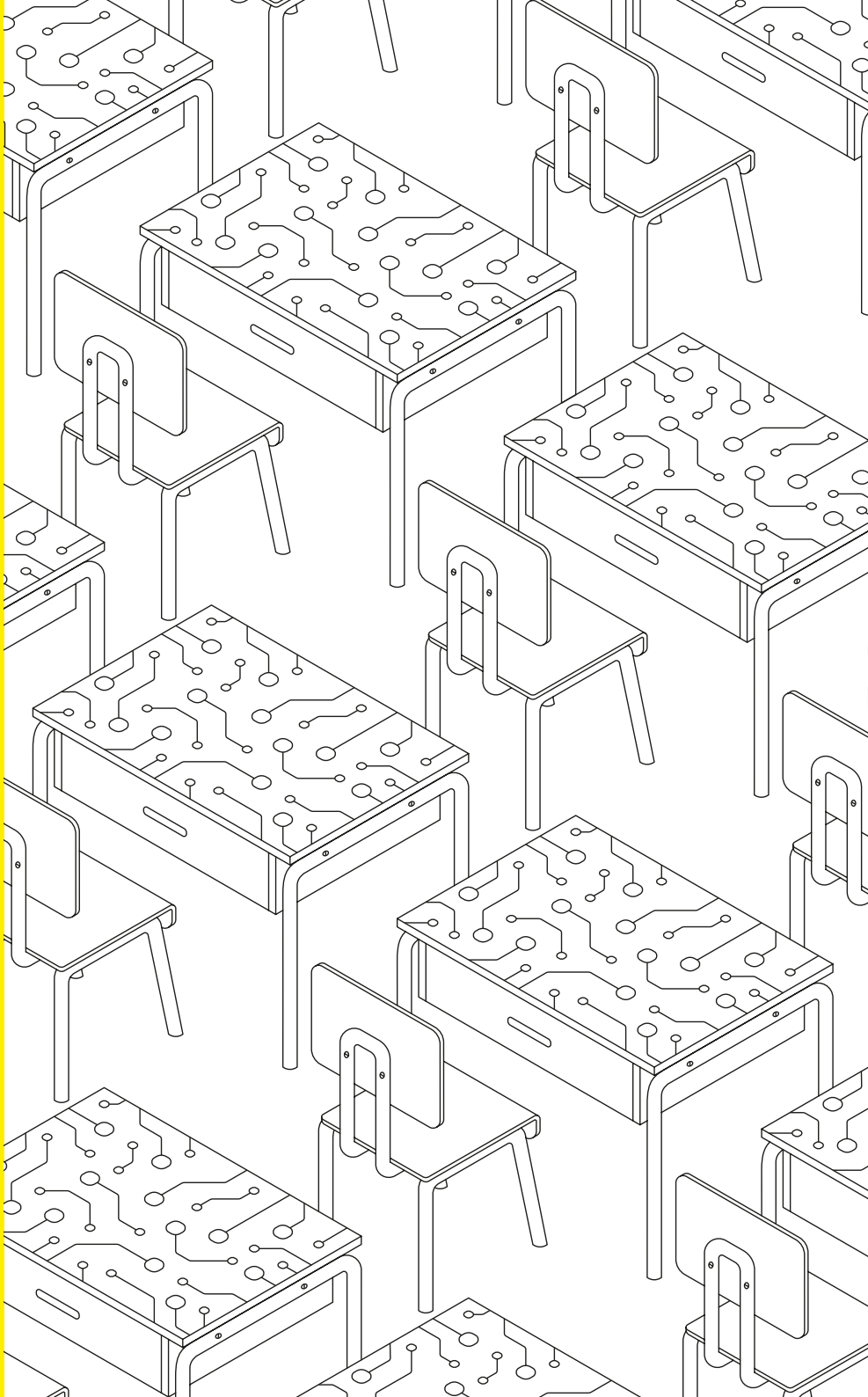
EDUCATION DATA FUTURES



Critical, Regulatory and Practical Reflections



EDUCATION DATA FUTURES
Critical, Regulatory and Practical Reflections

Edited by Sonia Livingstone and Kruakae Pothong, for the Digital Futures Commission

Published and distributed by 5Rights Foundation

Designed by David Weller and Daniel McGhee

Cover illustration by David Weller

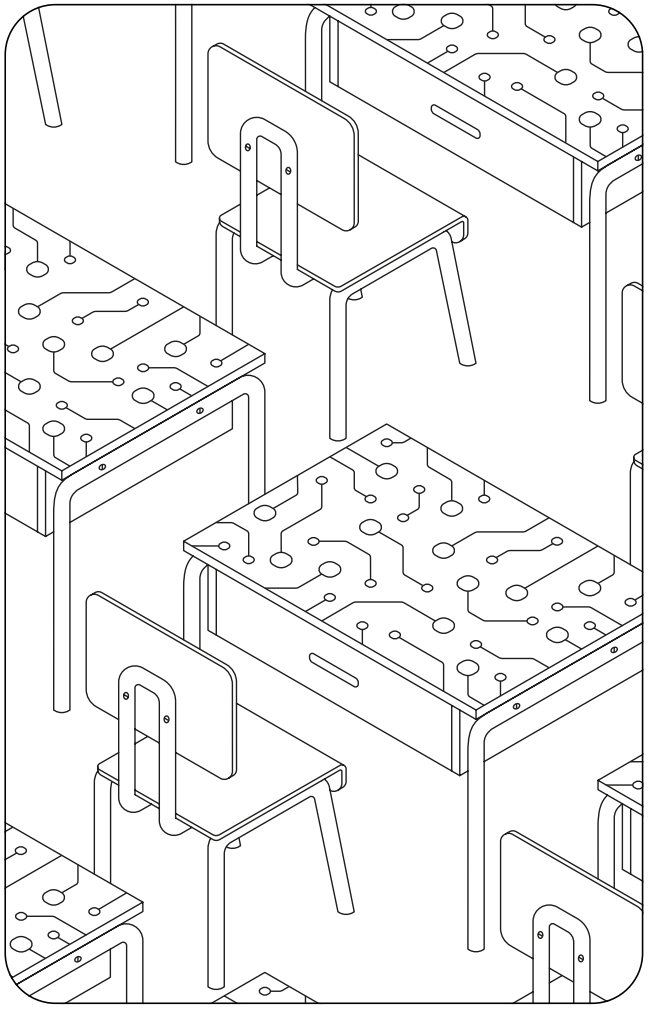
Printed by Park Communications Ltd.

5Rights Foundation
Charity number: 1178581
Company number: 11271356

https://5rightsfoundation.com

5RIGHTS FOUNDATION | DIGITAL FUTURES COMMISSION
Innovating in the interests of children and young people

EDUCATION DATA FUTURES: Critical, Regulatory and Practical Reflections



We would like to thank all the contributors and experts for our education data work and those who acted as anonymous reviewers.

The Digital Futures Commission is grateful for the support of the 5Rights trustees.

INTRODUCTION

Hegarty Maths is just memorising answers not actually helping you learn stuff (Boy, 15)

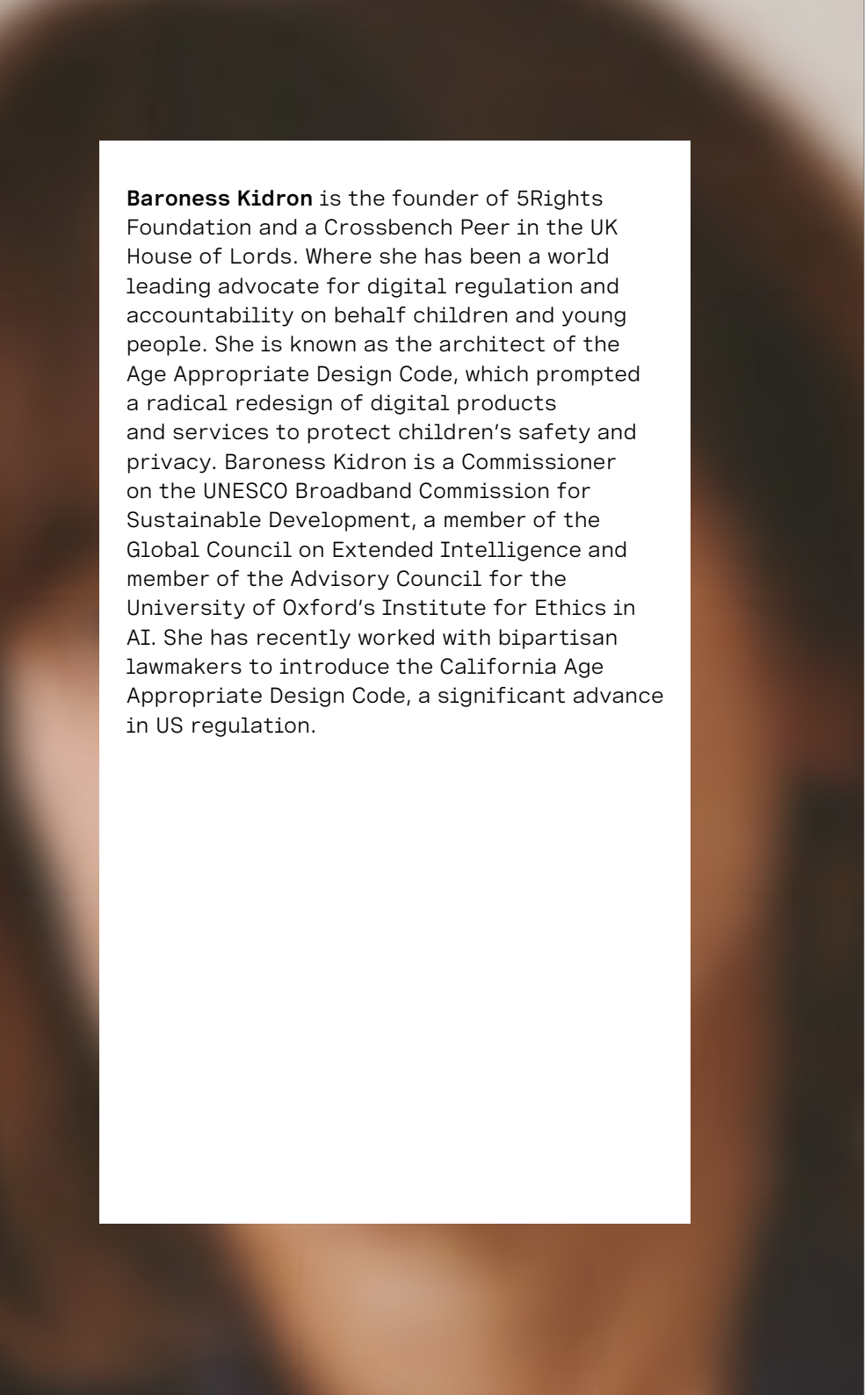
56% of 6-17-year-olds were asked by their school to use Google Classroom or Microsoft Teams this year

Times Tables Rock Stars makes it fun and competitive (Girl, 10)

9 in 10 children used one or more of the 11 apps in our EdTech survey

Seesaw is a bit more fun and interesting than a book (Girl, 9)

Digital Futures Commission (2022) EdTech survey



Baroness Kidron is the founder of 5Rights Foundation and a Crossbench Peer in the UK House of Lords. Where she has been a world leading advocate for digital regulation and accountability on behalf children and young people. She is known as the architect of the Age Appropriate Design Code, which prompted a radical redesign of digital products and services to protect children's safety and privacy. Baroness Kidron is a Commissioner on the UNESCO Broadband Commission for Sustainable Development, a member of the Global Council on Extended Intelligence and member of the Advisory Council for the University of Oxford's Institute for Ethics in AI. She has recently worked with bipartisan lawmakers to introduce the California Age Appropriate Design Code, a significant advance in US regulation.

Foreword

Beeban Kidron

This volume of essays offers refreshingly diverse perspectives on the state of education data. Overwhelmingly, we learn that, if the goal is to maximise the benefit for children in processing their education data, you would not start from here.

The scope of the essays is as broad as the data gathered - from the fingerprint in the school lunch queue; the lack of clear benefits to children of hotly promoted EdTech tools; the extraordinary obfuscation about what is collected, who owns what, and where it goes next; what education data processing is damaging and what would 'good' look like. Encouraged to imagine a better world, several authors tackle the failure of government and regulators to grasp the enormity of the issue while allowing an increasing role for private companies in school settings. Worryingly, it appears convenient to accept services that may or may not work and that most certainly gather data that are so intimate and yet shared so widely that it might impact on the outcomes of a child for a lifetime.

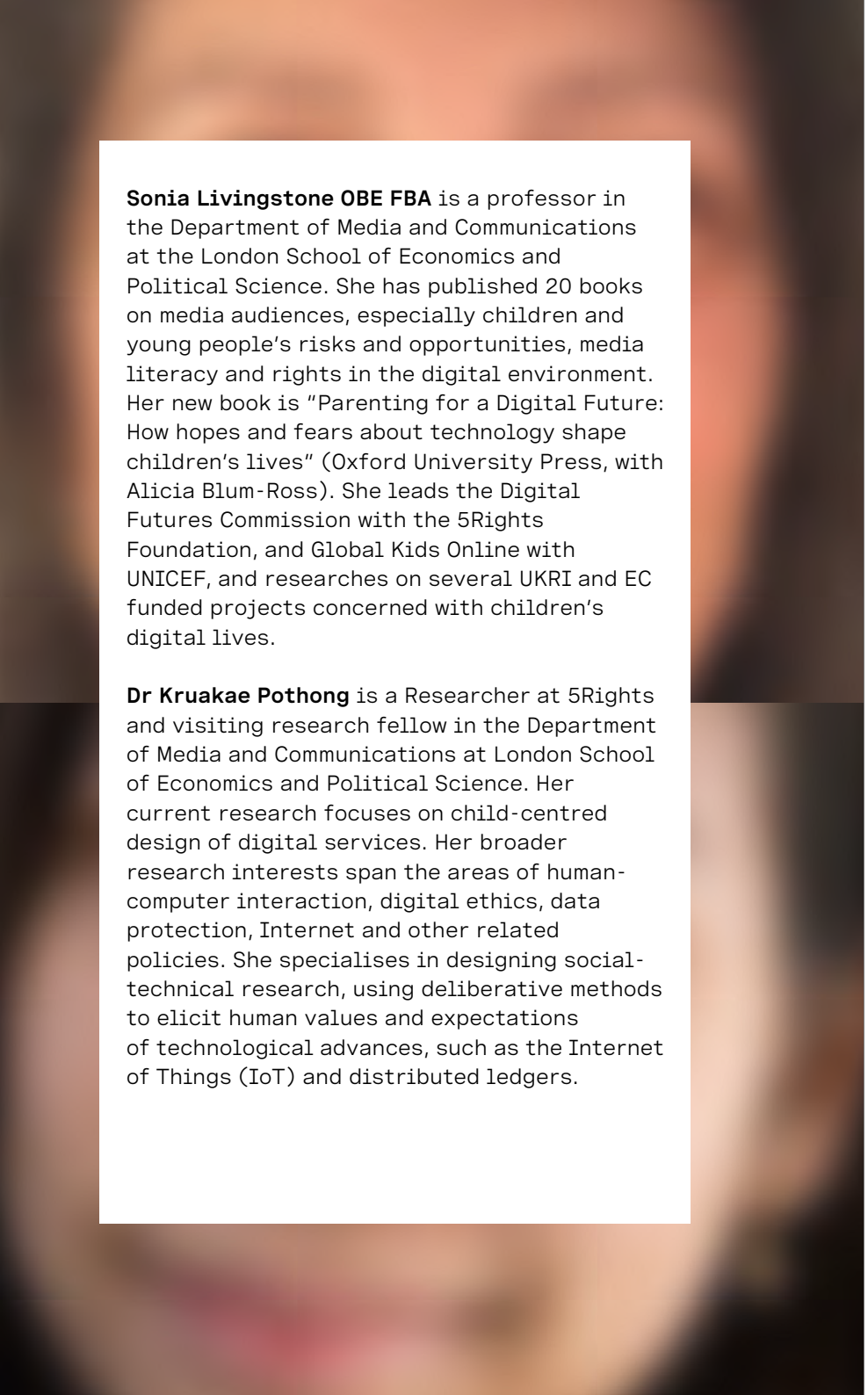
Some essays look at technological and social models that might give more agency to teachers and parents, others look more profoundly at what it would be for a child to be the ultimate owner of their own data. All agree that doing nothing is not an option.

Falling out of these pages is fair amount of frustration that we have not found a better way of unleashing the power of data processing to benefit children. There is no doubt that if technology was focused on wellbeing, learning, best interests (of the child and broader educational community) our ability to harness positive data driven outcomes would be transformed. Simultaneously, a drumbeat throughout the essays is that, unless those who have the power to insist on a more equitable system use that power effectively, we are on an inevitable path in which children at school are fodder for purely commercial interests.

The Digital Futures Commission has spent nearly three years looking at this issue and will in the New Year publish its cumulative findings in *A Blueprint for Regulating Education Data*. This blueprint will encapsulate the voices of school leaders, teachers, parents and caregivers and, of course, children, all of whom who point at the extraordinary asymmetry of power between them and the EdTech sector. It will therefore serve as a challenge to government to ensure that business innovates in the best interests of children.

Thanks are due to each of the contributors for giving so generously their time and wisdom, to Professor Sonia Livingstone and Dr Kruakae Pothong for working with the contributors on their essays, to Natasha Jetha of 5Rights Foundation who oversaw the publication of this volume and to the DFC Commissioners who provide a context and guidance for our work. As ever, our greatest thanks to the young people, their parents and teachers for their extraordinary engagement and support throughout the Commission's work.

Baroness Beeban Kidron OBE
Founder, 5Rights Foundation



Sonia Livingstone OBE FBA is a professor in the Department of Media and Communications at the London School of Economics and Political Science. She has published 20 books on media audiences, especially children and young people's risks and opportunities, media literacy and rights in the digital environment. Her new book is "Parenting for a Digital Future: How hopes and fears about technology shape children's lives" (Oxford University Press, with Alicia Blum-Ross). She leads the Digital Futures Commission with the 5Rights Foundation, and Global Kids Online with UNICEF, and researches on several UKRI and EC funded projects concerned with children's digital lives.

Dr Kruakae Pothong is a Researcher at 5Rights and visiting research fellow in the Department of Media and Communications at London School of Economics and Political Science. Her current research focuses on child-centred design of digital services. Her broader research interests span the areas of human-computer interaction, digital ethics, data protection, Internet and other related policies. She specialises in designing social-technical research, using deliberative methods to elicit human values and expectations of technological advances, such as the Internet of Things (IoT) and distributed ledgers.

The problem and the potential of children's education data

Sonia Livingstone and Kruakae Pothong,
Digital Futures Commission
and London School of Economics and
Political Science

The data collected from children at or through their participation in school are exponentially increasing in variety and volume. This is partly mandated by government, partly determined by schools, and partly driven by the commercial desires of educational technology (EdTech) companies of all kinds, large and small, national and global, user-facing and business-to-business. Increasingly, children's education data seem indispensable to public policy, planning and practice in education, health and welfare, and in schools, teaching, learning and assessment, safeguarding and administration. Meanwhile, commerce thrives on data - for research and development, advertising and marketing, and for many other valuable purposes within today's highly profitable data ecosystem.

Whose interests are served by the intensifying 'datafication' of education and childhood? Datafication - the quantification and analysis of human activity - is increasingly informing public and private sector decision-making (Mayer-Schönberger & Cukier, 2013). The economic interests in data-driven EdTech are considerable, fuelled by the UK Government's economic investment in the EdTech sector (DfE, 2019) and by the commercial ambitions of the rapidly growing

global EdTech industry. The political interests are more subtle and diverse, encompassing efforts to shape the nature of education itself as well as the role of the private sector in public provision. These, in turn, have been fuelled by the demand surge for technologies to support remote learning during the COVID-19 lockdown (Walters, 2021), consolidating the appeal of quick-fix technological 'solutions' to society's problems. Meanwhile, the public interest – including the interests of educators, the wider society and children in particular – is surprisingly little examined, even though grand claims abound about the transformative potential of technological innovation for education.

With daily news headlines announcing data breaches and cybercrime, experts debating 'surveillance capitalism' and algorithmic discrimination (Zuboff, 2019) and science fiction predictions of a society run by robots, it's easy to become dystopian. Yet, innovation continues apace, the public is unwilling to give up its tech, and children themselves relish their digital expertise and agency. At the heart of this dilemma is trust, and the need for a viable mechanism for building trust (Edwards, 2004). The Digital Futures Commission seeks to transcend the polarisation between technological optimists and pessimists by opening up a space for dialogue and deliberation about children's data-driven education futures. This space draws on the experiences of 'insiders' and critical 'outsider' perspectives from academia, industry, civil society and those working directly with children/data/schools. It must be an inclusive and creative space, for, in the short history of our digital society, certain views and interests have quickly come to dominate, closing down possibilities for independent analysis and fresh thinking.

In this space of dialogue and deliberation, it is often easier to diagnose problems than to identify what 'good' looks like. But it is vital to find ways to ensure that data-driven EdTech benefits children, especially since the technological infrastructure on which a digital society relies is privately owned. The UK's history of socioeconomic inequalities in education has already resulted in highly stratified childhood outcomes, which uses of technology tend to exacerbate (Helsper, 2021): can this be overcome or ameliorated? And its

history of unresolved debates over the very purposes of education has left society ill prepared to assert child-centred pedagogies over the instrumental approaches preferred by EdTech: can civil society rethink and redouble its advocacy? Regarding children's education data, key questions include:

- What data are collected from children at or through their participation in school, why, and how are they used?
- How can we share data in the public interest, including to support children's learning or welfare, without undermining their privacy?
- Do uses of education data privilege some children over others, and can we design innovations specifically for those who are disadvantaged?
- Should we better regulate, or differently incentivise, the EdTech market to benefit children's education without commercially exploiting them?

The Digital Futures Commission is grounded in a clear human rights framework, namely, the *United Nations Convention on the Rights of the Child* (UNCRC, see UN General Assembly, 1989). As General Comment No. 25 by the UN Committee on the Rights of the Child asserts, children's rights in relation to the digital environment require efforts on many fronts to mitigate risks, optimise opportunities and meet new challenges. And many of these efforts, in turn, demand critical attention to data. Thus far, we have found that child rights experts have paid too little attention to data, and data experts have paid too little attention to children and their rights. Meanwhile, educators and education policy often attend more to the digital products and services that can support learning than to the data processed by these technologies or the interests thereby served.

To advance the debate, drawing on the best available evidence and ideas, we invited essays from experts, including the data protection regulator, academia, private sector, non-

governmental organisations and civil society. Within the broad remit of examining the potential for beneficial uses of children's education data, each contributor was free to define the challenge as they saw fit. Some prioritise academic sources; others practical experience or professional insights. Some take a deliberately neutral stance; others are more critical, or political. Together, we believe they make a unique contribution towards a rights-respecting pathway for the uses of education data that benefits everyone.

Competing interests in education data

Education data can yield insights of many kinds. However, with the increasing datafication of children's learning (Lupton & Williamson, 2017; Williamson, 2019), critical questions arise over whose interests are served by processing children's education data. Two concerns have come to the fore. First, public and civil society bodies are being prevented from using education data in children's best interests by risk-averse data protection regulation or bureaucratic practice. For example, even de-identified data is rarely shared in circumstances that would help a child or children at risk. Second, the EdTech sector finds itself relatively free to use even personally identifiable and sensitive data from children to pursue its commercial interests. This is because its complex data ecosystems are highly opaque, and its powerful players easily dwarf the capacity of a school to negotiate or even grasp the scale of their operations. The irony of this situation is painful, and children are doubly the losers.

Our first pair of essays set out how greater data sharing could improve a host of child protection interventions. Indeed, **Mark Mon-Williams**, **Mai Elshehaly** and **Kuldeep Sohal** argue that, by combining datasets across institutions to piece together the needed information to warrant individual interventions, high-profile instances of systematic social care failures resulting in child deaths may have been prevented (Butler, 2021). They explore the potential of connected data to target efforts to mitigate risk and disadvantage and overcome the problematic fragmentation among services meant to safeguard children. The authors' telling case studies illustrate how linking education and health datasets, combined with

intersectional indicators of inequality, have informed policy and practice - for instance, providing for young children with undiagnosed autism - in ways not otherwise possible. Recognising data protection risks and potential privacy infringements of creating ever-larger and more centralised databases about identifiable children, Mon-Williams et al. commit to the co-production of acceptable solutions with affected communities. While this adds to a project's workload, it also lightens it by gaining community insights and ensuring community trust.

Leon Feinstein makes the case for state-mandated data collection from the most vulnerable children - to provide for them, as is their right, and to hold government to account for so doing. His case study of the lack of robust and comprehensive data on children with insecure immigration status shows that without such data collection these children are invisible to the system that is meant to support them. Hence their needs go unmet. Nor can society analyse the drivers of children's problems or evaluate the interventions designed to improve their situation. Nonetheless, recording children's immigration status at school and then sharing it with the Home Office or other government agencies has proved controversial.

Acknowledging the risk to the individual of sharing sensitive personal data, Feinstein advocates sharing only de-identified, aggregated data for explicit public purposes via secure services such as the Office for National Statistics' (ONS) Five Safes framework. Also important are data ethics: this means taking seriously children's right to be heard, including in practices of data collection and use, and weighing these according to a rights-based framework with their best interests, individually and collectively.

Yet, the business models that drive EdTech and education data processing are not designed to meet these concerns. Indeed, they are attracting considerable concern for pitting commercial interests against children's best interests, as the following two essays examine. **Michael Veale's** analysis of the vertically integrated business models of the major players - combining hardware, operating systems, cloud services and educational platforms - reveals how EdTech businesses far

more than public, educational or child rights considerations set the standards and determine the rules of the game for the education system. And it is a long game they are playing, locking students early into particular tech practices and norms, providing schools with 'free' systems with profitable add-ons from which it is difficult to extricate themselves, and shaping the offer of content vendors to fit particular platform functionalities over others. Meanwhile, competition law and data protection regulation focus on consumer protection, which fails to take account of the particular needs of the education sector.

Alternative approaches exist, Veale suggests: more collaborative EdTech systems, national procurement frameworks, open source technologies and community-based projects – although these are difficult to scale or sustain, especially at low or no cost. Even these can be appropriated by major platforms able to adjust to and profit from diverse circumstances. But the increasingly global financial power brokers behind the EdTech brands already embedded in UK classrooms exert a very different influence, as **Huw Davies**, **Rebecca Eynon**, **Janja Komljenovic** and **Ben Williamson** examine. Crucially, the major investors in EdTech are generalists or tech evangelists rather than education experts, and their decisions are financially motivated.

Digital education platforms, Davies et al. show, play a crucial role in connecting young users to the surveillant and extractive data economy, guaranteeing what's seen as a reliable revenue stream from cradle to grave. EdTech investors are also political actors promoting normative educational futures in which learning is conceived as on-demand, personalised, lifelong and provided at scale via so-called 'weapons of mass instruction'. Such visions prioritise efficiency gains and drill-and-skill over deep or child-led learning, encouraging external rather than intrinsic rewards and profoundly disintermediating the school as the public education system relies on EdTech platforms and companies appeal directly to parents and caregivers.

The struggle to make education data serve children's best interests is not only fought in national policy circles but also the everyday life of families and schools. Education data is also occasioning plenty of trouble here, undermining children's rights in ways examined in the next section.

The trouble with data

Those working with data in practical settings are also raising the alarm about the complexities of education data and the difficulties of ensuring data underpin rather than undermine children's needs and rights. Concerned that the everyday practices of schools now contribute, however inadvertently, to unregulated and risky data lakes, even data swamps, **Heather Toomey** documents a host of easily overlooked problems that demand rectification. Careful not to blame already over-pressed schools for 'failing' in their arguably impossible task, she highlights ways in which school cultures contribute to the datafication of childhood.

Teachers, administrators, safeguarding officers and other professionals set out to be conscientious in complying with regulations and respecting children's rights. But they are busy, rushed, under-resourced, lacking relevant guidance or training, ever hopeful of finding a useful shortcut or workaround, and tempted to follow the usual practice rather than think things through from first principles. Dealing with EdTech can too easily take teachers' attention from their primary task of educating the children in front of them. Moreover, not only is the complex data ecology they must navigate hardly transparent, but the very EdTech companies that pose schools with difficulties also proffer 'solutions' that can supposedly ease their path. And yet, broader uses of education data in children's best interests are on offer – Toomey gives the example of how safeguarding needs may be met by interagency data sharing. Whether this can be enabled without further commercial exploitation of children's data remains to be seen.

Education data may, for multiple reasons, often unintentional, enable discrimination, exclusion or inequality on multiple grounds, including gender, ethnicity, sexuality, disability, refugee status and more. Arguably, schools are the institutions to redress rather than perpetuate inequalities among children. Yet research reveals many biases, inaccuracies, distortions and other harms in the operation of data-driven and automated technologies that amplify and

accentuate pre-existing sources of disadvantage in society. **Najarian Peters** examines the 'dirty data' processes that discriminate against Black children in the USA and UK, now perpetuated through EdTech. She charts a range of adverse outcomes from educational practices that result in Black children being recorded as less innocent or vulnerable and more aggressive or disruptive than their white classmates. No wonder Black parents more often choose home education for their children. Other than opting out, what are the prospects of righting the wrongs in education data and its uses? Peters calls for fair data practices, data subject rights, improved regulation and recognition of the Black Data Traditions by which Black communities seek to preserve their rights.

A common retort is that their parents have signed the necessary permissions with the school, and they are, in any case, responsible for their children - and their children's data. Yet parents, too, are little informed about data-driven EdTech or able in practice to exercise their responsibilities. **Rosalind Edwards**, **Val Gillies** and **Sarah Gorin** commissioned a nationally representative survey of UK parents, which found that while parents were aware of data collected from their children, they were less aware of the uses to which data are put, including data sharing across agencies. Once made aware, parents expected to be asked for their consent since only half trusted public services - including schools - to use information in children's best interests. Inequalities matter - parents from relatively disadvantaged or discriminated-against groups, especially Black parents and lone parents, considered data linkage less legitimate, were less trusting of agencies and had more experiences of problematic uses of data regarding their child. Edwards et al. call for a public moratorium on data linkage while a meaningful national dialogue is held to ensure legitimacy.

Yet, far from any moratorium, the quantity and range of data collected from children in a typical day is escalating, as **Jen Persson** maps in her *State of data 2020* report. That report highlighted the struggles of UK schools to manage education data and comply with an at-times unclear or inadequate data governance landscape (Defend Digital Me, 2020). In her essay in this volume, she grounds her energetic call for a better system of education data processing within a holistic child rights framework. Different types of data and data processing are linked to different concerns and rights, few of which attract sufficient attention from the duty bearers - government, regulator, schools and businesses - charged with respecting children's rights. Yet, she points out, at school, children have particularly little agency to determine what happens to them or their data. By contrast with many business-to-consumer uses of data, rarely can children consent or withdraw consent from particular EdTech uses by the school. Nor have they opportunities to exercise their data subject rights or to be consulted on the school's education data policy.

The call for improved regulation is mounting on all sides, and we examine this next. It is notable, however, that the UK Government has recently proposed an alternative approach (Data Protection and Digital Information Bill, 2022). Whether the proposals amount to smarter regulation or fewer data protection constraints on the market remains open to debate. Below we consider the value of better regulation before alternative approaches to respecting children's rights in relation to data-driven EdTech.

The value of better regulation

Much of the work of the UK's data protection authority focuses on preventing risky and unlawful sharing of personal data, including data from children, for which Age Appropriate Design Code applies. However, tackling the thorny question of what 'good' data sharing looks like, **Stephen Bonner**, **Melissa Mathieson**, **Michael Murray** and **Julia Cooke** from the Information Commissioner's Office also recognise the risks of not sharing children's data when such sharing would be in children's interests for early intervention to prevent harm. To balance the risks of sharing with those of not sharing, in accordance with both the Age Appropriate Design Code (or Children's Code) and their Data Sharing Code of Practice, they advocate a seven-point strategy: build on existing best practices; adopt a multistakeholder approach; ensure organisational accountability for data protection; prioritise data minimisation; promote transparency both in schools and from digital businesses; support confident data sharing

through training and regulatory compliance; and underscore the importance of data accuracy and data subject rights.

Now that the UK is reconsidering its adherence to Europe's General Data Protection Regulation (GDPR), what can be learned from European and international legal and human rights frameworks? Or even from data protection in other sectors? For **Ingrida Milkaite**, the critical challenge is less technological innovation than EdTech's business model, which fuels increasing commercial data processing more than educational goals. Given that the promised educational benefits remain unproven, and with costs to children's rights also likely, the Council of Europe advocates the precautionary principle, especially regarding children's sensitive and biometric data. The considerable power imbalance between children and even schools in relation to EdTech businesses cannot be redressed through digital literacy education alone, important though this is. Consequently, Milkaite calls on national data protection authorities to strengthen their actions to enforce existing regulations, take a precautionary approach to technological innovation and underpin children's rights in all contexts, including education.

One of the earliest laws to protect student privacy is the US Family Education Rights and Privacy Act (FERPA, 1974), passed reactively when state more than commercial misuse of education data was occasioning concern. Half a century on, can other countries learn from the US experience? FERPA's protections include the right to correct inaccuracies and prevent unauthorised sharing. **Amelia Vance** sets out the rationale for sector-specific protections - in this case, that students are required to attend school, that children are uniquely vulnerable to privacy harms, and that data processing is an integral part of school responsibilities. Yet, Vance argues, FERPA contains so many exceptions that, in practice, it has proved confusing and weak. Also problematic is its reliance on parental consent as a mechanism for data collection and sharing since parents may give consent ill-advisedly or against their child's interests for a host of practical reasons. Viewed from the UK, which has already benefited from the provisions of the GDPR still lacking in the USA, the main lesson appears to be to avoid the mistakes made with FERPA.

Education is but one focus of innovation in our digital society. **Riad Fawzi** examines how the financial services sector has responded to financial technology (FinTech), with its promise of more diverse, tailored and affordable consumer services, but suffering from low trust among the public. As with EdTech, FinTech has harnessed modern technology to innovate business-to-business and business-to-consumer services. Yet both state and self-regulatory efforts are more advanced than appears to be the case for EdTech, resulting in both a more mature regulatory ecosystem for FinTech and greater oversight and transparency. Yet, for both sectors, greater efforts are needed to merit public trust: Fawzi advocates the combination of regulation and self-regulation, as well as standards for security, privacy and digital identity, a commitment to customer service and, last but not least, provision of a truly valuable service - in this case, EdTech that meets children's genuine educational needs.

In addition to complying with regulation, what else can and should EdTech businesses do? Whether or not regulation is an enabler or a brake on innovation, EdTech is innovating fast, and the drivers are not only commercial but also social and educational. So what better digital products and services can be hoped for, and are they in evidence?

Seeking design solutions

The most often mentioned benefit of education data is personalised learning - the promise of providing exactly the teaching materials that each child needs at just the moment when they need them. It is hoped that personalised learning can motivate, enable and reward all children as they learn while relieving teachers from the effort to support each child individually - and the guilt of attending to 'difficult' children while others lose out. **Natalia Kucirkova** weighs the evidence for the added value of deploying often-automated, data-driven, adaptive EdTech in the classroom, finding this not only weak but, where it exists, mainly focused on drill-and-skill learning.

Two problematic design principles underpin much of this technology - exponential growth (the idea that more data is always better) and recommendation systems that promote

more similar content. But since educational theories instead value teachers' knowledge of their pupils and a diversity of learning resources, better principles would minimise data collection, keep the 'human in the loop' and recommend multiple alternative opportunities. Redesigning EdTech with educational principles and learner agency at its core will require a substantial rethink by businesses.

One group for whom the benefits of data-driven technologies are eagerly anticipated is disabled children (Alper, 2017). Using the phrase 'disabled children' to emphasise the social theory of disability, namely that any deficit lies not in the child but in society's provision for all children, **Sue Cranmer** and **Lyndsay Grant** argue that, while there is evidence of digital technologies being used to benefit disabled children's learning, traditionally such technologies have not been data-driven. When it comes to data-driven EdTech, there are growing critical concerns regarding the biases, stigma and inequalities that can affect this group from automated uses of education data. Is there scope for empowering data-driven interventions to supplement long-standing efforts toward inclusive education?

The authors offer five suggestions to this end: systematic data collection to inform and target government actions; personalised learning provision that responds to accessibility or other disability-related difficulties; monitoring progress to identify when greater support is required; sharing data with relevant agencies for effective decision-making; and using data to represent diversity and redefine norms. In each case, however, they note potential risks as well as the lack of robust evidence for beneficial outcomes. They also note how rarely disabled children are themselves consulted or provided with genuine choices.

It may not be obvious that what education needs is a greater focus on students' emotions. But through the advent of 'affective computing' or 'emotional AI' – or what **Andrew McStay** terms 'automated empathy' – an industry has grown to monitor and respond to children's emotions at school. The technology now exists to record children's facial expressions, keyboard presses and bodily movements and analyse the resulting data to segment, profile and score children on their attention,

interest, uncertainties and feelings during learning. And already on the horizon are educational uses of automated biometric empathy in the metaverse. While not yet in operation in UK schools, McStay examines these developments as part of the broader agenda of personalised learning.

His essay sets out three critical concerns. First, he argues that the technology is inaccurate, being underpinned more by pseudoscience than robust evidence. Second, it infringes children's rights to privacy, including freedom from surveillance, profiling and commercial exploitation. Third, it is unlikely to work in practice, for not only does it not meet a genuine educational need, but it is likely to generate unintended and adverse consequences as children seek to evade such scrutiny of their every move.

What of those working in EdTech itself? By design solutions for safety, privacy and security are currently being sought in multiple domains, including in education, bringing into focus the role of designers and developers in protecting children's interests. **Ari Beckingham** and **Larissa Pschetz** rethink the assumptions that underpin much EdTech design, concerned that too often design is motivated to maximise user attention rather than encourage deep understanding holistically across formal and informal learning contexts. Instead, their research attends to the pace of learning, embedding ethical data practices in technology (for example, augmented reality [AR]) designed to encourage children to pay careful attention to the world around them and engage reflectively in their learning process. Such research seems to herald a promising alternative to the dominant focus of EdTech, foregrounding attention to pedagogy and inviting deliberation over the educational vision that data could and should serve.

Rethinking data futures

Without public trust in EdTech's ambitions, policies and practices, scepticism about commercial uses of education data will likely grow rather than diminish. The final section of this volume is the most radical, exploring technical and market-led alternatives to privacy-invasive systems of data harvest among data oligopolies. In other words, rather than placing ever-greater reliance on the regulator, can a new ecosystem of

trusted data management technology (or a personal data store) and a new data management service (or a data trust) offer data subjects more effective control? The concept of data trusts as a solution for privacy protection was introduced over a decade ago (Edwards, 2004). Yet, data trusts as technical and market solutions are only now gaining traction with concrete proposals coming to the fore, as discussed in the essays in this section: the authors explore the hope that data trusts can help to realise the benefits of sharing education data for the public, including children.

Expectations of the data protection regulator in a digital society are becoming impossible, not least because people want personalised services. But, as **Roger Taylor** argues, people wish to be protected not from personalisation in and of itself, but from harmful or exploitative use of data by providers of data-driven services. His proposed alternative is to separate the management of data (or data stewardship) from the provision of data-driven services and applications via the creation of data trusts as a service. As a service, data trusts manage individual users' data on their behalf, and must be governed independently in ways that respect the interests of individual data subjects with other public and private sector benefits likely to follow, as emerging good practice cases suggest. As several authors note, however, the political, policy and business challenges are notable.

Defining data trusts as legal entities that provide independent stewardship of data, **Jim Knight** and **Timo Hannay** base their optimism on the experience of founding a data analytics company to examine the effects of remote learning on children's outcomes during the pandemic. This taught them that however valuable the insights from education data, these cannot be obtained when public trust in technology companies to manage data fairly is dropping (Wisniewski, 2020). Trust is of particular importance in relation to big data and artificial intelligence (AI), where it is implausible that the public can understand and scrutinise the uses of their data. This applies especially to children and those responsible for them. Avoiding simple solutions, Knight and Hannay are careful to argue for data trusts as part of a wider mix of legislative, self-regulatory and other actions to promote the common good in a digital world.

Also responding to calls for ever tighter data protection regulation, which he sees as resulting from fears linked to surveillance capitalism, **Bill Thompson** advocates an innovative technical approach to data management - the personal data store. With several different forms available, and more experimentation underway, the heart of this alternative is that the data owner - potentially, the child - stores their own data and controls access to it. The personal data store, he suggests, could be embedded in a trusted public service data ecosystem such as that developed at the BBC. Although the technical potential has existed for a while, with interest now growing in response to the extensive datafication of childhood (Barassi, 2020; Mascheroni, 2020), challenges remain, including gaining informed consent from minors, data breaches and the difficulty of rectifying poor choices. Nonetheless, with greater transparency and user control on offer, there are grounds for optimism.

Jun Zhao also addresses the potential of data trusts in calling for a new decentralised data governance structure for children's data and data sharing. Recognising the host of data governance problems set out by the Digital Futures Commission (Day, 2021), and concerned not to burden individuals with excessive vigilance and comprehension regarding the commercial data ecology, Zhao joins those seeking a technical rather than a regulatory solution, whether through data commons, data trusts or data cooperatives. She examines existing and hypothetical cases in education to highlight how a data trust, and its trustees dedicated to acting in children's interests, can provide a needed intermediary between schools, students and EdTech companies. Can this model work at scale? If so, what legal framework is required, how might it be funded and who will be liable if something goes wrong? As new questions arise, the space for debate over children's education data is expanded, and the potential for rights-respecting approaches is kept alive.

A rights-respecting approach to children's education data

This volume does not position EdTech - or the data it generates - as either good or bad in and of itself. Instead, we emphasise

human actions and values in determining how technological design and systems, business logics, communities of practice and other socioeconomic and political factors (Ihde, 2002; Arthur, 2009) 'serve human beings in the accomplishments of their individual and collective purposes' (Buchanan, 2001, p. 9). Or fail to serve them. By examining the forces shaping children's learning lives, we hope to identify the steps needed to better realise their rights in a digital world.

Talk of rights often focuses on particular areas of children's lives, but the UNCRC insists on a holistic approach to children's rights to participation, education, information, privacy, play and fullest development, among other rights, for rights cannot be ranked. Crucially, the UNCRC emphasises the child's best interests as a primary consideration. This outweighs commercial interests and demands a comprehensive assessment of the needs and rights of each child and children collectively.

Also significant in the UNCRC are what is called the general measures of implementation. These specify how the state should act as duty bearers, taking all necessary steps, including ensuring that business and other actors meet their responsibilities to children. Indeed, it is notable that many authors in this volume have paid more attention to the organisations that process education data than to the data flows. They have emphasised the importance of establishing appropriate legislation and implementing it effectively to prevent harm, improve provision and participation, and stimulate innovation that opens up new opportunities for society, including children.

This volume builds on the Digital Futures Commission's recent critique of the UK's governance of children's education data (Day, 2021), followed by a multistakeholder roundtable discussion (Livingstone et al., 2021), a deep dive into the data-related challenges faced by schools (Turner et al., 2022), consultations on children's hopes and concerns for their digital lives (Mukherjee & Livingstone, 2020), and sociolegal analysis of the problematic practices of prominent EdTech companies (Hooper et al., 2022). Here, our purpose is to look forward.

This volume offers critical, practical and creative reflections that can guide society in harnessing education data for good. It weaves together often-disconnected policy conversations about technologies as a means to support education (UN, 2022) with regulatory, market and technical solutions for data governance. It highlights the fundamental principles that should guide state and business activities across the essays – transparency, accountability, legitimacy, fairness and non-discrimination, appropriate remedy, consultation with those affected, ensuring public trust, and innovation in children's best interests. These principles have been widely overlooked in relation to children's education data and it is time to prioritise them. The Digital Futures Commission is proud to have brought together these insightful essays. These will surely inform and advance the public and policy debate. They also provide a sound basis on which to develop our forthcoming blueprint for regulatory and practical change to ensure that future uses of education data serve children's best interests.

Alper, M. (2017). *Giving voice: Mobile communication, disability, and inequality*. MIT Press

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon & Schuster

Barassi, V. (2020). *Child data citizen: How tech companies are profiling us from before birth*. MIT Press

Buchanan, R. (2001). Design research and the new learning. *Design Issues, 17*(4), 3-23

Butler, P. (2021). Anger over child deaths should not trigger knee-jerk overhaul of social care policy. *The Guardian*, 19 December

Data Protection and Digital Information (UK) Bill (2022)

Day, E. (2021). *Governance of data for children's learning in UK state schools*. Digital Futures Commission, 5Rights Foundation

Defend Digital Me (2020). *The state of data 2020: Mapping a child's digital footprint across England's state education landscape*

DfE (Department for Education). (2019). *Realising the potential of technology in education*

Edwards, L. (2004). Reconstructing consumer privacy protection on-line: a modest proposal. *International Review of Law, Computers & Technology, 18*(3), 313-344

FERPA (Family Educational Rights and Privacy Act). 20 USC § 1232g (1974)

Helsper, E. (2021). *The digital disconnect: The social causes and consequences of digital inequalities*. SAGE

Hooper, L., Day, E., Livingstone, S., & Pothong, K. (2022). *Problems with data governance in UK schools: The cases of Google Classroom and ClassDojo*. Digital Futures Commission, 5Rights Foundation

Ihde, D. (2002). *Bodies in technology* (Vol. 5). University of Minnesota Press

Livingstone, S., Atabey, A., & Pothong, K. (2021). *Addressing the problems and realising the benefits of processing children's education data: Report on an expert roundtable*. Digital Futures Commission, 5Rights Foundation

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society, 19*(5), 780-794

Mascheroni, G. (2020). Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology, 68*(6), 798-813

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt

Mukherjee, S., & Livingstone, S. (2020) *Children and young people's voices*. Digital Futures Commission, 5Rights Foundation

Turner, S., Pothong, K., & Livingstone, S. (2022). *Education data reality: The challenges for schools in managing children's education data*. Digital Futures Commission, 5Rights Foundation

UN (United Nations). (2022). More than 150 ministers of education discuss key elements for transforming education ahead of September Summit. Press release, 30 June

UN General Assembly. (1989) *United Nations Convention on the Rights of the Child*. Treaty Series, Vol. 1577, p. 3, 20 November

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

Williamson, B. (2019). Brain data: Scanning, scraping and sculpting the plastic learning brain through neurotechnology. *Postdigital Science and Education, 1*(1), 65-86

Wisniewski, G. (2020). Losing faith: The UK's faltering trust in tech. *Edelman*, 30 January

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books

COMPETING INTERESTS IN DATA EDUCATION

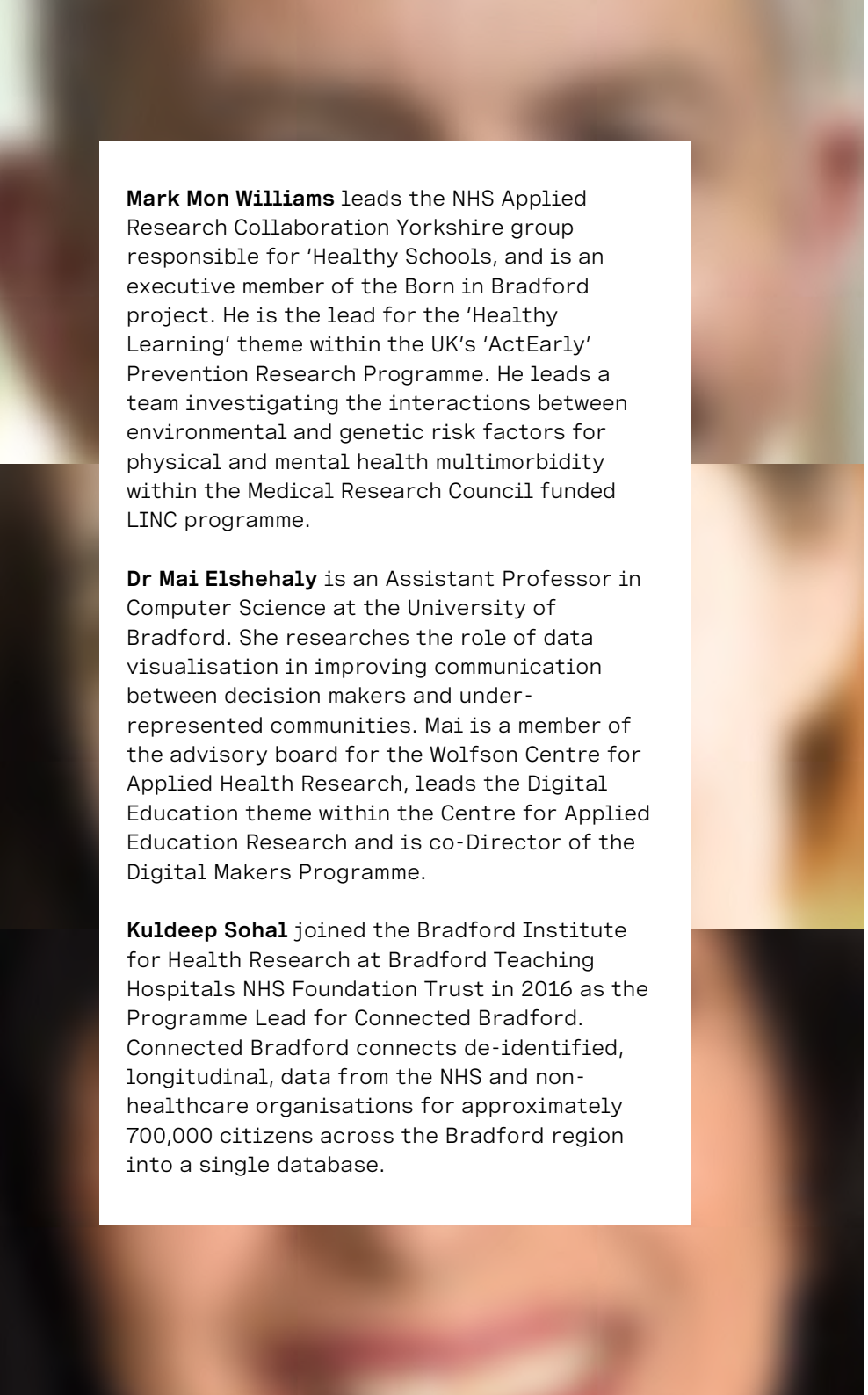
Anything I need to learn, I can learn from YouTube (Boy, 9)

I get work sent through from my teachers [through Microsoft Teams] so I can keep up if I have missed a class (Girl, 14)

Times Tables Rock Stars helps me practise my times tables in a fun way (Girl, 9)

Satchel One is useful for some things but I don't like it (Boy, 11)

Digital Futures Commission (2022) EdTech survey



Mark Mon Williams leads the NHS Applied Research Collaboration Yorkshire group responsible for 'Healthy Schools, and is an executive member of the Born in Bradford project. He is the lead for the 'Healthy Learning' theme within the UK's 'ActEarly' Prevention Research Programme. He leads a team investigating the interactions between environmental and genetic risk factors for physical and mental health multimorbidity within the Medical Research Council funded LINC programme.

Dr Mai Elshehaly is an Assistant Professor in Computer Science at the University of Bradford. She researches the role of data visualisation in improving communication between decision makers and under-represented communities. Mai is a member of the advisory board for the Wolfson Centre for Applied Health Research, leads the Digital Education theme within the Centre for Applied Education Research and is co-Director of the Digital Makers Programme.

Kuldeep Sohal joined the Bradford Institute for Health Research at Bradford Teaching Hospitals NHS Foundation Trust in 2016 as the Programme Lead for Connected Bradford. Connected Bradford connects de-identified, longitudinal, data from the NHS and non-healthcare organisations for approximately 700,000 citizens across the Bradford region into a single database.

Connected data for connected services that reflect the complexities of childhood

Mark Mon-Williams, Mai Elshehaly and Kuldeep Sohal, University of Leeds

The COVID-19 pandemic has revealed myriad problems with the systems, services and processes that are supposed to protect and support children and young people. We argue that a fundamental problem relates to fragmented public services, including education, health and social care. Data science provides analytical systems that can allow us to address this connection problem and facilitate a holistic approach to meeting the needs of children and young people. In this context, the sharing of data across services is an essential step in creating efficient systems capable of safeguarding children effectively, providing timely support for vulnerable children, removing structural inequalities and enabling a whole-system approach to tackling the wider determinants of physical and mental health, educational outcomes and social mobility.

Here, we present three case studies drawn from the connected routine datasets of over 13,500 children within the Born in Bradford (BiB) project to illustrate the benefits of data sharing. These provide the rationale behind the creation of the West Yorkshire Integrated Data Engine for Analytics (IDEA) centre - dedicated to connecting routine datasets and applying the power of science to develop systems that are fit for purpose.

Context

In December 2021, *The child of the north* report (Pickett et al., 2021) provided a harrowing account of the appalling situation facing huge swathes of children growing up in poverty within the North of England (with the same problems affecting disadvantaged children regardless of where they live). The resulting cost to families is enormous, but so is the financial public burden that arises from our collective failure to support vulnerable children. It is extremely difficult to believe that anyone could digest *The child of the north* report and not be motivated to join the voices demanding radical change to our systems and services. We suggest that connected datasets can allow us to start understanding the root causes of the problems (by revealing the intersections and interactions between different factors such as health, education and social care), and can enable the genuine multiagency responses that are required if the complex needs of disadvantaged children are to be addressed effectively.

The consequences of failing to use information effectively in public service delivery are catastrophic. The Children's Act 2004 requires *a serious case review* after the death of a child where abuse or neglect is known or suspected. It is rare to read a serious case review that does not conclude that failure to share information across professionals and organisations was a contributing factor to the death. In December 2021, the tragic cases of Arthur Labinjo-Hughes and Star Hobson (Child Safeguarding Practice Review Panel, 2022) provided a depressing reminder that our systems and processes are antiquated and do not take advantage of scientific advances in data science. Arthur and Star are the personification of a wide malaise affecting a multitude of children within the UK's most deprived areas. Our central argument is that we could tackle these problems by connecting the information available through the routine datasets held by the different organisations that have a share of responsibility for the wellbeing of children and young people (e.g., health, education, social care, policing etc.).

We start by defining information as the state of the Universe, relative to an observer, measured as the logarithm of the number of its possible states. In order to avoid a tedious scientific exposition of this definition, we consider information in the context of Wordle, a popular game that illustrates the key points we wish to make. Wordle requires players to guess five letters chosen from the 26 letters of the English alphabet. Wordle selects different letters each day and gives the player six opportunities to guess the five letters. In this description of the game, the player is trying to guess which one of 7,893,600 permutations is the target. Fortunately, Wordle (as the name implies) constrains the task by only using words within the English language.

It is clearly helpful for a Wordle player to use their existing knowledge about the English language when attempting to guess the target word, which brings us to another key feature of Wordle - information is provided after each guess is made. A correct letter in the right location is highlighted in green, whereas a correct letter in the wrong location is shown in yellow. It follows that a good Wordle strategy is to use an initial word that will yield maximum information. This means using the most frequent letters and placing those where they most frequently appear. Thus, there are two possible approaches to playing Wordle. One is to use all available information and enjoy a daily average Wordle performance of between three and four guesses. The alternative is to ignore the available information. Our argument is that the current approach to public service delivery is equivalent to this alternative - and frankly bonkers - Wordle strategy.

The child of the north report establishes what 'bad' looks like when data are not used effectively. Importantly, rich datasets exist, but the information is split across education, health, social care, and so on. The fact that these datasets remain disconnected is deeply concerning as they can describe the intersecting and interacting factors impacting on a child's life. Moreover, the combined data can support the creation of powerful data analytic systems for tackling childhood vulnerabilities. Nevertheless, no organisation within the UK currently uses connected data to address inequality or even to meet its legal responsibilities to protect children and young people. On the basis of first principles, it can be argued that failure to connect information across public services means that we are missing opportunities to better serve children. We

do not need to rely on logical arguments alone, however, as we have accumulated a wealth of empirical data within the Bradford district that show the benefits of connecting data.

Connected routine datasets

Bradford is uniquely positioned to show the power of data as routine administrative records have already been connected through the Born in Bradford (BiB) project. BiB is one of the world's largest longitudinal birth cohort studies and has linked routine data for over 30,000 Bradfordians.† Frequent engagement with the families and children allows us to collect informed consent for continued routine data linkage (e.g. health, social care and education records).

The success of BiB in using connected data has led to the creation of the 'Connected Bradford' database containing the records of citizens across the Bradford District. Connected Bradford combines a number of records, including primary care (e.g., appointment history, prescribing and clinical data), community care (e.g., mental health, school nurse, health visitor interactions), secondary care (e.g., maternity, outpatient), social care, children's centres data, education, housing and benefits, crime, housing data and data from the National Child Measurement Programme (see Sohal et al., 2022).‡

In order to connect the health and education data, we obtained Confidentiality Advisory Group approval for individual data linkage of health records to National Pupil Data education records held by the Department for Education. We used non-unique personal identifiers because the Unique Pupil Number (that identifies each pupil in England) is not available to healthcare organisations.

† The BiB cohort comprises 12,453 women recruited at 28 weeks of pregnancy, who gave birth at the Bradford Royal Infirmary to 13,857 children between the period 2007 and 2011 (see https://borninbradford.nhs.uk). Half of all BiB families live within wards classed among the 20% most deprived within England and Wales and 45% of families are of Pakistani origin. The BiB families provided informed consent for their routine electronic records to be linked and used for scientific purposes. These are supplemented with detailed testing (on measures including physical health, cognitive ability, sensorimotor function, household demographics etc.) on a regular basis, providing one of the richest available descriptions of a population's genotype and phenotype.

‡ For the interested reader, Sohal et al. (2022) provide detailed information on the different datasets, the legal pathways through which linkage occurred, data security and storage, ethical arrangements etc.

We will now use three case studies to demonstrate the usefulness of the resulting connected data (i.e., our examples will focus on linked health and education records).

Glasses in classes

Our first study demonstrates how the existence of a connected dataset enables a proper understanding of the complex factors that contribute towards poor outcomes for children (including poor physical health, mental health, educational attainment and social mobility) - in this instance, poor reading skills.

There are a large number of children in Bradford who fail to learn to read at an acceptable rate (DfE, 2017). The natural response to this situation is to improve school leadership around reading or to provide approaches such as phonics programmes. The BiB data revealed, however, that a fundamental health problem might explain the unsatisfactory levels of reading. The connected data showed that many children identified with an ophthalmic deficit (i.e., they needed a pair of eyeglasses) were not taken to the hospital eye service or the local optometrist despite a letter informing the relevant carer that there was a problem with the child's eyesight. Moreover, the data showed that children with uncorrected eyesight were at increased risk of delayed reading skills (DfE, 2021). These insights were available because the ophthalmic status of the children could be obtained from the health records (i.e., the children's medical records) while the child's reading abilities were available through the connected education data (i.e., information from the Department for Education). This simple example demonstrates the power of connected datasets in flagging important intersections between education and health, showing where we need to address health barriers that impact on education.

Classroom Air Cleaning Technologies (Class-ACT) research

Our second case study relates to the usefulness of connected data when testing interventions targeted at improving outcomes for children. The COVID-19 pandemic highlighted the importance of good ventilation in the prevention of airborne diseases. The pandemic also revealed that a number of classrooms don't have adequate ventilation, which could

potentially increase the risk of a child or teacher contracting an airborne illness. One possible solution to poorly ventilated classrooms is the provision of 'air cleaning technologies'. For example, filtration technologies remove particles from the circulating air including the COVID-19 virus and other pathogens. The technologies also remove particulate matter that can cause asthma and the pollen that can cause hay fever. It follows that - *in principle* - these technologies might decrease illness in children (and teaching staff) and thereby reduce school absences (with all of the educational benefits accrued through increased time in school). However, the use of these technologies involves substantial financial investment, and there is currently little evidence available that the potential benefits will actually translate into real world impact.

The fact that Bradford has linked health and education data allowed the creation of the Class-ACT (Classroom Air Cleaning Technologies) project, where we could conduct a randomised trial and use a combination of health and education data to understand fully the impact on children attending schools fitted with these technologies. Class-ACT allows a holistic investigation into the data on childhood infections available from the health records combined with information on school absences available from the education system. In the absence of the connected datasets, it would only be possible to obtain a piecemeal picture of the potential for COVID-19 transmission to be reduced through fitting air-cleaning technologies within schools.

Identifying children with undiagnosed autism through education records

Our third case study highlights the usefulness of connected data in addressing the problem of undiagnosed autism. There is overwhelming evidence to show that identifying autism in the early years of life has great benefits for the child and their family (e.g., French & Kennedy, 2018). Unfortunately, many children do not have their needs identified until the end of primary school, or sometimes not even until they are in secondary school or beyond (Department of Health and Social Care, 2021). The issue of undiagnosed autism places health and education services under great strain, and creates long-term financial costs that could have been avoided through early action. Many areas have lengthy waiting lists for autism assessment, with children often waiting for many years before they receive the support they need.† Furthermore, societal inequalities are reflected within the autism assessment process, with children from disadvantaged backgrounds waiting much longer than their more affluent peers (Kelly et al., 2019). Notably, children from disadvantaged backgrounds with undiagnosed autism are far more likely to also have additional needs that will require a holistic response from a number of different organisations (Pickett et al., 2021).

The BiB dataset showed that routine educational data can be used to identify undiagnosed autism in children, and tested novel approaches to address the problems associated with this and other developmental disorders. We were able to show that the Early Years Foundation Stage Profile (EYFSP) scores given by teachers at the end of Reception Year (age 4-5) can be used to help identify neurodevelopmental problems, including autism (Wright et al., 2019). Once again, these insights were only made possible because the children's health records (identifying patients with autism) were linked with their education records (allowing us to explore what 'red flags' in education data might be indicative of children at risk of undiagnosed autism).

The data-driven research then led to a study in 10 Bradford primary schools, involving in-school screening of 600 pupils to identify 'at risk' pupils faster and more accurately (Wright et al., 2021). The study identified children who would benefit from a formal autism assessment. A multiagency team, including Child and Adolescent Mental Health Services (CAMHS) and educational psychology services, then attended the relevant schools to help conduct assessments quickly, share information instantly with teachers, parents and caregivers, and facilitate the development of a single support plan.

† For example, NICE Quality Standards for autism stipulate that the wait between referral and first diagnosis appointment should be no more than 13 weeks, but in a government survey just 18% of local authorities in England reported meeting this target (All-Party Parliamentary Group on Autism, 2019).

Building Integrated Data Engine for Analytics (IDEA) centres

These three case studies demonstrate how connected data can allow holistic evidence-based solutions to be implemented when tackling issues related to childhood vulnerability. This is equivalent to using the feedback provided by Wordle when guessing the target word rather than ignoring the information. It is worth emphasising that the current systems for identifying and supporting children with vulnerabilities (e.g., autism) are ignoring the information available across the system because each stakeholder is failing to share data with its partner organisations - despite their shared statutory responsibility for children and young people.

The creation of a connected dataset is essential, but the optimal use of information requires community engagement, intelligent analysis and visualisation of data. This is why we created the West Yorkshire Integrated Data Engine for Analytics (IDEA) centre that brings together experts in data analytics, community engagement, ethics, law, economics and visualisation across Leeds, York and Bradford. Our West Yorkshire IDEA centre is now leading a regional effort to tackle inequality through the Digitally Acting Together As One (DATA 1) programme. This follows the COVID-19 pandemic lifting the lid on the costs of the current fragmented system. For example, the pandemic highlighted the large number of vulnerable children 'under the radar' of organisations with safeguarding responsibilities. The unavailability of connected information during lockdown made coordinating multiagency responses extremely difficult, despite the same families requiring support from multiple organisations. These problems played out against the backdrop of rising inequalities, with service providers finding it increasingly hard to deliver the holistic support needed to address the root causes of many needs (Pickett et al., 2021).

DATA 1 aims to create data analytics tools that can improve service delivery across different providers, including health, education and social care. These tools will: (a) allow early identification of need and (b) enable frontline practitioners to organise efficient and effective multiagency responses to children who would benefit from support. The creation of data analytics tools capable of connecting practitioners will transform public service delivery and improve the support offered to hundreds of thousands of people. Moreover, they will connect policymakers, communities and practitioners, and empower them to tackle the numerous problems that currently plague our society.

The creation of data analytics tools requires us to find solutions to the following challenges: technical issues associated with connecting and visualising data; ethical and legal issues of data protection; engagement with the communities served by the tools; and the imperative of producing tools that can be readily used by practitioners from a range of different organisations.

The technical issues in developing data analytics tools requires us to tackle several challenges, including: integrating heterogeneous data sources; understanding the decision-making tasks and priorities of practitioners from a range of different organisations; and measuring the impact that these decisions can have from the perspective of the communities. This dictates a high level of continued engagement to ensure that the designed technologies benefit and respect these diverse stakeholders and their lived experiences.

Instead of trying to resolve these challenges across a range of different domains, our strategy is to focus on creating data analytics systems that can help clear the queue of children on waiting lists for autism assessment, allow earlier identification of undiagnosed autism and enable children with autism to receive multiagency support as soon as their needs are recognised. Our rationale is that focusing on one specific problem will ensure that we can work through the technical, ethical, legal and practical issues in a manageable manner. It also enables us to coproduce our data analytics solutions with the relevant communities, and allows us to communicate why we are connecting data.

In our opinion, the processes of coproduction and community engagement require the same level of attention as the technical aspects of data science. While we are aware that many people are worried about the misuse of education data information (e.g., Kolkman, 2022), we share their concerns and are passionate about using our connected datasets to help children understand their data rights. Coproduction and

engagement lie at the heart of BiB, and we believe this explains why our wonderful families are content for their data to be linked and used in the manner described here. We have teams who lead work around ethics and legal pathways, and we have developed specific programmes of work that directly support coproduction and engagement. For example, our Digital Makers programme involves the 30,000 young people involved in the next phase of BiB (known as 'Age of Wonder'). Digital Makers is working with the young people's schools to provide digital upskilling and help young people understand their data rights. We capture their voice through a 'Youth Summit' that directly feeds into our research endeavours.

Scaling up

In the peer review process, we were asked how the Bradford approach could be replicated elsewhere. The answer is that any region within the UK can *choose* to commit to using data science to tackle the dreadful inequalities affecting our most disadvantaged communities. The question is whether there is the political will to tackle this source of inequality. West et al. (2022) have identified four key aspects needed to develop and sustain such approaches: leadership, resource and capacity, culture, and partnerships. Effective partnerships (between the police, local authorities, health systems, schools, universities etc.) are founded on strong, shared principles, which shape decisions and interactions through planning and delivery.

The connection of services through linked data requires an unprecedented breadth of collaboration across organisations, commitment to community engagement (see Islam et al., 2022) and an openness to change both culture and practice. We would urge every area across the UK to commit to the necessary partnership working and explore - at pace - how their routine datasets can be connected to improve outcomes for families. Our experience is that successful implementation needs to be driven at a regional level (allowing community engagement), with coordination and support provided through central government.

Conclusion

We have deliberately focused on examples of connecting education with health data to make the case for the use of education data in improving outcomes for children. In an alternative approach, we could have shown the benefits of linking education data with other datasets. We could also have shown the power of using routine educational data *per se* in learning how we can better support children. For example, we used the educational records from 8,130 participants in the BiB study to explore the predictive utility of the EYFSP. We found that the school readiness measure ('good level of development') predicted performance in reading, writing, maths and science at the end of Key Stage 1 (age 6-7) and later special educational needs (SEN) status (Atkinson et al., 2022). This means that the EYFSP could be used as a screening tool to identify children at risk of poor academic achievement and/or requiring SEN support. Thus, the EYFSP has the potential to be an effective trigger for early identification (and support) of SEN, and provides an exemplar for the possible use of improved provision through the use of education data.

We hope this essay shows why it is important to use education data effectively in efforts to tackle inequality and provide support to our most vulnerable families. The work in Bradford has demonstrated unequivocally the immense potential of data analytics to generate societal benefit. The connected data have also revealed the huge need for transformation across all of our organisations, systems and processes. Our firm conclusion is that connecting education data with other routine datasets is an essential first step towards reducing inequality and improving life outcomes for children and young people.

Acknowledgements

Born in Bradford (BiB) is only possible because of the enthusiasm and commitment of the children, parents and caregivers in BiB. We are grateful to all the participants, practitioners and researchers who make BiB happen. The work was conducted within infrastructure provided by the Centre for Applied Education Research (funded by the Department for Education through the Bradford Opportunity Area) and ActEarly, a City Collaboratory approach to early promotion of good health and wellbeing funded by the Medical Research

Council (grant reference MR/S037527). Mark Mon-Williams was supported by a fellowship from the Alan Turing Institute. The Centre for Applied Education Research receives core infrastructure funding from the Wellcome Trust (WT101597MA) and the National Institute for Health Research Yorkshire and Humber ARC (reference: NIHR20016). Linkages with education data are further supported by joint grants from the UK Medical Research Council and UK Economic and Social Science Research Council (MR/N024397/1) and the Strategic Priority Fund 'Tackling multimorbidity at scale' programme (MR/W014416/1) delivered by the Medical Research Council and the National Institute for Health and Care Research in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council.

All-Party Parliamentary Group on Autism (2019). *The Autism Act 10 years on: A report from the All-Party Parliamentary Group on Autism on understanding, services and support for autistic people and their families in England*

Atkinson, A. L., Hill, L. J. B., Pettinger, K. J., Wright, J., Hart, A. R., Dickerson, J., & Mon-Williams, M. (2022). Can holistic school readiness evaluations predict academic achievement and special educational needs status? Evidence from the Early Years Foundation Stage Profile. *Learning and Instruction, 77*, 101537

Child Safeguarding Practice Review Panel (2022). *Child protection in England: National review into the murders of Arthur Labinjo-Hughes and Star Hobson*

Department of Health and Social Care (2021). *National strategy for autistic children, young people and adults: 2021 to 2026*

DfE (Department for Education) (2017). *Bradford opportunity area delivery plan*

DfE (2021). Opportunity areas insight guide: Health and education

French, L., & Kennedy, E. M. M. (2018). Annual research review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: A systematic review. *Journal of Child Psychology and Psychiatry, 59*, 444-456

Islam, S., Albert, A., Haklay, M., & McEachan, R. (2022). *Co-production in ActEarly: nothing about us without us*. Bradford Institute for Health Research & University College London

Kelly, B., Williams, S., Collins, S., Mushtaq, F., Mon-Williams, M., Wright, B., Mason, D., & Wright, J. (2019). The association between socioeconomic status and autism diagnosis in the United Kingdom for children aged 5-8 years of age: Findings from the Born in Bradford cohort. *Autism, 23*, 131-140

Kolkman, D. (2022). The (in)credibility of algorithmic models to non-experts. *Information, Communication & Society, 25*(1), 93–109

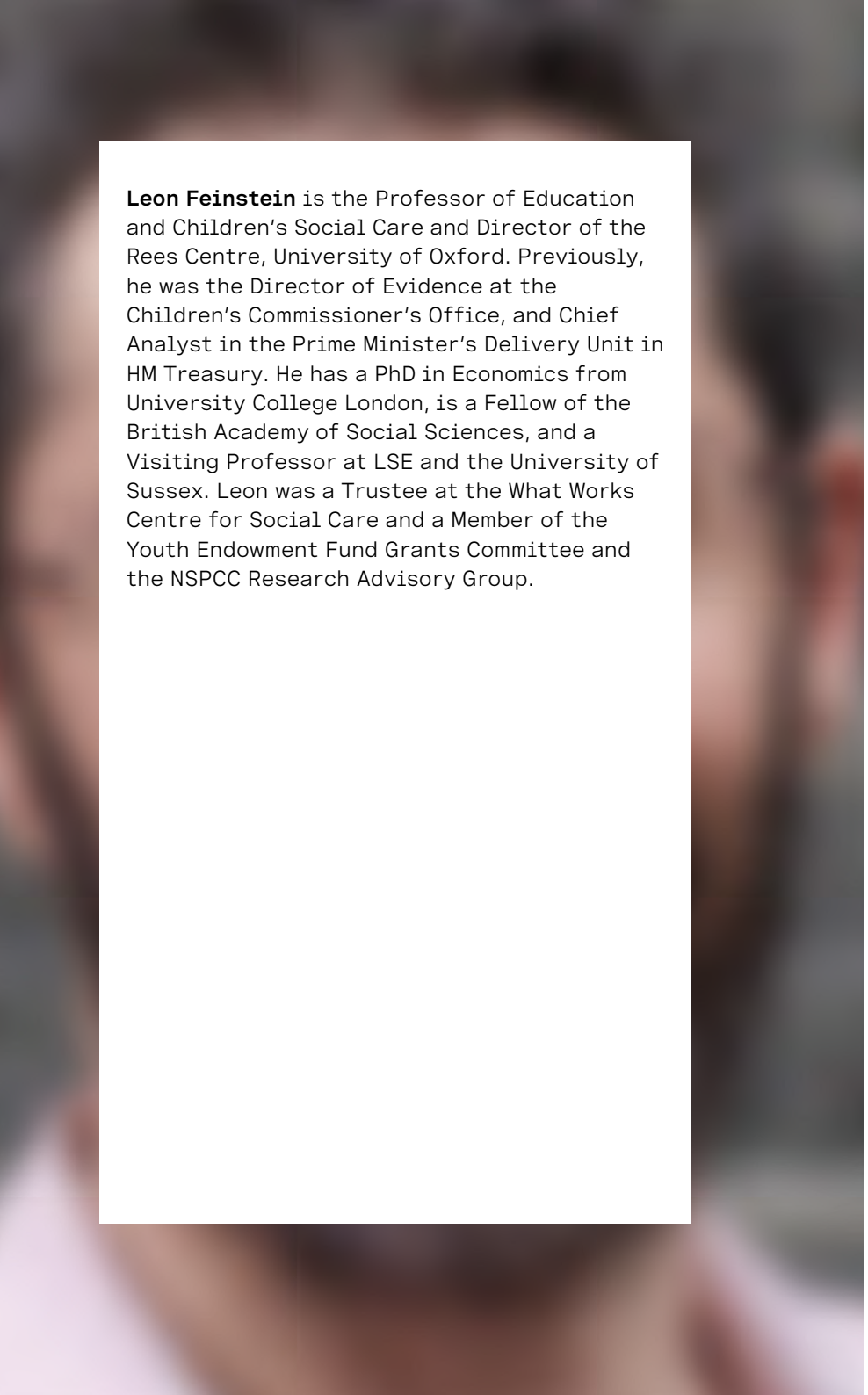
Pickett, K., Taylor-Robinson, D., et al. (2021). *The child of the north: Building a fairer future after COVID-19*. NHSA (Northern Health Science Alliance) and N8 Research Partnership

Sohal, K., Mason, D., Birkinshaw, J., West, J., McEachan, R. R. C., Elshehaly, M., Cooper, D., Shore, R., McCooe, M., Lawton, T., Mon-Williams, M., Sheldon, T., Bates, C., Wood, M., & Wright, J. (2022). Connected Bradford: A whole system data linkage accelerator. *Wellcome Open Research, 7*(26)

West, J., Wright, J., Bridges, S., Cartwright, C., Ciesla, K., Pickett, K. E., Shore, R., Witcherley, P., Flinders, M., McEachan, R. R. C., Mon-Williams, M., Bird, P., Lennon, L., Cooper, D., Muckle, S., England, K., & Sheldon, T. (2022). Developing a model for health determinants research within local government: Lessons from a large, urban local authority. *Wellcome Open Research, 6*, 276

Wright, B., Mon-Williams, M., Kelly, B., Williams, S., Sims, D., Mushtaq, F., Sohal, K., Blackwell, J. E., & Wright, J. (2019). Investigating the association between early years Foundation stage profile scores and subsequent diagnosis of an autism spectrum disorder: A retrospective study of linked healthcare and education data. *BMJ Paediatrics Open, 3*, e000483

Wright, B., Konstantopoulou, K., Sohal, K., Kelly, B., Morgan, G., Hulin, C., Mansoor, S., & Mon-Williams, M. (2021). Systematic approach to school-based assessments for autism spectrum disorders to reduce inequalities: A feasibility study in 10 primary schools. *BMJ Open, 11*, e041960



Leon Feinstein is the Professor of Education and Children's Social Care and Director of the Rees Centre, University of Oxford. Previously, he was the Director of Evidence at the Children's Commissioner's Office, and Chief Analyst in the Prime Minister's Delivery Unit in HM Treasury. He has a PhD in Economics from University College London, is a Fellow of the British Academy of Social Sciences, and a Visiting Professor at LSE and the University of Sussex. Leon was a Trustee at the What Works Centre for Social Care and a Member of the Youth Endowment Fund Grants Committee and the NSPCC Research Advisory Group.

An argument for better data about children

Leon Feinstein, University of Oxford†

This essay makes a case for attempts to link data about children across periods and multiple sources and to data about the adults they live with, so as to get a better handle on levels and characteristics of child vulnerability to harm and deprivation. It contends that there is both a rights-based ethical foundation for better aggregate data about children and families as well as a practical argument that it can be used to enable government and society to improve the lives, experiences and outcomes of children and young people, although these benefits do not come without risks, preconditions and costs. The use of children's statistical data should be balanced with concerns about data protection and the prevention of misuse of data, but there is also a positive agenda about linking data to the voices of children and their families, using data to address the research concerns and issues of children and families.

† This essay draws heavily on work undertaken when the author was Director of Evidence at the Children's Commissioner's Office, 2017-19. It also prefigures work at the Rees Centre in Oxford with partners in the University of Sussex and the London School of Economics and Political Science, and with Research in Practice supported by the Nuffield Strategic Fund, working with four local authorities in England to test and explore ways to link and use children's information that emphasise, recognise and celebrate ethics and the views and perspectives of children and families alongside statistical concerns and the uses of information to improve policy and practice for children and families.

The essay begins with an argument for the importance and value of aggregate data about children and their families and wider contexts as a means for influencing policy and practice in ways that can improve children's experiences and outcomes. Drawing on the example of immigration status as a risk factor causing vulnerability for some children, I first describe the weakness of current aggregate data for assessing the needs of children in the UK. I use this example because insecure immigration status is an important form of disadvantage and risk for some children, and influences their experience of education as well as other parts of their lives.

I then consider some of the real and perceived risks of improving data, and raise the issue of the importance of public trust and hence of transparency about data use and adequate mechanisms for ensuring meaningful public interest. The conclusion emphasises the vital role of government in balancing the benefits of aggregate data for holding government and society to account against the genuine risk of use of individual data in ways that would conflict with the rights of children.

A rights-based requirement for good aggregate data about children

Starting from a rights perspective, Article 3, para 2 of the United Nations Convention on the Rights of the Child (UNCRC) commits States Parties 'to ensure the child such protection and care as is necessary for his or her wellbeing, and, to this end… take all appropriate legislative and administrative measures.'

How are we to know if this is achieved and for whom? Wellbeing is not unidimensional, and the UNCRC recognises multiple aspects of protection and care, for example in relation to children who are at risk of 'physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment or exploitation, including sexual abuse' (Article 19), removed from their family environment (Article 20), adopted (Article 21), a refugee (Article 22), disabled (Article 23), economically exploited (Article 32), sexually exploited or abused (Article 34), abducted or trafficked (Article 35) or in detention (Article 37).

The vulnerability framework developed at the Children's Commissioner's Office 2016-20 (e.g., CCO, 2017) was an attempt to measure the number of children with different characteristics related to protection and care across the domains identified above, and to assess what is known about their views and experiences through forms of qualitative research, their characteristics and outcomes in terms of statistical measures and what they receive in terms of government support. This was a deliberate attempt to span the terrain of risk and harm and to provide a general overview of issues so that the Commissioner could report to Parliament on the state of the nation's most vulnerable, hidden and invisible children and the quality of the response of government and society to them. This enabled the Commissioner to target reviews on specific issues of concern from an evidence-based and informed perspective.

The Commissioner's power to request data and visit sites where children were resident was then used to further probe experiences for groups of children for whom the data indicated high levels of concern or for whom there was inadequate information (e.g., CCO, 2020a, 2020b). This method was used to identify particular areas of concern on which subsequent lobbying by the Commissioner pushed for improvements in their care and treatment, including young carers (CCO, 2016), children in detention in mental health settings and other settings (CCO, 2020a) and homeless children (CCO, 2020b). Advocacy could have occurred without good quantitative data on the numbers of children affected, but the Commissioner found good data essential for engaging public attention. As the former Home Secretary the Rt Hon Jacqui Smith put it in a blog about the 2019 annual CCO report: 'Politics is about priorities. The Children's Commissioner has shown our politicians what's happening to our most vulnerable children. Now they must choose to do the right thing' (CCO, 2019b).

An example of the need for improved aggregate data

How many children in the UK have insecure immigration status?

One subgroup of concern in the CCO analysis was children who were vulnerable to harm by virtue of immigration status. The 2019 report included in this regard the category of refugees

and a separate category of children and young people with 'unresolved immigration status', with a further four subcategories of children and young people who were: unaccompanied asylum-seeking; arriving under Dublin regulations; in families seeking asylum; or undocumented (CCO, 2019a).

Subsequent work (Feinstein et al., 2022) has revised these categories to assess risk of harm for three distinct categories of immigration status: children and young people without leave to remain, with limited leave, or with indefinite leave. These can be distinguished from children and young people with UK citizenship in terms of level of risk of harm associated with each status, on an underlying continuum of risk, with no leave to remain the most associated with risk.

We found that it is not possible to estimate from the official statistics the number of children in the UK without leave to remain, with limited leave or with indefinite leave. This is partly because no attempt is made by the Office for National Statistics (ONS) to track children through the system across years, and partly because so many children are missing from the official statistics as undocumented, 'invisible' to the system or below the radar (Chase, 2009; Kohli, 2006).

Much is known from case histories and qualitative research about how as a nation the UK treats children and young people with insecure immigration status (see, for example, Dexter et al., 2016; Price & Spencer, 2015). However, if we cannot even say how many children are currently going through the legal system, how they are in general and what their outcomes are, how sure can we be that we are meeting our obligations to them?

Children in need

These questions about accountability represent one kind of argument for the value of aggregate data. Arguments are also made about how to understand the effectiveness of policy. As Chancellor of the Duchy of Lancaster Michael Gove put it in the 2020 Ditchley Annual Lecture on 'The privilege of public service' (Gove, 2020):

> If Government ensures its departments and agencies share and publish data far more, then data analytics specialists can help us more rigorously to evaluate policy successes and delivery failures. People's privacy of course must be protected. But once suitably anonymised, it is imperative that we learn the hugely valuable lessons that lie buried in our data.

When we try to understand levels of need more generally, for example in relation to the number of children in England living in households or families with characteristics or locations that indicate higher potential likelihood of current and future harm, we also find considerable difficulties of measurement. Some of this results from inherent empirical challenges, but the bigger difficulty comes from the lack of activity to link data about children to data about adults, so as to know how many children are living in households characterised by high levels of drug or alcohol misuse, mental health difficulties, disabilities, material deprivation, prison, abuse or other characteristics likely to increase the risk of harm or disadvantage for children. There are some important beneficial examples of linkage but, as the CCO work has indicated, very substantial gaps.

The National Audit Office (NAO) made the argument for better data in a comment in 2019 on the Department for Education (DfE): 'The Department... still does not fully understand what is driving demand for children's social care or why there is such wide variation between local authorities in their children's social care activity and costs' (HC 1868, Session 2017-19, 23 January 2019).

When the government responds to the MacAlister review of the Children's Social Care System (MacAlister, 2022) we will know more about how this administration intends to address the continued increasing pressures on the care system. The options comprise combinations of raised thresholds, reduced rights, improved prevention and increased spend. Better data on the level of need will be required if we are to know in aggregate what the results are for children and families.

The point of the CCO framework is that it is general and holistic. It is for governments to decide on priorities, but we may still wish to know the outcomes and experiences of

children impacted by the actions or inactions of government. For governments that choose to act to meet need, better data enables focus, clarity of strategy and transparency of outcome.

Data risks

Invisibility

Returning to the issue of immigration status, we must probe further into what is meant by 'invisibility' and whom it serves. A 2016 Freedom of Information request† by Pippa King asked 'if the Police or Home Office have requested data, or whether data has been sent to them, from the National Pupil Database and or any other Department of Education held pupil level database.' The answer indicated that: 'Since April 2012, the Police have submitted 31 requests for information to the National Pupil Database. All were granted, however only 21 resulted in information being supplied.'

As a result of legal action in 2017 the DfE stopped requesting data on nationality. However, in a country with a hostile environment for people with insecure immigration status there are evident, perceived risks to the provision of information about immigration status for children and families. Concerns about these risks increase the degree to which these families are forced to hide from official agencies, increasing their vulnerability to trafficking and other risks of harm (House of Commons, 2020).

Therefore, this essay doesn't argue for an open and blanket approval for all forms of use of children's data, but seeks to place the use of data in a context of participation and democracy. How are concerned families to know that if data is linked to enable beneficial aggregate analysis it will not be used at individual level to target children and families for punitive measures, immigration controls or other forms of sanction?

As a practical matter it is not technically hard to link data in secure, encrypted environments de-identifying individuals by removing personal, identifying information and replacing this with non-identifiable data keys that enable statistical analysis. If risks of small numbers are appropriately handled, no individuals can be re-identified. In principle, the Five Safes (UK Data Service, 2020) approach developed at the ONS enables data to be made available in non-identifying ways for research. This is similar to approaches adopted around the world (e.g., Hanafin, 2020) that provide secure, trusted and legal bases for linking administrative to longitudinal data, generating the potential for informative research. Researchers including myself have often experienced frustration at how slow and difficult it is to get access to aggregate and de-identified data held by the DfE, even when necessary safeguards are in place and a clear case has been made that the analysis is likely to add clear insight and value in support of children's interests.

However, individuals in any de-identified or anonymous dataset can be re-identified if suitable re-identifying information is provided. Moreover, it is hard for individuals and families to know how their individual data are being used and processed in the first place, and that individual data are not being shared with criminal justice or other agencies without their knowledge or agreement.

Therefore, trust is critical. However, a 2020 report by the Information Commissioner's Office (ICO) provided 'a comprehensive review of data protection practices, governance and other key control measures supporting' the National Pupil Database and other databases held by the DfE, which leads on data about children within Whitehall. The review made 139 recommendations, and found:

> There is no formal proactive oversight of any function of information governance, including data protection, records management, risk management, data sharing and information security within the DfE which along with a lack of formal documentation means the DfE cannot demonstrate accountability to the GDPR.

Members of the public might recognise the benefits of ethical and careful use of children's data for improving and evaluating policy and practice, but how can we be confident in systems for linking data about children when, according to the ICO report (2020, p 5):

† Reference number 2016-0038372

> Information risks are not managed in an informed or consistent manner throughout the DfE or in line with the Risk Management Framework. Information assets are not assessed with sufficient frequency to ensure that the process is effective and resulting risks are not recorded with sufficient granularity or detail on the Information Risk Log to enable meaningful control and monitoring. Not all information risks are recorded and where they are, they do not always identify actual risks or control measures.

It is of concern that the ICO finds 'There is an over reliance on using public task as the lawful basis for sharing which is not always appropriate and supported by identified legislation' (2020, Executive Summary, p 6). Thus, it seems that data management in the key government department charged to be custodian of children's data in central government is neither adequately open and transparent to inspire trust, nor adequately resourced to enable timely, secure and effective access to data.

Ethics

Ethical practice requires a further level of reflection. Ethics, as Leslie et al. (2020, p. 19) put it in their review of the ethics of using machine learning techniques in children's social care, 'is both about justifying morally correct conduct and about motivating and setting a direction of travel for that conduct.'

If we are to take a rights-based approach to requesting that data about children be available for statistical analysis, we must also recognise Article 12: 'States Parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in all matters affecting the child.' This requires us both to seek the views of children about uses of data and also their views and perspectives on their circumstances, experiences and conditions as a form of information in data[†], when we use data to identify and address unmet needs, for example. This is not straightforward.

† I am grateful to Professor Elaine Sharland at the University of Sussex for this formulation. There is much more to be said about this.

There is an important tension between the Article 12 right to be heard and the Article 3 duty on states to ensure that 'in all actions concerning children... the best interests of the child shall be a primary consideration.'

Taylor (2016) describes the legal and operational difficulties of achieving the objectives of Article 3 when children are not at the centre of legal and policy systems. Although children may not be best placed to assess the value of aggregate data for policy decisions, this does not mean the conversation should not be had - just that Article 12 must be balanced with Article 3. As Lundy (2007) emphasises in a classic paper on the depth of meaning of Article 12, children, families and practitioners rarely have opportunities to shape the collection, interpretation and uses of data about their lives, and there is no recognised way to assess the quality or impact of this work (see also Bakketeig et al., 2020). Even where vulnerable children's voices are elicited, they are often not well heard or acted on (Kennan et al., 2018). Parents, caregivers and wider communities also have a role in setting priorities for data collection and in considering the meaning and implications from the findings of data.

In addition, to claim that there are material benefits to aggregating children's data and linking to information on parents and caregivers, we must establish both that improved data is necessary to realise the rights of children, and that appropriate safeguards are in place so that children's digital rights are not impinged. Recognising the benefits of collecting, collating and analysing children's data we must also adequately resource the work of securely, safely and transparently handling it, and of engaging children and young people, families and wider communities on the questions for analysis. On these issues there is clearly a long way to go.

Conclusion

As with children who are vulnerable by virtue of immigration status, the implicit response of government to unmet need may be to enhance invisibility and restrict rights rather than meet need. Aggregate data is a means for government and society to know something of how it is doing in meeting need. This essay argues that such data are necessary if we are to

deliver on the rights of children, but this must be balanced with a genuine and deep engagement with children, young people, families, practitioners and wider communities about what data are used and how, and more progress must be made on transparency and demonstrable safeguards.

Government can take advantage of the opportunities provided by better data about children and families, but also has a responsibility to ensure the data are used in the best interests of children, and that must involve deep, wide and meaningful dialogue that enhances trust by recognising the risks and biases of data as well as the benefits. We might recognise the limits of children and families to understand all aspects of government and data, but providing real opportunities to shape data use would not only improve legitimacy; it would also improve policymaking.

Bakketeig, E., Boddy, J., Gundersen, T., Østergaard, J., & Hanrahan, F. (2020). Deconstructing doing well: What can we learn from care experienced young people in England, Denmark and Norway? *Children and Youth Services Review, 118*

CCO (Children's Commissioner). (2016). *The support provided to young carers in England*

CCO. (2017). *Childhood vulnerability in England 2017*

CCO. (2019a). *Childhood vulnerability in England 2019*

CCO. (2019b). Former Home Secretary and member of Children's Commissioner advisory board Jacqui Smith responds to our vulnerability report

CCO. (2020a). *Who are they? Where are they? Children locked up*

CCO. (2020b). *No way out: Children stuck in B&Bs during lockdown*

Chase, E. (2009). Agency and silence: Young people seeking asylum alone in the UK. *British Journal of Social Work, 40*(7), 2050-2068

Dexter, Z., Capron, L., & Gregg, L. (2016). *Making life impossible: How the needs of destitute migrant children are going unmet*. The Children's Society

Feinstein, L., Aleghfeli, Y. K., Buckley, C., Gilhooly, R., & Kohli, R. K. S. (2022). Conceptualising and measuring levels of risk by immigration status for children in the UK. *Contemporary Social Science, 16*(5), 538-555

Gove, M. (2020). The privilege of public service [Transcript of the speech]. Ditchley Annual Lecture

Hanafin, S. (2020). *Scoping review of literature: Best international practices in linking administrative and longitudinal study data*. Dublin: Department of Children, Equality, Disability, Integration and Youth

House of Commons. (2020). *Windrush lessons learned review*. Independent review by Wendy Williams. HC 93

ICO (Information Commissioner's Office). (2020). *Data protection audit report*. Department for Education

Kennan, D., Brady, B., & Forkan, C. (2018). Supporting children's participation in decision making: A systematic literature review exploring the effectiveness of participatory processes. *British Journal of Social Work, 48*(7), 1985-2002

Kohli, R. (2006). The sound of silence: Listening to what unaccompanied asylum-seeking children say and do not say. *British Journal of Social Work, 36*(5), 707-721

Leslie, D., Holmes, L., Hitrova, C., & Ott, E. (2020). *Ethics review of machine learning in children's social care*. What Works for Children's Social Care

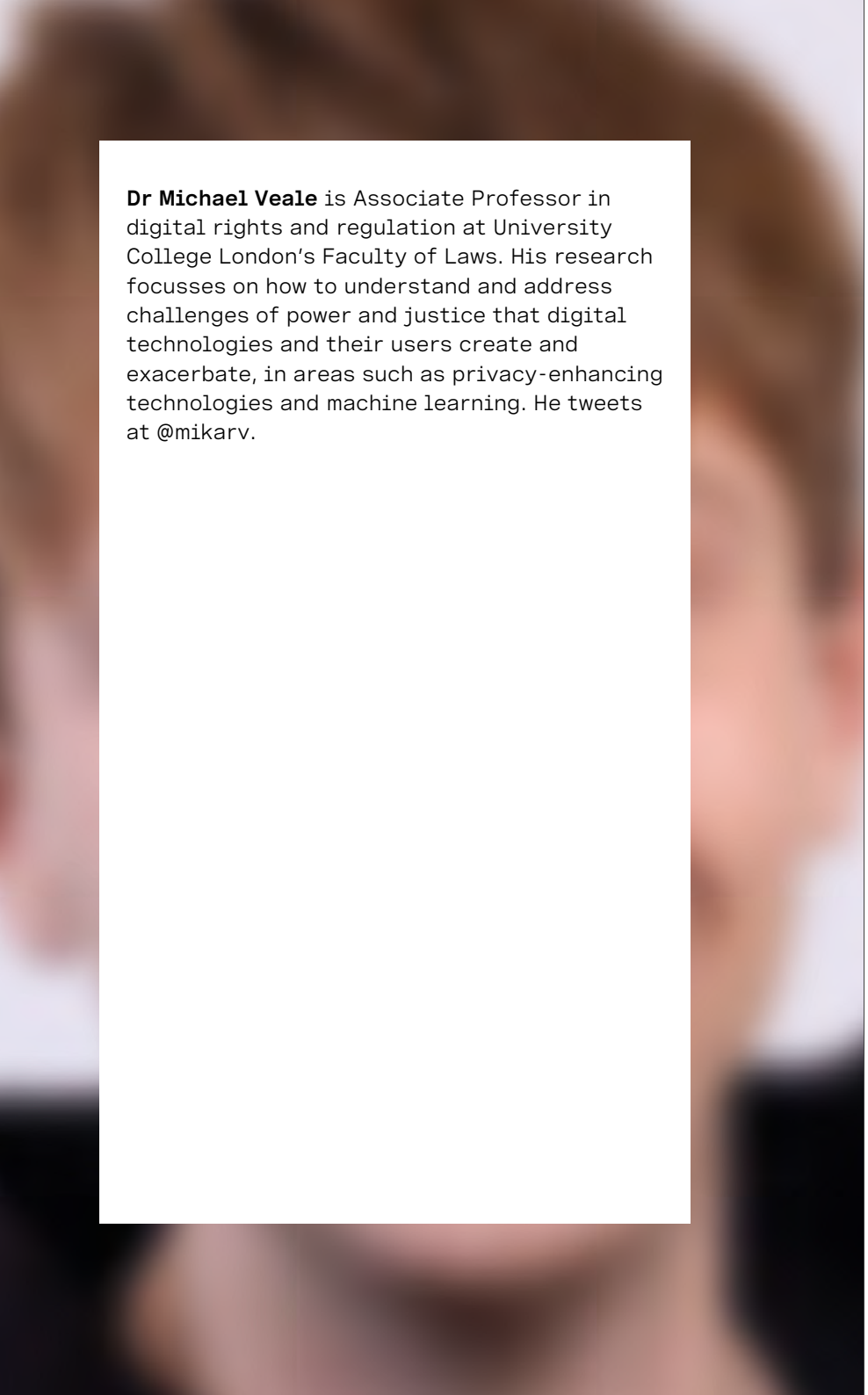
Lundy, L. (2007). 'Voice' is not enough: Conceptualising Article 12 of the United Nations Convention on the Rights of the Child. *British Educational Research Journal, 33*(6), 927-942

MacAlister, J. (2022). *Independent review of children's social care. Final report*

Price, J., & Spencer, S. (2015). *Safeguarding children from destitution: Local authority responses to families with 'no recourse to public funds'*. COMPAS

Taylor, R. (2016). Putting children first? Children's interests as a primary consideration in public law. *Child and Family Law Quarterly, 28*(1)

UK Data Service. (2020). What is the Five Safes framework?



Dr Michael Veale is Associate Professor in digital rights and regulation at University College London's Faculty of Laws. His research focusses on how to understand and address challenges of power and justice that digital technologies and their users create and exacerbate, in areas such as privacy-enhancing technologies and machine learning. He tweets at @mikarv.

Schools must resist big EdTech - but it won't be easy

Michael Veale, University College London

Over the last decade, school-level education across the world has seen the growing involvement of a small number of large technology firms prevalent across all sectors of the global economy. Three of the biggest in the West are Google, Apple and Microsoft,† all of which have a vertically integrated business model, meaning that they produce interlinked hardware, operating systems, a range of cloud services, and crucially for this essay, educational platforms.‡ Their educational platforms - including Google Classroom, Apple Classroom and Schoolwork and Microsoft Teams and OneNote for Education - are tied in varying constellations to their well-known general-purpose hardware (e.g., Chromebook, iPad, Surface) and operating systems (ChromeOS, iOS/MacOS, Windows). Such arrangements are often described as technology stacks, where the upper layers, such as the application-level functionality seen by users, relies on lower-level capabilities such as networking, cloud services or even

† This essay owns up to an implicit focus on Western education sectors, and does not consider large players elsewhere in the world, such as Alibaba, or the institutional conditions in which they operate.

‡ This contrasts with horizontal integration, where complementary services would be offered on the same layer of a technology stack rather than up and down it, such as offering many educational content services.

specific chips or sensors. Educational uptake of all types of layers this stack provided by 'big EdTech' giants saw a further global boost from 2020 due to the remote learning demands of the COVID-19 pandemic (Williamson, 2021).

These firms are far from the only digital services in education (see Decuypere et al., 2021, p. 3), but their vertical integration, deep infrastructural roots into devices, the foundational technologies underpinning computing and the internet, and influence outside the education sector sets them apart. Importantly the infrastructural nature of the power and influence they wield, particularly over medium-to-long-term horizons, appears harder for education actors to reason about than other important educational technology (EdTech) actors such as content providers that provide services more analogous to recognised educational activities and functions.

Plenty has been written on concerns over these and related developments, including a specific focus on issues such as student privacy and surveillance (Hope, 2016; Zeide, 2018) or pedagogical transformation (Perrotta et al., 2021; Zeide, 2020). However, in this short essay I wish to analyse and appraise what policymakers and educational institutions can *do* to respond to these giants' strategies.

What do platforms *want*?

The platform business model centres on connection, grouping and intermediation. By placing themselves in between many types of information and communication flows, platforms obtain economic and political power (Srnciek, 2017). A main way they do this is by creatively designing and deploying infrastructures that add stickiness into networks, and make interacting within the platformed part of the network easier than interacting across its boundaries (Cohen, 2019, p. 40). Collecting data and using predictive systems is often thought of as the core tool for platforms in establishing these boundaries (Srnciek, 2017), but many other effective strategies are used in concert, such as setting standards, controlling the development and functionality of hardware or software, binding users through technical standards or contracts, the effort of learning alternative systems, or providing often heavily subsidised bundles of complementary services. As the platform decides many of the rules of the game for the actors within the bounded part of the network, they can configure systems to extract economic rent or power for a range of purposes. As a result, despite the ambitions of internet pioneers of flatter, less hierarchical governance, for platforms, networks have proven to be a lucrative organising principle.

Impacts on platform participants

Participants interacting with platforms in the education sector play different parts in platform strategies. Here I focus on the impacts that interact directly with those strategies, rather than other (hugely) important issues such as pedagogical or social outcomes.

Students

Students are an obvious starting point. As they grow up, the platform-related choices made for them earlier on in life can stick with them due to non-material factors such as training and comfort, and material factors such as continued use of devices invested in by schools, parents or guardians. The strategy is familiar to old digital giants. Design software from Adobe, or operating system and office software from Microsoft, has been remarkably easy to pirate throughout its history, even when greater security techniques were possible. Rather than lost revenue, this has been characterised as a way to dominate emerging markets and consumers and create barriers to entry to defend against similarly featured, free and open source alternatives, such as GIMP or Linux (Karaganis, 2011, pp. 51-52). Bill Gates stated in 1998 that 'as long as they're going to steal [software] we want them to steal ours' (quoted in Grice & Junnarkar, 1998). In 2006, Hal Varian, now Google's chief economist, compared software usage to drugs: 'the first dose is free … once you start using a product, you keep using it' (quoted in Piller, 2006). The choices that schools make cannot easily be separated from the governance of platforms in society more broadly.

Teachers and other school staff

Schools, teachers and administrators can also find themselves tied into a single platform's ecosystem. This may be because

technologies are sold as a bundle; it is hard for a school to justify purchasing a separate email or office or administrative system if roughly suitable technologies are bundled in with classroom technologies, such as Google Workspace for Education Fundamentals or Office 365 Education. Notably, both of these tools have entirely globally free tiers, including cloud storage, videoconferencing and office software, which makes it hard to justify any software spend on competitors that do not have this broad, cross-subsidised, horizontal integration. Yet these free tiers may not meet all future needs; nor is there a firm guarantee they will be free forever given changing basic requirements. By the time they are integrated into technical systems and social routines, schools are likely to find it easier to upgrade and begin paying rather than to consider all potential options from scratch.

Wholesale reliance can be reinforced by the lack of funding or technical capacity in education. Remote technical support for both software and hardware can be part of platforms' offerings to underfunded sectors, taking this role away from schools and making local IT support staff difficult to justify. Remotely managed, packaged services also look appealing to schools faced with cybersecurity threats, particularly ransomware, the prevalence of which has increased in the UK education sector year-on-year (NCSC, 2021). Yet this trend means that where IT professionals do exist in schools, they are more likely to turn into 'licence managers' than have the organisational, practical and technical know-how previously expected of them (Balayn & Gürses, 2021, p. 110). Expertise bundled into platforms' cloud packages is hardly likely to diagnose the issues of lock-in, nor provide independent counsel for taking action such as diversifying or migrating away that would be inconvenient to the platforms.

In a similar manner, teachers can be reconfigured by platforms as part of strategies to increase their indispensability. Perrotta et al. (2021, p. 12) argue platforms transform teacher-student practices and relations from 'actual teaching' to facilitating and coordinating the 'slotting' in of automated tasks and modules. Insofar as these skills can be platform-specific - and, given the integration with a huge amount of other types of software for communication and content creation, they will tend to be - teachers can be reconfigured as agents perpetuating certain platforms' dominance.

Third party content providers

Education content vendors are coerced to design for ever-closer integration with these platforms. This predominantly occurs through these vendors using application programming interfaces (APIs) of the platforms, such as Google's Classroom API, which they need to use to connect with the systems schools are using. Integration with large platforms' APIs is typically thought of by developers as a risky business. They tether the firm to a platform that can expand and contract functionality and alter contractual terms with the aim of allowing certain operators to shut down and others to flourish (Bucher, 2013). Even if the large platforms themselves choose not to become content providers, this type of infrastructural integration extends platforms' power over other EdTech providers in national and international markets, and imposes structural decisions shaping the kind of pedagogy or interactivity that is facilitated, and the types that are not. In the future, it is not difficult to imagine that this may extend to directly facilitating interactions between pupils across schools, or even internationally - yet, on current trajectories, when and how an impactful shift such as this occurs will be on the platforms' terms, not that of schools' or potentially even local content providers'.

Can anything be done?

Some of the challenges of platforms in education are common to the general regulation of 'big tech', and can be seen through the lens of that literature (e.g., Moore & Tambini, 2021). This essay is not the space to unpack all of those general strategies and initiatives. Instead, here I focus on two possible approaches to providing countervailing forces to platformisation in education in particular.

Collective agreements against a credible threat of withdrawal

It does seem possible to force changes to educational platform practices in extreme situations, if collective measures are

taken, and backed by a credible threat of withdrawal.

In May 2021, the Dutch Data Protection Authority warned that schools could not use Google Workspace for Education in the new 2021 academic year, as a report by consultancy Privacy Company in 2020 indicated high privacy risks emerging in particular from the telemetry and diagnostic data that Google collected for its own purposes and analysed beyond the context of the contracting school. Against this backdrop of a potential prohibition, the Dutch cooperative of school boards for ICT purposes, SIVON (and the equivalent organisation for higher education, SURF) engaged in negotiations with Google. It obtained an agreement that Google would move from being a (joint) data controller, and that it would process personal data of students and staff for 33 of its own purposes, to a data processor, where it would only be able to process data for three narrow, pre-agreed-upon purposes on the explicit instruction of the school (Nas & Terra, 2021). Interestingly, this agreement appears to be only operational in the Netherlands through a contractual amendment, indicating the reluctance of Google to distribute the negotiated benefits elsewhere.

Such developments were only possible because they occurred against a backdrop of this platform operating illegally to the point that the regulator threatened a prohibition and gave a timeline for improvement. Platform behaviours damaging the long-term independence of the educational sector, rather than the immediate misuse of personal data, typically fall short of unambiguously breaking current law. No obvious regime exists to protect the education sector against powerful, informationalised business interests. Competition law, and even new ex ante competition-like instruments such as the EU's Digital Markets Act or the UK's proposed Digital Markets Unit still centre on consumer welfare rather than the specific interests of the educational sector. Even where data misuse can be pointed to, despite the UK having near-identical data protection law to the Netherlands, the lack so far of the Information Commissioner's Office threatening to prohibit Google Workspace for Education means negotiation in the UK would start from a much weaker position.

Furthermore, we can see that where the law is not broken, even countries like the Netherlands with strong collective agreements struggle to protect against data misuse by platforms. A parallel tale to the success of the Dutch Data Protection Authority and Privacy Company is the tale of interoperability of EdTech in the Netherlands. While there was significant proactive coordination to ensure national EdTech vendors signed up to governed interoperability requirements, in practice, the largest platforms, such as Google and Microsoft, have subverted such requirements by engaging with different education platforms as an identity provider, gaining a nodal position while not having to adhere to the interoperability requirements all other Dutch EdTech providers do (Kerssens & van Dijck, 2021).

From these two tales, it seems a credible threat of withdrawal is needed to fuel successful collective action. This highlights the urgency of seeing the educational sector through the lens of the new regimes to regulate platforms, such as the draft Digital Markets Act in the EU and the forthcoming Digital Markets Unit awaiting a statutory footing in the UK. Educational actors must discuss with regulators how to put the sector on legal notice, how to collectively agree a vision for the future, and then, together, consider how they can achieve change.

Layers of alternative generative and maintenance capacities

Where there is a demand for more advanced technology stacks in education - such as for videoconferencing to continue forms of education during lockdowns - platforms can appear the only technically feasible option. To stop this becoming inevitable, viable alternatives are required, and schools and educational decision-makers have key roles in bringing them into existence and keeping them there.

Yet even attempting to do so requires cooperation, collaboration and collective action between educational organisations, which is lacking in some jurisdictions. Individual schools cannot invent and maintain modern technology stacks alone. Platformised alternatives now develop, test and maintain modularised software at internet-scale (Gürses & van Hoboken, 2018). Alternatives do not need the billions of users Apple,

Google or Microsoft claim, but without some scale it is difficult to create comparable functionality, security or be responsive to new needs and developments.

Several countries have collective or membership organisations representing the education sector in ICT-related negotiations, such as SIVON in the Netherlands. The UK lacks a general-purpose overarching organisation for primary or secondary education, relying on a patchwork of capacities at multiple levels. Nationally, procurement frameworks and guidelines exist. Local authorities or coalitions of them may establish shared IT support initiatives. Scotland, for example, organises online services for its state schools through a national credential management system called Glow.

Yet, as it stands, these patchworks of support levels are becoming conduits for platforms such as Google Workspace and Office 365. Many are 'Google for Education Partners'. One of Glow's main contemporary functions is providing access to schools to both systems across Scotland. There is a need for these organisations to have a longer term strategy role in representing the sector's interests *against* platformisation, rather than acting as a conduit, further subsidising already questionably cross-subsidised services and ensuring their ubiquity as a foundation for all other EdTech. Yet the fragmented and privatised nature of procurement, technology assessment and management organisations for schools in the UK, and particularly outside of Scotland, inhibits meaningful possibilities of representation at a level that will be able to apply any pressure to platforms at all.

A wide variety of school types, chains of accountability and governance mechanisms further fragment and limit the possibility or impact of sector-wide cooperation in jurisdictions like England. A 'divide-and-conquer' approach suits platforms well as *they* can be the unifiers, and benefit from the distinction between the low-friction bounded platform zone they create in a network and the residual background stickiness and friction created by fragmentation.

In areas where larger-scale coordination exists, or were it to be further supported, such bodies might try creating or supporting alternatives to mainstream platforms, potentially through international collaboration. For example, open source

technologies such as BigBlueButton, a specialist, web-based videoconferencing platform for teaching, and Moodle, a widely used open source learning management system, both have community-supported business models, supported either directly by users or specialised contractors. A further genre of 'community source' projects takes a more structured approach between clubs of institutions working together on software development, which may vote on their progress or development. German universities provide software platforms for administrative tasks through a jointly owned cooperative company, HIS (Hochschul Informations System eG) (Kerres, 2020, p. 691), while in the US higher education domain, Sakai and Kuali are foundations developing open source learning management systems and financial and administrative software respectively (Jisc, 2013).†

Initiatives to study and promote open source development in education have existed in the UK, such as OSS Watch, and some work and funded projects in higher and further education by sectoral education charity Jisc, but there appears to be less momentum in this area than in the early 2010s.[1] Potentially due to a lack of expertise or scale, this approach has also been more common in universities than schools - although similar platformisation trends are empirically visible in higher education (Fiebig et al., 2021).

Retaining and building IT expertise, developing alternatives and resisting cross-subsidised educational software bundles like Google Workspace or Office 365 has a cost. If educational institutions choose to put their own constellations of systems and software together, someone has to be around to maintain it and its bespoke quirks. In contrast, while platforms like iOS or Chromebook internally arrange their software development in a highly modular manner, they do not offer users the same granularity of choice, instead bundling components together and constantly updating, changing and managing these homogeneous bundles at vastly more economic internet-scale (Gürses & van Hoboken, 2018). The difficulty of customising less vertically integrated software to local needs while

† Germany generally has much more adoption of open technologies in education; see, for example, an empirical study of the university sector by Fiebig et al. (2021), which considers, inter alia, BigBlueButton and Moodle.

providing economical, scalable (and thus likely remote) support, should not be underestimated. It is this area of configuration, customisation and support that perhaps needs the most focus and institutional experimentation to get right and not frustrate or create excessive labour for local users.

A final challenge is that coalitions investing in the development of alternative, open software can be easier to justify supporting when proprietary alternatives cost a licence fee that goes down a sinkhole, rather than when the alternative is 'free'. At its core, making the case for spending to limit the homogenisation of the education sector by big EdTech firms requires governments to take a mature approach to understanding the value of such investments. Can we put a price on the significant loss of control of a country's education infrastructure?

Conclusion

Schools are slowly becoming extremely reliant on a few large companies' entire technological stacks in order to operate. In turn, these stacks are reshaping what schooling is and could be, and exercising unaccountable control over students, teachers, administrators and content providers alike. It is not on the cards for the educational sector to become technologically independent or 'sovereign'. Technologies will reshape the sector, and not all of those decisions will ever be able to be made by individual schools. But that is not to say that the sector cannot summon countervailing forces that allow it to stay strong in the face of these developments. This essay has been a modest effort to stimulate further thought in that direction, focusing on collective negotiations and joint collaboration on alternative technologies and support systems. This will not be an easy task in many areas, and may require rethinking underlying institutional conditions to give schools a more coherent and collective voice and resource.

Going forwards, there are reasons to think that big EdTech will become bigger and more vertically integrated still, particularly in relation to the use of artificial intelligence (AI) in education. AI's true pedagogical use is still questionable, but its political economy is much clearer. A small number of companies, many, like Google and Microsoft with significant EdTech interests, spend tens of millions of a dollars at a time to train models to analyse or generate text or multimedia, which bring a range of daunting policy challenges (Bender et al., 2021; Cobbe & Singh, 2021). Insofar as deployment of AI in education is broadly yet-to-come, it is crucial that schools and related decision-makers grapple now with the political economy of the technology stacks they are enmeshed in, in order not to lose further control of key pedagogical choices in the years to come.

Balayn, A., & Gürses, S. (2021). *Beyond debiasing: Regulating AI and its inequalities*. European Digital Rights (EDRi)

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability and Transparency in Computing Systems (FAccT 2021)*, 610-623

Bucher, T. (2013). Objects of intense feeling: The case of the Twitter API. Computational Culture: *A Journal of Software Studies, 3*

Cobbe, J., & Singh, J. (2021). Artificial intelligence as a service: Legal responsibilities, liabilities, and policy challenges. *Computer Law & Security Review, 42*, 105573

Cohen, J. E. (2019). *Between truth and power: The legal constructions of informational capitalism*. Oxford University Press

Decuypere, M., Grimaldi, E., & Landri, P. (2021). Introduction: Critical studies of digital education platforms. *Critical Studies in Education, 62*(1), 1-16

Fiebig, T., Gürses, S., Gañán, C. H., Kotkamp, E., Kuipers, F., Lindorfer, M., Prisse, M., & Sari, T. (2021). Heads in the clouds: Measuring the implications of universities migrating to public clouds. *ArXiv:2104.09462* [Cs]

Grice, C., & Junnarkar, S. (1998). Gates, Buffett a bit bearish. CNET, 2 July

Gürses, S., & van Hoboken, J. (2018). Privacy after the agile turn. In E. Selinger, J. Polonetsky, & O. Tene (Eds.), *The Cambridge handbook of consumer privacy* (pp. 579-601). Cambridge University Press

Hope, A. (2016). Biopower and school surveillance technologies 2.0. *British Journal of Sociology of Education, 37*(7), 885-904

Jisc. (2013). Shared services. 1 July

Karaganis, J. (Ed.). (2011). *Media piracy in emerging economies*. Social Science Research Council

Kerres, M. (2020). Against all odds: Education in Germany coping with Covid-19. *Postdigital Science and Education, 2*(3), 690-694

Kerssens, N., & van Dijck, J. (2021). The platformization of primary education in the Netherlands. *Learning, Media and Technology, 46*(3), 250-263

Moore, M., & Tambini, D. (Eds.). (2021). *Regulating big tech: Policy responses to digital dominance*. Oxford University Press

Nas, S., & Terra, F. (2021). *Update DPIA report: Google Workspace for Education*

NCSC (National Cyber Security Centre). (2021) Further ransomware attacks on the UK education sector by cyber criminals. 4 June

Perrotta, C., Gulson, K. N., Williamson, B., & Witzenberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education, 62*(1), 97-113

Piller, C. (2006). How piracy opens doors for Windows. *Los Angeles Times*, 9 April

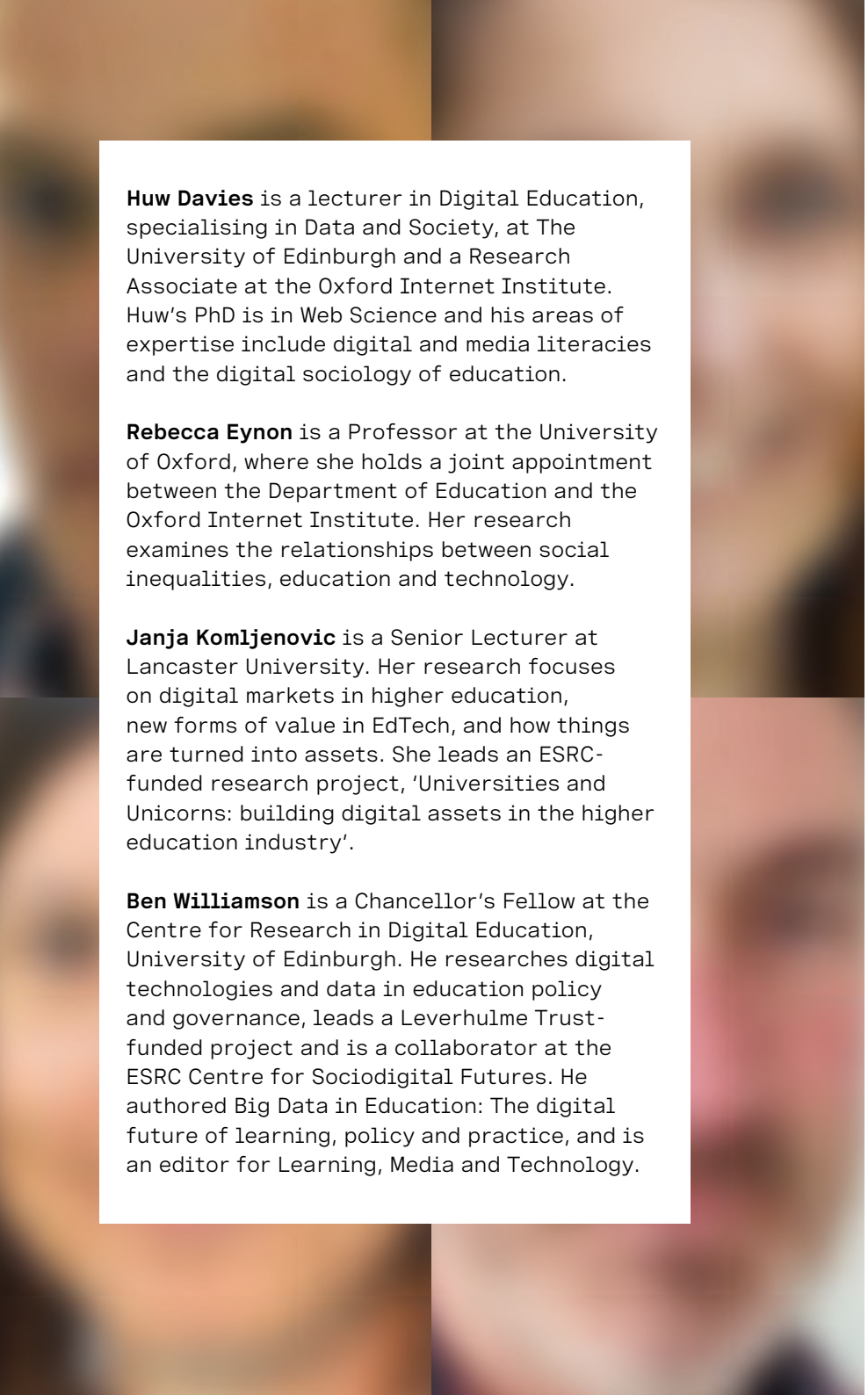
Srnciek, N. (2017). The challenges of platform capitalism: Understanding the logic of a new business model. *IPPR*, 20 September

Williamson, B. (2021). Education technology seizes a pandemic opening. *Current History, 120*(822), 15-20

Zeide, E. (2018). Education technology and student privacy. In E. Selinger, J. Polonetsky, & O. Tene (Eds.), *The Cambridge handbook of consumer privacy* (pp. 70-84). Cambridge University Press

Zeide, E. (2020). Robot teaching, pedagogy, and policy. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 787-803). Oxford University Press

1 http://oss-watch.ac.uk



Huw Davies is a lecturer in Digital Education, specialising in Data and Society, at The University of Edinburgh and a Research Associate at the Oxford Internet Institute. Huw's PhD is in Web Science and his areas of expertise include digital and media literacies and the digital sociology of education.

Rebecca Eynon is a Professor at the University of Oxford, where she holds a joint appointment between the Department of Education and the Oxford Internet Institute. Her research examines the relationships between social inequalities, education and technology.

Janja Komljenovic is a Senior Lecturer at Lancaster University. Her research focuses on digital markets in higher education, new forms of value in EdTech, and how things are turned into assets. She leads an ESRC-funded research project, 'Universities and Unicorns: building digital assets in the higher education industry'.

Ben Williamson is a Chancellor's Fellow at the Centre for Research in Digital Education, University of Edinburgh. He researches digital technologies and data in education policy and governance, leads a Leverhulme Trust-funded project and is a collaborator at the ESRC Centre for Sociodigital Futures. He authored Big Data in Education: The digital future of learning, policy and practice, and is an editor for Learning, Media and Technology.

Investigating the financial power brokers behind EdTech

Huw Davies, University of Edinburgh;
Rebecca Eynon, University of Oxford;
Janja Komljenovic, Lancaster University;
Ben Williamson, University of Edinburgh

The education technology (EdTech) market is diverse and powered by finance from a growing EdTech investment sector (Pau, 2021). It is a global market, but its wealth is concentrated, with the fastest growing and wealthiest companies located in just a few countries, primarily China, India and the USA, although European countries are starting to catch up (HolonIQ, 2022a). Similarly, EdTech investors are based primarily in China and the USA, and increasingly in Europe and the UK (Brighteye Ventures, 2022). According to the education market intelligence company HolonIQ, global venture capital investment in EdTech totalled US$500 million in 2010, rising to US$16 billion in 2020, and to more than US$20 billion in 2021 alone (HolonIQ, 2022b). This is expected to continue escalating, with the coronavirus pandemic fuelling increasing investment and interest in the EdTech market. The number of EdTech companies valued at more than a billion dollars ('unicorns') has surpassed 30 worldwide, most of them 'minted' in 2021, with venture capital investors seeking to identify other potential future unicorns that they can invest in now to secure economic returns in the future (Barosevcic, 2022).

Despite the significant funds unicorns spend on EdTech, they are relatively inconspicuous to the public or even to those

working in education. However, they deserve attention because they are power brokers in the EdTech economy, with their investment decisions ultimately determining what products and services are funded into existence or not; they are therefore consequential in shaping the future of education (Feher, 2018).

EdTech investors are not only financial actors; they are also political actors, using their financial power to influence what they think should be happening both within and after school. Investors take a lifelong learning view, 'from pre-K to gray',[1] based on a model of education that involves a person's whole life course and that is increasingly integrated with private technology services as well as the entertainment industry. EdTech investors are therefore setting the digital scene for the future of education in schools and beyond (Regan & Khwaja, 2019).

Investors calculate their future earnings, or so-called return on investment (ROI) (Muniesa et al., 2017). Increasingly, EdTech investors' vehicle of choice for the greatest ROI is the platform model. Platforms act as new intermediaries in education, providing access to learning content, allowing students to upload assignments or complete tasks online (Kerssens & van Dijck, 2021). A prominent example is Google Classroom, used extensively in schools, which offers both a free-to-access platform for online teaching and learning, as well as for-pay subscription features including integration with thousands of other services, data-driven insights about students, and increasing capacity to automate selected educational tasks (Perrotta et al., 2021).

Investors consider such digital platforms a particularly reliable revenue stream because they can be regularly upgraded with new features, integrated with other platform services and continuously collect user data, to be used as intelligence for future product and functionality development (Komljenovic, 2021). As in the wider digital economy, user data provides a source of value creation and subscription fees from users while simultaneously amassing data to develop further derivative products and services for future financial yields (Sadowski, 2020). Digital education platforms therefore act as new kinds of intermediaries, connecting users to educational services and extracting data traces from every interaction as a route to monetisation (Decuypere et al., 2021).

Beyond their investment strategies, some investors deliberately and actively promote a particular view of education to serve their interests. They operationalise specific ideas about what is wrong with education, how it should be disrupted or fixed, and what future education should materialise. A familiar trope is that EdTech will deliver a 'digital transformation' of education and modernise it towards the digital future (Marmol Queralto, 2021). This is widely shared across business, policy, technology and finance sectors, particularly since the onset of school disruptions related to COVID-19 and a global explosion in the use of private EdTech products across public education systems (Williamson & Hogan, 2020). EdTech investors have exploited this discourse, claiming they are investing in companies and platforms that will 'transform how the world learns' (HolonIQ, 2021).

The range of investors in EdTech includes venture capital, private equity and strategic investors, and may be generalist (investing in all or several economic sectors) or EdTech-specific (Komljenovic et al., 2021). We hypothesise that some investors might simply pursue ROI while others act as powerful and influential political actors. Here, we provide an insight into EdTech unicorns' investors through a focus on four investors that have invested in five or more unicorns – two EdTech-specific (Tiger Global Management, GSV Ventures) and two general (Owl Ventures and SoftBank Vision Fund).[†] We highlight the work investors do in reframing the practices, values and meaning of education, and consider the implications for the governance of education data.

† We took the list of unicorns from HolonIQ's website on 26 November 2021 and searched for investors in those unicorns on CrunchBase. We prepared a list of 424 investors. The huge majority (361 investors, or 85%) invested in only one unicorn. We found six investors that invested in five or more unicorns: Tiger Global Management, Coatue, Tencent, GSV Ventures, Owl Ventures and SoftBank Vision Fund. We took these six investors as the focus for our analysis. After an initial investigation, we removed Tencent due to limited sources and it being a tech company and not an investor per se, and Coatue as a general investor, to maintain a comparative balance between the cases. We propose that investors that invest in multiple unicorns are playing the most powerful role in shaping the future of EdTech.

EdTech-specific investors: GSV Ventures and Owl Ventures

GSV Ventures and Owl Ventures are both based in the USA, with international investments. They are major investors in EdTech unicorns, and also promote particular normative visions of the future of education – visions that their investments and promotional work are intended to enact.

GSV Ventures claims a need for massive digital disruption, and views education as an investment opportunity similar to e-commerce in the past. It claims, 'we invest in education technology leaders positioned to achieve disproportionate gains'.[2] GSV Ventures engages in 'pre-K to gray' thematic investing, seeking gains across the full lifecycle, including the schools sector. It invests in technologies used in schools in the UK, such as ClassDojo and Duolingo, although it invests in models of platform-based EdTech that it assumes have global market potential. The unicorns it has invested in include Andela, Degreed, Guild Education, Handshake and MasterClass. It uses the calendar analogy of 'BC' and 'AD' to position the pandemic as a transformative opportunity for investment and disruption in education. 'BC' refers to 'before corona', where education is seen as targeting the traditional student population, is organised in classes, temporally runs in semesters, consists of lectures and exams, is done during the daytime, is segmented in the curriculum and teaches theory. Alternatively, 'AD' stands for 'after disease', and is framed to cater for everyone across their life. For schools, AD education is characterised by peer-to-peer learning, on-demand teaching, massive scale operations (called 'weapons of mass instruction'), personalisation and increasing use of data and artificial intelligence (AI) ('RoboEd'). Learning is not only lifelong but also life-wide, with the rise of 'whole self-education' (Moe & Rajendran, 2020).

GSV Ventures does political work through reports, podcasts and other promotional materials, and by organising large events starring notable politicians and celebrities promoting their vision of education. It also facilitates start-up challenges to lead entrepreneurs in developing particular technologies and business models, and evaluates, rates and publicises the 'most transformational' EdTech start-ups in terms of their revenue scale and growth, user reach, geographic diversification and profit margins profile.[3]

Founded in 2014, Silicon Valley's Owl Ventures, with over US$2 billion of EdTech assets, is the world's most significant EdTech venture capitalist company. 'We believe there is a digital revolution rapidly unfolding in education', the company claims. 'This revolution is creating a historic opportunity to invest in companies that are disrupting and improving the over $6 trillion global education market... Hundreds of millions of students and teachers around the world can now leverage innovative learning platforms'.[4] Like GSV Ventures, Owl Ventures is celebrating new post-pandemic opportunities, highlighting how long top-down sales cycles and entrenched publishing incumbents made schools difficult markets. But now, thanks to government cash injections and the 'aggressive' adoption of technology during the pandemic, schools offer a lucrative source of income.

Owl Ventures' investments in EdTech unicorns Apna.co, BYJU's, Degreed, Greenlight, MasterClass, Newsela, Quizlet and Stash highlight its strong focus on employability, workplace skills and its interest in direct-to-consumer (DTC) EdTech. It also promotes a narrative of an outdated education system, and envisions the future of education in similar terms to GSV Ventures, such as short courses offering skills validated by micro-credentials and promoting wrap-around cradle-to-grave models of learning, arguing the need to constantly reskill to stay relevant in ever-evolving job markets. It also sees data as an important fix to inequality, encouraging companies to collect as much data as possible about their learners to facilitate social justice (Owl Ventures, 2021).

'Generalist' investors: Tiger Global Management and SoftBank Vision Fund

Tiger Global Management (USA) and SoftBank Vision Fund (UK) have international investments, primarily in technology

Tiger Global focuses on investments in China, India, Russia and the USA across multiple sectors.[5] It is seen as a major player in shaping the technology sector globally, especially internet-driven consumer and financial technology (FinTech) companies. Although education is a relatively minor segment in its

portfolio, it has invested in five EdTech unicorns: BYJU's, Kajabi, Outschool, Unacademy and Zuoyebang. One of its most recent EdTech investments is in the GoGuardian platform for schools, taking its valuation to more than US$1 billion, supporting its ambitions to scale to new markets outside the USA (Bergen, 2021). These highlight its interests in data-intensive personalised learning and creating a new platform marketplace beyond formal educational provision. It is not engaged in any kind of promotion of a particular vision of current or future education; it is the investments themselves that are important in shaping how data-intensive systems are to be used in education.

Similarly to Tiger Global, SoftBank Vision Fund's EdTech investments constitute a relatively minor part of its portfolio.[6] Indeed, education is a relatively new venture, with most EdTech investments made since 2020 (albeit it already invested in 2001 and 2004). As with Tiger Global Management, its influence is primarily in what it invests in rather than in the promotion of any particular vision of education. Its investments in 24 EdTech companies (including Atom Learning, Kahoot, Zuoyebang, EdCast, GoStudent, Riid and Unacdemy), of which seven are unicorns, tend to focus on personalisation and increasing educational access. Most investments are in DTC EdTech, thus its focus on reconfiguring and creating new educational markets.[7] Its corporate philosophy aims to 'promote the Information Revolution to contribute to the wellbeing of people and society' and to 'bring happiness and give inspiration to people.'[8] It invests in companies across the economy that promise to bring change with technology.

Personalisation, deinstitutionalisation and scale

EdTech-specific investors are political actors that construct, promote and operationalise particular ideas of education and its future. They institutionalise these ideas through investment in particular products and the promotion of particular visions of education through events, blogs, reports and other activities. Generalist investors do not engage in the promotion of EdTech, and appear mostly concerned with ROI regardless of the sector-specific interests. Indeed, they tend to invest in the later stages of the investment cycles, and with larger investment sums. Therefore, EdTech-specific investors engage in the first-order work of financing specific products and services, and also engage in the second-order work of shaping discourses and practices of education more widely. Generalists step in later to enable scaled growth and the institutionalisation of these visions. The meaning and values promoted by EdTech investors have significant implications for education. We now review three key themes and their implications for schools in the UK: personalised learning across the life course, bypassing educational institutions, and scaling up.

First, investors see EdTech as having a role in every learning stage of a person's lifecycle ('pre-K to gray'). EdTech investment presents itself as pursuing a moral purpose by financing the solutions required for the challenges facing learners of any age, especially in a digital-first economy. Across very different sectors of education, investors project similar ideas of education. In particular, platforms are promoted to allow personalised learning, which requires continuous data extraction, prediction of progress and tailored recommendations to improve future performance. Personalised, platform-based education is positioned as a moral imperative in the face of apparent curriculum standardisation, outdated modes of assessment and educators' incapacity to address individual learners' needs on demand and just in time. The implication is that schools are failing, unable to modernise, and don't personalise learning.

Personalisation is also envisaged to cater for individuals beyond institutionalised forms of learning. Platforms are imagined to help individuals find and engage with the most appropriate micro and other learning opportunities throughout their life, support parents, caregivers and children in tutoring or home schooling, provide digital credentials for career-readiness, connect learning to work more seamlessly, and even improve individuals' wellbeing and personal growth through tailored opportunities (Davies et al., 2021). Investors often claim such investments are 'fixing' education, supporting each individual based on their needs, and improving society. They offer what they construct as benevolent capitalism that can save education and alleviate geographical, social and economic inequalities through intelligent use of data and strategic investments. However, to address these inequalities, investors

say they need as much biographical and behavioural data as possible. This produces a moral tension between data for social good, and data for surveillance and profit.

Second, investors increasingly finance products and platforms that bypass the gatekeepers within the school system to deliver EdTech direct to young people, their families and lifelong learners. Again, such strategies reinforce views widely shared by EdTech investors, that formal education is slow to innovate, outdated, not fit for purpose and in need of transformation. This is reflected in rapidly increasing investment in platforms that mobilise and promote peer-to-peer learning, and the view that anyone can become a teacher, creating lectures and other content on digital platforms. Increasingly, education systems such as schooling in the UK and elsewhere are being decentralised and opened up to private contractors, making technical platforms into one of the new centralising powers, uniting schools as a national school system (Hillman et al,, 2020). This suggests EdTech investors are financing technology services and platforms that function as 'shadow' powers to shape what happens in schools, with little involvement of official governance or regulatory agencies (Williamson, 2019).

In a clear effort to build investor confidence in education, these investors construct this new deinstitutionalised sector as a largely untapped trillion-dollar source of revenue, which is less constrained by local legislative, regulatory and other messy challenges if focused directly on individuals. The data implications here are highly significant and are more likely to be based on the interests of the EdTech sector than children, young people and other learners.

However, while this appears to be a *de*institutionalisation of education - bypassing schools and universities, quality control and national, institutional and democratic standard setting - it is more of *re*institutionalisation of education around private platforms. This mirrors other domains of society such as news media. Publishers in the UK are regulated by public bodies such as the Independent Press Standards Organisations. Platforms are able to act as de facto publishers without similar independent oversight.

Neither personalisation nor reinstitutionalisation is possible without the third theme: scale. Only very large-scale platforms are able to produce massive disruption. Investors are therefore investing in technologies and markets that offer them rapid scale and revenue promise. For example, as a result of a series of inward capital investments, BYJU's in India is estimated to be worth more than US$22 billion. It has established a laboratory for AI innovation, hired technology experts from companies such as Amazon, acquired numerous other EdTech companies, and sought to expand to new markets internationally through major partnerships (such as Disney in the USA and Google in India). Its main source of revenue, which its investors expect to generate future ROI, is growing subscription payments from students for access to its platform services (Thathoo, 2022). It is therefore pursuing a simultaneous strategy of fundraising, technology innovation and expanded scale for its particular model of platform-based, AI-enhanced education, with a business plan that emphasises market and revenue growth. BYJU's example is quite representative of other unicorns' activities and strategies.

EdTech investors also strategically support coordinated and consolidated ecosystems of interoperable platforms that are both 'lifelong' and 'life-wide' in their pursuit of scale. An exemplar case is the learning management system Blackboard, acquired by Anthology, a student information systems company, in 2021, in a multibillion-dollar deal between their respective private equity owners. The newly merged company aims to 'break down data silos' and 'create the most comprehensive and modern EdTech ecosystem at a global scale for education' (Ballhaus, 2021). The merger was driven by revenue growth opportunities associated with cross-selling, increasing international reach and combining products and data to create new value (Hill, 2021).

Although the value of these new forms of data integration and cross-platform integration may not be obvious to educational institutions, it is an appealing model for vendors and investors because it promises to unlock value from data while locking in customers to the ecosystem. As with BYJU's, it also seems to indicate a trend towards increasing monopolisation, with the EdTech economy increasingly characterised by mega-EdTech corporations and vast

interoperable platform ecosystems.

Conclusion

We have highlighted the ways investors in EdTech unicorns operate in the education sector. Their primary influence is on the EdTech economy as they allocate finance to selected companies, but this exerts a second-order influence on the education sector more generally. Investors are not only creating the future uses of data-intensive systems in education through their financial power; they are also actively creating the future by shaping the discourse around EdTech in ways that will benefit their actions (Watters, 2016). Well-financed EdTech companies are enabled to expand their offerings, grow market share and deepen their penetration into education. Moreover, investors and EdTech companies are pursuing new models of education that can reshape conventional practices of teaching, learning and management, or produce new competitive alternatives, such as DTC learning platforms that bypass educational institutions altogether. By avoiding any official gatekeepers, including government and school bureaucracy, these solutions offer investors the rapid scaling that their large investments demand, with as yet unknown consequences for students, families, education and wider society.

The financial and discursive role investors play gives them a significant role in how the future of education should unfold and how data should be used. We are not convinced by the intertwined educational and data futures that investors are creating, as it is a reductionist view of the purposes of education, focusing on efficiency gains and learning as an individual activity that is primarily carried out for economic purposes. When it comes to data collected at or through school, who determines what is good for children, young people, schools and society? Whose interests are served, and who gets to write the agenda?

In the UK it is notable how much of the future vision of EdTech in schools has been shaped by the visions of the EdTech industry (see, for example, Ball, 2021; DfE, 2019). At the same time, there are significant regulatory and implementation gaps in the ways that data are used in schools (Day, 2021). It is important not only to address these gaps from a legal perspective, but also to ask if the kind of EdTech that gets developed and supported is the kind schools want and need. For example, what educational impact do particular digital products and services have, what kinds of algorithms run its operations and on what principles, and what is the pedagogic principle behind the innovation?

The primary motive of the investors driving the EdTech market is ROI. This influences, for example, data governance policy in education. Governments and teachers must have a far stronger role in setting both the educational and regulatory agenda so that education serves the interests of whole society: children, young people and life-learners, not just private capital investors.

Ball, S. J. (2021). *The education debate*. Policy Press

Ballhaus, B. (2021). The next big step in our journey. Blackboard Blog, 13 September

Barosevcic, M. (2022). The future unicorns of EdTech, intro: New article series from Emerge. LinkedIn, 20 January

Bergen, M. (2021). Tiger Global plows $200 million into EdTech firm GoGuardian. *Bloomberg*, 5 August

Brighteye Ventures. (2022). *The European EdTech funding report 2022*

Davies, H., Eynon, R., & Salvensen, C. (2021). The mobilisation of AI in education: A Bourdieusean field analysis. *Sociology, 55*(3), 539-560

Day, E. (2021). *Governance of data for children's learning in UK state schools. Digital Futures Commission*, 5Rights Foundation

Decuypere, M., Grimaldi, E., & Landri, P. (2021). Critical studies of digital education platforms. *Critical Studies in Education, 62*(1), 1-16

DfE (Department for Education). (2019). *Realising the potential of technology in education: A strategy for education providers and the technology industry*

Feher, M. (2018). *Rated agency: Investee politics in a speculative age*. Zone Books

Hill, P. (2021). The end of Blackboard as a standalone EdTech company. Phil on Ed Tech, 13 September

Hillman, T., Rensfeldt, A. B., & Ivarsson, J. (2020). Brave new platforms: A possible platform future for highly decentralised schooling. Learning, *Media and Technology, 45*(1), 7-16

HolonIQ. (2021). $16.1b of global EdTech venture capital in 2020. 5 January

HolonIQ. (2022a). Global EdTech 1000: 2021 stats and 2022 applications open. 24 February

HolonIQ. (2022b). Global EdTech venture capital report - Full year 2021. HolonIQ, 3 January

Kerssens, N., & van Dijck, J. (2021). The platformization of primary education in the Netherlands. *Learning, Media & Technology, 46*(3), 250-263

Komljenovic, J. (2021). The rise of education rentiers: Digital platforms, digital data and rents. *Learning, Media & Technology, 46*(3), 320-332

Komljenovic, J., Sellar, S., & Birch, K. (2021). *Mapping emerging EdTech trends in the higher education sector: Companies, investment deals and investors*. Lancaster University

Marmol Queralto, J. (2021). *A critical analysis of investors' logic in business discourse Universities and unicorns: Building digital assets in the higher education industry*. Lancaster University

Moe, M., & Rajendran, V. (2020). Dawn of the age of digital learning: An acceleration of trends that have been building for years. 6 May

Muniesa, F., Doganova, L., Ortiz, H., Pina-Stranger, Á., Paterson, F., Bourgoin, A., Ehrenstein, V., Juven, P. A., Pontille, D., Sarac-Lesavre, B., & Yon, G. (2017). *Capitalization: A cultural guide*. Presses des mines

Owl Ventures (2021) 2021 education outcomes report

Pau, S. (2021). Raising investments in EdTech: Trends in the market and tips for success. Nesta, 7 July

Perrotta, C., Gulson, K. N., Williamson, B., & Witzenberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education, 62*(1), 97-113

Regan, P. M., & Khwaja, E. T. (2019). Mapping the political economy of education technology: A networks perspective. *Policy Futures in Education, 17*(8), 1000-1023

Sadowski, J. (2020). The internet of landlords: Digital platforms and new mechanisms of rentier capitalism. *Antipode, 52*(2), 562-580

Thathoo, C. (2022). Byju backs BYJU's - EdTech unicorn founder invests $400 mn, ups stake to 25%. Inc42, 11 March

Watters, A. (2016). The best way to predict the future is to issue a press release. Hack Education, 2 November

Williamson, B. (2019). New power networks in educational technology. *Learning, Media & Technology, 44*(4), 395-398.

Williamson, B., & Hogan, A. (2021). *Commercialisation and privatisation in/of education in the context of Covid-19*. Education International.

1 https://gsv.ventures

2 https://gsv.ventures

3 https://www.asugsvsummit.com/gsv-edtech-150

4 https://owlvc.com/about.php

5 https://www.tigerglobal.com

6 https://visionfund.com/uk

7 https://edtech.dealroom.co

8 https://group.softbank/en/philosophy/corporate_philosophy

THE TROUBLE WITH DATA

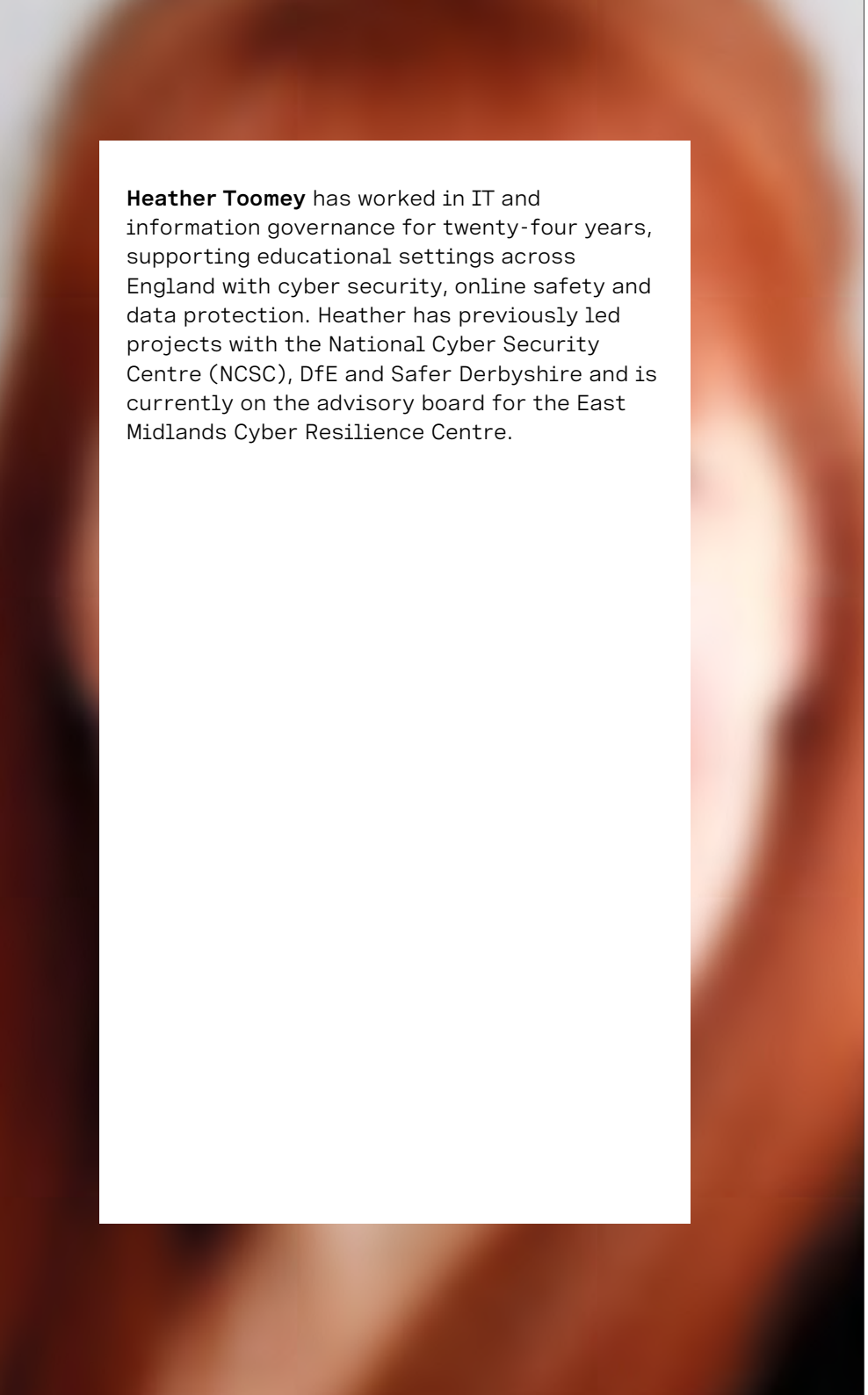
I didn't like [Microsoft Teams] because I liked being in class (Boy, 10)

You are rewarded Dojo points if you do excellent homework or complete tasks to a certain level. I don't think it helps learning because kids just want the points. However, they don't retain the information (Girl, 13)

The downside [of Google Classroom] is that children could make their own classroom, and you don't know what they could be posting. Because they can pretend to be a teacher. I've already done it (Boy, Year 5)

I get Dojo points for good work (Girl, 7)

Digital Futures Commission (2022) EdTech survey and Child Rights by Design Consultation (2022)



Heather Toomey has worked in IT and information governance for twenty-four years, supporting educational settings across England with cyber security, online safety and data protection. Heather has previously led projects with the National Cyber Security Centre (NCSC), DfE and Safer Derbyshire and is currently on the advisory board for the East Midlands Cyber Resilience Centre.

Turning data into insight and why data sharing is as vital as it can be concerning

Heather Toomey, Cyber Security and Information Governance specialist

Data has always been used in schools, with registers and class lists associated with school life.† The digital age has, however, seen data uploaded into more systems with fewer controls. This is the case across the UK, as data is required to monitor pupil attainment and evidence progress. Senior leaders use it to inform school improvement planning, and Ofsted (2021) uses the School Self-Evaluation Form (SEF) to help inform judgements on schools. There has also been an increased emphasis on attainment analysis using, for example, gender and deprivation indices, leading to more potential infringements of privacy.

While overflowing filing cabinets historically led to a natural need to purge data for practical reasons, the ever-expanding storage presented by large hard drives and cloud servers has led to data lakes,‡ or more often, unmanaged swamps, with the ability to store ever-increasing electronic and intangible personal data without an easy way to evaluate or control it.

---

† See Education Act 1996, Sections 434(1)(3)(4) & (6) and 458(4) & (5) and the Education (Pupil Registration) (England) (Amendment) Regulations 2016 (www.legislation.gov.uk/uksi/2016/792/contents/made).

‡ A data lake is a centralised system or repository of data that allows the storage of structured or unstructured data.

Staff can now copy and amend files in a way that was more difficult with hard copies, but if they lose track of version controls, there are also risks to the management of that data. Busy school staff may lack the opportunity to review the value of the data they are holding, and to ensure they only retain useful information.

After many years working in and with schools, my experience has been that of a natural hierarchy, with safeguarding data being generally well protected and the need for strong access controls recognised. Special educational needs (SEN) data, while ultimately shared with staff and key stakeholders to ensure accessibility needs are met, has tended to be matched by an understanding of the sensitivity surrounding it. However, data used for day-to-day administrative purposes may be shared too extensively, with teaching and non-teaching staff having increased MIS (Management Information System) access that can be poorly controlled and protected.†

Attainment data, at the core of teaching and learning, is created, collected and shared as the basis of progress monitoring, but during audits I have seen this on staff room walls and 'achievement charts', clearly visible and not seen as sensitive, despite young people's self-image being strongly associated with their view of their achievement. The premise is sound, but the visual representation of potential failure is stigmatising, and balancing the needs and rights of children against the need to share data to generate insights and protect their wellbeing is a constant struggle for school staff, who must decide what it is necessary to share in an ever-changing landscape.

Generating insights from data collection in schools

As reliance on data has grown, schools have purchased more systems and software solutions to collect, store, share and analyse data. Staff generally lack the expertise or time to make the most of them after purchase, and so data languishes in legacy systems as staff move on and school management focuses shift. Many schools lack a thorough understanding of which systems are currently in use, what data they hold, who has access and at what level, and how information is secured. This makes it impossible to create a comprehensive information asset register, and if you don't know you have it, you can't protect it.

Adding contextual information, such as prior attainment and free school meals eligibility, to seating plan software can enable the use of artificial intelligence (AI) to aid behaviour management (Lynch, 2019) depending on the system chosen. Recent research (Sailer et al., 2022) considered the use of AI in helping student-teachers to identify pupils with potential learning difficulties. Pupils can be tracked by their attainment, subgrouped by key indicators, such as gender or perceived disadvantage, and seated in class by algorithms that determine the statistical likelihood of one child disrupting another seated near them. AI comes with the risk of reaffirming a bias that has been hardcoded into algorithms by the design process or by biased training datasets, but the benefits are believed to be considerable in improving outcomes and supporting students in their learning journey (Zhang & Aslan, 2021).

Under the Education Act 1996,[1] it is a legal requirement for schools to provide national school data to the Department for Education (DfE). For state schools, this currently takes the form of the school census, carried out three times a year. In January 2022, the DfE asked schools to sign up to a daily attendance trial, as there is no doubt that the DfE needs to understand trends across the education sector and ultimately, improve outcomes and safeguard pupils. Following on from the successful EDSET (Educational Settings) daily collection form, which helped the government to understand the impact of the pandemic on both schools and the sector in general at regional and national level, the trial will collect real-time registration data from the school MIS. The data from registers will be used to help address absences more quickly and to better understand the long-term implications.

If the trial yields good results, this automated system could be used to collect other forms of data. The data will automatically be collected from school systems, processed and shared by EdTech company Wonde. However, while this

† See reports on breaches to the Information Commissioner's Office (ICO), National Cyber Security Centre (NCSC) and Jisc. See NCSC (2021a) for case study material.

approach will no doubt be more efficient and help reduce the administrative burden on schools, extensive checks will be needed to assess Wonde's suitability; although the company holds ISO 27001 certification, the international standard for information security, this is not the case with all EdTech vendors.

Meanwhile, collection of biometric data is increasing in schools, despite concern from privacy professionals and regulators (Green, 2021).† Cashless biometric catering (ParentPay, 2022) and biometric attendance systems are relatively common, particularly in the secondary sector and Trust schools, but the data protection implications of using these systems is neither well recognised nor understood. The DfE has guidance around biometric use (2012), and the Information Commissioner's Office (ICO) lists the use of biometrics as 'likely to result in high risk' to a data subject's rights and freedoms, requiring a data protection impact assessment (DPIA).[2]

The Protection of Freedoms Act 2012 requires schools and colleges to notify all parents, including birth parents and those with parental responsibility for a child, of their school's use of biometric data. This can be difficult if the data hasn't been provided to the school on entry. Further, school staff are often sold systems without referring back to the guidance or accurately assessing the risk. Questions about the use of biometric systems in schools have been discussed in the House of Lords,[3] and in October 2021, nine schools in North Ayrshire, Scotland, paused the rollout of facial recognition systems (FindBiometrics, 2021) following enquiries by the ICO.

The rights of the child vs schools' data practices

The UK signed the United Nations Convention on the Rights of the Child (UNCRC) in 1990. This sets out the rights that all children everywhere are entitled to, including the right to privacy, encompassed in Article 16, which states:

† Chapter 2 of the Protection of Freedoms Act 2012, 26(5) states that 'if, at any time, the child - (a) refuses to participate in, or continue to participate in, anything that involves the processing of the child's biometric information, or (b) otherwise objects to the processing of that information, the relevant authority must ensure that the information is not processed, irrespective of any consent given by a parent of the child under subsection (3).' See https://legislation.gov.uk/ukpga/2012/9/part/1/chapter/2/enacted

> (a) no child shall be subjected to arbitrary or unlawful interference with his or her privacy, family, home or correspondence, nor to unlawful attacks on his or her honour and reputation.
>
> (b) the child has the right to the protection of the law against such interference or attacks.

This is not clearly understood by schools. Parental consent is taken as overriding any objections from a child, and children are vulnerable to breaches of their privacy because of this imbalance of power. There have also been documented issues when separated parents have had differing opinions on consent, leading to difficulties for school staff in determining whether consent is confirmed or not.

The culture of data collection in schools is so heavily embedded that staff frequently collect raw scores and statistics that have little relevance or meaning in practice. In over 20 years of working in schools, pupil referral units and educational establishments, I have experienced staff inputting dozens of scores into spreadsheets and mark books that are never reviewed or subsequently evaluated. When pupils transfer to new settings, the receiving school will often call up and ask for baseline performance data on entry. At times, it was only these types of calls that would highlight missing data or data that had been entered incorrectly, demonstrating the lack of oversight and under-utilisation of the information gathered.

Raw scores on a test cannot determine whether or not a pupil has performed well; that requires context such as prior attainment, key indicators, pastoral needs and attendance. Turning data into real insight must be the priority, but the irony is that, in doing this, we need to collect and input more data to add this context. As datasets grow larger and more complicated, this necessitates the use of analytical tools and systems to inform and support the judgements that staff make. Consequently, schools turn to EdTech suppliers and third-party systems to process that data and support decision-making.

Problems with the use of EdTech in schools

Tools utilising AI are powerful, providing faster analysis and

insights into data, predicting potential outcomes and monitoring trends in behaviour against attainment. The ability of seating plan software to analyse where children sit *and* who they sit next to, and to predict which groups of pupils work better together, is intended to minimise the likelihood of specific pupils constantly interrupting the lesson and distracting those nearest to them. These disturbances to lessons are commonly referred to as 'persistent disruption' and have been evidenced to have a major impact on the attainment of the disrupter and the class as a whole (EEF, 2021). Ofsted first raised this issue as a problem in 2014, but behaviour management continues to be a real challenge for educators, with persistent disruption still the reason for over a third of permanent exclusions in 2019/20.[4] However, as automated decision-making creeps into pedagogy, privacy and pupil rights need to be considered. Gone are the days of graph paper and handwritten pupil names; today the most popular software vendors offer colourful pictograms and confirm their intention to share data with third parties in privacy notices that are often not fit for purpose and do not make it clear what data is collected or where it is shared.

Despite the type and level of data being added, processed and retained in these systems, schools tend to make procurement decisions based on school finances or choose a system based on popularity or by its use in other settings. During school audits I have been told numerous times that school staff have implemented a system due to the number of other schools who also use it. Relying on this 'safety in numbers' principle, rather than carrying out their own due diligence, it may lead to settings not even having a contract in place with suppliers, or having little understanding of system security and vendor data protection obligations.

Staff need an awareness of which systems hold personal information, for what purpose, and who has access. This requires schools to keep a full inventory of systems and applications and a complete information audit.[5] This also relies heavily on communication with suppliers and obtaining reliable information from them about their own internal processes. This often becomes time-consuming and arduous, with staff coming under pressure to make prompt decisions on provision without a complete understanding of how a system is transmitting, processing, storing and securing personal data.

Sometimes schools are unaware of the extent to which companies are utilising the data they upload or the levels of privileged access that third-party employees are provided with. Technical support teams and subcontractors might access pastoral issues and safeguarding concerns. While this may be referred to in the support contracts, school staff may be unaware that system administrators have such access (NCSC, 2020). Schools therefore need to ensure that appropriate due diligence and DPIAs consider privileged (administrator) access, and under what circumstances this access might be necessary.

Information held in electronic systems, like all other data stored electronically, may also be vulnerable to cyber-attack, and supply chain threats are emerging as a genuine concern.[6] As cyber-criminals target software developers and suppliers, if those suppliers have access, the criminal may gain access to third-party connected systems, in this case, schools. Many well-known software applications are commonly found in high numbers of schools, meaning that the implications of an attack on any one of them would be far-reaching. Suppliers to schools must have appropriate security to minimise the risks to schools.

Adversarial foreign governments are increasingly using hackers to target and disrupt organisations across the globe. These hackers, known as nation-state actors, are penetrating even the most secure systems. Schools are collateral damage in this worldwide cyber war, with many being affected by attacks meant for more significant targets. The drive for schools to transition from storing data in-house and from on-premises servers to the cloud is growing. The security of most cloud servers is certainly far more robust and reliable than the security seen routinely within school settings, but with schools using swathes of smaller applications, it is hard to reliably assess the risk of all of them.†

EdTech vendors must now meet the requirements of the Age Appropriate Design Code (AADC),[7] also known as the

† Ninety per cent of applications contain open source code, and open source applications are at equal risk (Sonatype, 2021), with the Apache Log4j vulnerability highlighted by the NCSC in December 2021 (NCSC, 2021b).

Children's Code, if their product or service is likely to be accessed by children. The code is currently not directly applicable to schools, although some EdTech vendors (Groopman, 2020) and privacy professionals have contested this limited scope. It does, however, have implications for school procurement of EdTech services, including those that are offered without charge. It is not yet clear how many schools understand their obligations in this regard and the need to have a contract in place, even when no money changes hands. The AADC and the proposed online safety bill aims to protect the rights of children at a time when privacy has come second to provision.[8]

Safeguarding - a growing EdTech subsector

As safeguarding systems are increasingly implemented in schools, more personal data is added to systems hosted by third parties, which are out of the direct control of the data controller. These record wellbeing concerns, referrals to outside agencies, hold copies of documents including photographs, and record qualitative opinions. This data may be exported to form safeguarding chronologies and provide information to the courts.

It is imperative that staff have a firm understanding of when it is necessary to share this type of data. Too often school staff struggle to determine the legal basis for processing personal data, under Article 6 of the General Data Protection Regulation (GDPR). Ensuring data sharing is lawful, proportionate and transparent is central to balancing data protection and privacy with the need to protect the vital interests† of data subjects and safeguard pupils.

In May 2021, Chief Constable Simon Bailey QPM, the National Police Chiefs' Council lead for child protection at the time, said the failure of schools to share information with the police was one of the most significant obstacles in tackling child sexual exploitation. This follows the publication of the Jay Report in 2014 and the subsequent Independent Inquiry into Child Sexual Abuse in Rotherham (House of Commons, 2018).

† These relate to processing personal data to safeguard and protect their life. See https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-code

Published serious case reviews (NSPCC Learning, 2022) demonstrate the need for interagency working and data sharing, and the heavy reliance on data collection and review to inform the extent of specific risk factors. The need for robust information sharing and oversight is often cited in the learning from such case reviews, but the lack of interoperability between systems used by various agencies and departments makes seamless sharing a challenge. Ultimately, children's futures, and possibly their lives, are at stake.

Systems and procedures for monitoring, as required under the *Prevent duty guidance* for England and Wales (Home Office, 2021a), are a key example of systems that suffer from 'scope creep' in schools. Section 26 of the Counter-Terrorism and Security Act 2015[9] includes a duty to have 'due regard to the need to prevent people from being drawn into terrorism', yet monitoring is frequently much more extensive than the recommended risk-based approach would require (Home Office, 2021b).

Frequently, however, internet and classroom monitoring solutions include remote screen watching, screen capture, communications monitoring and key logging. Services purchased by schools may also involve monitoring third parties, and analysing and categorising activity across an entire network, including Wi-Fi-attached devices. These services state compliance with the Prevent duty, Ofsted regulations and keeping children safe in education guidance (DfE, 2021b), but omit any reference to compliance with data protection laws.

Conclusion

EdTech is a huge business, with an estimated spend on school EdTech up by 72% since 2019 (BESA, 2021), and the estimated value of the UK EdTech market at almost £3.5 billion (Walters, 2021). Technology was a crucial enabler of remote provision during the COVID-19 pandemic, and this led schools to accelerate planned procurement for software solutions or invest in systems that had not been planned. In a bid to ensure accessibility and inclusion for all, these rushed implementations led to a lack of time for due diligence and staff training. The pandemic left teachers 'learning on the job', changing ways of working in days, when implementation of such systems would

usually take years. Mistakes were made and ICO reports show incident reported by the education and childcare sector were second only to the health sector (ICO, 2022).

At present, EdTech companies have access to a huge amount of children's data, with very little understanding by schools as to what is ultimately processed and why. The benefits of improving pupil outcomes, by gaining better understanding, demonstrating progress and increasing attainment, are obvious, and the need to safeguard pupils is, undeniably, vital. However, ensuring pupil rights and privacy is a challenge. The majority of headteachers (88%) and teachers (84%) indicated that technology had or would contribute to improved pupil attainment (DfE, 2021a), and it is this perceived benefit that leads schools to invest so heavily in EdTech. Safeguarding and data concerns were highlighted by 23% of school staff, surveyed as part of the DfE's EdTech Survey 2020-21, but this was considered a 'small barrier' to the increased uptake of technology (DfE, 2021a).

The data and information schools collect is vital for informing individual safeguarding requirements and strategies to address wellbeing across the country. Persistent absenteeism (DfE, 2022a) can have a detrimental impact on children long after they exceed school leaving age (Lolly & Bermingham, 2020). Chronic absenteeism correlates with unauthorised absence rates, with pupils missing education without an adequate reason, increasing year on year.[10] The Timpson review of school exclusion found that every extra percentage point of school sessions missed due to unauthorised absence was associated with an increase of one percentage point in the likelihood of permanent exclusion (DfE, 2019). The collection of this essential data needs to be matched with well utilised analysis and planned interventions to ensure young people are all provided with the opportunities they deserve, especially following the return to the classroom after the COVID-19 pandemic.

Data sharing with the DfE has enabled the construction of pseudonymised datasets that track education data with the employment, benefits and earnings data of adult members of the public. The Longitudinal Educational Outcomes (LEO) data (DfE, 2022b) aims to use de-identified, person-level data to analyse the effectiveness of education policy and provision. The dataset connects an individual's education data with their employment, benefits and earnings. While these aims appear to be in the public interest and children's best interests, the tracking of individuals' academic progression as the means to measure their 'success' and the effectiveness of education policy and provision should be proportionate to the government's objectives. Success can also be measured in many ways that are not directly linked to academic performance, and there are many reasons why an individual's earnings, and their employment choices, may not always directly correlate to their academic achievement.

The term 'EdTech' is the combination of education and technology, but this intersection between teaching and technology can be a misnomer. Teachers are generally public sector workers. This is a sector that includes social workers, healthcare professionals, law enforcement and the armed forces - people we trust. EdTech vendors are not public bodies; they are commercial companies, and the level of access they have to children's data is astonishing. Many of these companies will utilise, or attempt to utilise, this data, to meet with their own strategic objectives. As we live through this digital revolution, we must be sure to balance our reliance on technology with a determination to protect the children it serves.

BESA (British Educational Suppliers Association). (2021). ICT in UK maintained schools 2021. Insights, 3 September

DfE (Department for Education). (2012). *Protection of children's biometric information in schools*. Guidance

DfE. (2019). *Timpson review of school exclusion*. May

DfE. (2021a). *Education technology (EdTech) Survey 2020–21*. May

DfE. (2021b). *Keeping children safe in education*

DfE. (2022a). *Statistics: Pupil absence*

DfE. (2022b). *Apply to access the Longitudinal Education Outcomes (LEO) dataset*

EEF (Education Endowment Foundation). (2021). *Behaviour interventions*

FindBiometrics. (2021). North Ayrshire suspends controversial in-school face payments program. 26 October

Green, A. (2021). Biometrics in education supports the new normal. Future Identity Blog, 17 September

Groopman, J. (2020). The pros and cons of biometric authentication. *TechTarget*, August

Home Office. (2021a). *Prevent duty guidance*

Home Office. (2021b). *Revised Prevent duty guidance: For England and Wales*

House of Commons. (2018). *The Rotherham independent review: A review into information passed to the Home Office in connection with allegations of child sexual abuse in Rotherham (1998–2005)*

ICO (Information Commissioner's Office). (2022). Data security incident trends, Q4 2021/22

Jay, A. (2014). *Independent inquiry into child sexual exploitation in Rotherham 1997–2013*

Lolly, C., & Bermingham, R. (2020). COVID-19 and the disadvantage gap. UK Parliament Post, 1 September

Lynch, M. (2019). Using machine learning to modify student behaviour. *The Tech Edvocate*, 21 October

NCSC (National Cyber Security Centre). (2020). How to do secure system administration, 16 September

NCSC. (2021a). Cyber security training for school staff. 21 April

NCSC. (2021b). Alert: Apache Log4j vulnerabilities. News, 10 December

NSPCC Learning. (2022). *Recently published case reviews*

Ofsted. (2014). Below the radar: Low-level disruption in the country's classrooms. September

Ofsted. (2021). *Education inspection framework*. Guidance

ParentPay. (2022). Efficient cashless catering in 2022

Sonatype. (2021). *State of the software supply chain*

Walters, R. (2022). *EdTech: The hyper-accelerator: The disruptive potential of the infant tech sector*. Roger Walters Tech Series

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2022). *Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. Learning and Instruction*, 101620

Zhang, K., & Aslan, A. (2021). AI technologies for education: Recent research and future directions. *Computers & Education*, 2, 100025

---

1 https://legislation.gov.uk/ukpga/1996/56/contents

2 https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/examples-of-processing-likely-to-result-in-high-risk

3 https://hansard.parliament.uk/Lords/2021-11-04/debates/26FB2DF4-8D5A-456B-AFDA-73501D1CCBD3/BiometricRecognitionTechnologiesInSchools

4 https://explore-education-statistics.service.gov.uk/find-statistics/permanent-and-fixed-period-exclusions-in-england

5 https://ico.org.uk/for-organisations/accountability-framework/records-of-processing-and-lawful-basis

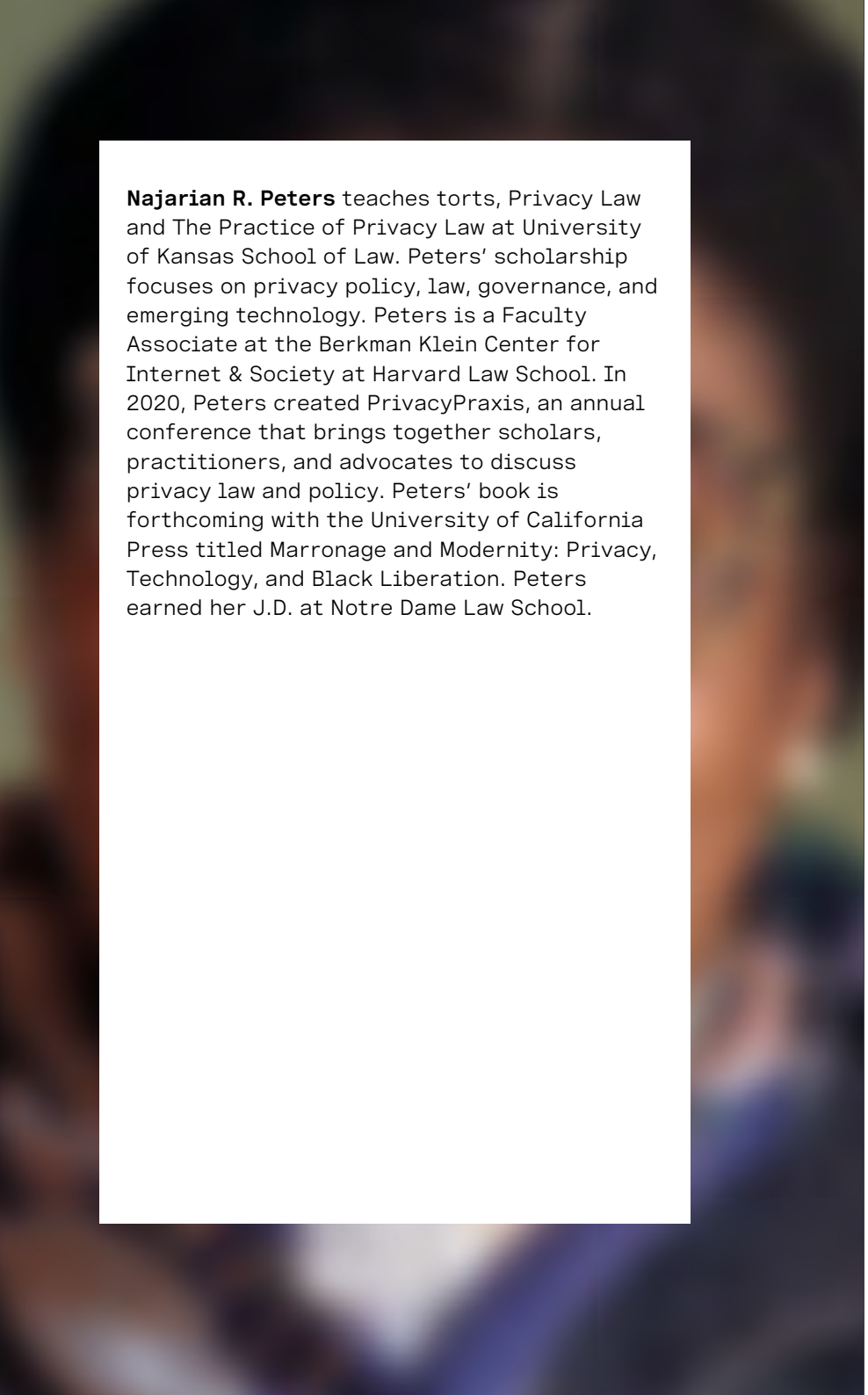
6 https://ncsc.gov.uk/collection/supply-chain-security/supply-chain-attack-examples

7 https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-code

8 https://ico.org.uk/for-organisations/childrens-code-hub/faqs-on-the-15-standards-of-the-children-s-code

9 https://legislation.gov.uk/ukpga/2015/6/section/26

10 https://explore-education-statistics.service.gov.uk/find-statistics/pupil-absence-in-schools-in-england-autumn-and-spring-terms/2020-21-autumn-and-spring-term



Najarian R. Peters teaches torts, Privacy Law and The Practice of Privacy Law at University of Kansas School of Law. Peters' scholarship focuses on privacy policy, law, governance, and emerging technology. Peters is a Faculty Associate at the Berkman Klein Center for Internet & Society at Harvard Law School. In 2020, Peters created PrivacyPraxis, an annual conference that brings together scholars, practitioners, and advocates to discuss privacy law and policy. Peters' book is forthcoming with the University of California Press titled Marronage and Modernity: Privacy, Technology, and Black Liberation. Peters earned her J.D. at Notre Dame Law School.

Black Data Traditions and the praxis of childhood preservation and anti-subordination in child education in the USA and UK

Najarian R. Peters, University of Kansas School of Law

The right to data accuracy is fundamental to data integrity, data ethics and the protection of privacy rights. However, the history of racial and other types of social marginalisation in both the USA and UK reflects how equity-based data practices have often evaded policy and law. Many of the formal processes of subordination that disproportionately impact racially marginalised people begin with harmful datafication in childhood through contact with formal educational systems.

Much has been written about how data policy and legal regimes that interact with racially marginalised individuals maintain the racial subordination of broader communities. However, little has been said about the traditions of Black communities to preserve educational opportunity and childhood in direct opposition to those systems. Black Data Traditions† provide rich examples of anti-subordination customs and practices for policy and lawmakers to follow – if they mean what they say when they claim they want to eliminate racial discrimination and inequity in educational environments.

† The term 'Black Data Traditions' or 'BDT' is used to describe how Black people have developed practices to protect, shield and otherwise combat the onslaught of distortionist proclivities of the dominant culture in Western societies that stigmatise Blackness.

We are not going to eliminate racial marginalisation in formal education with diversity and inclusion. In fact, as currently configured, research has shown that diversity and inclusion programmes exist across a variety of industries and sectors including education. Yet disparities and discrimination continue to exist, and in some instances are increasing. Some of the reasons for failure of these programmes include a lack of understanding and white resistance in implementation (Allen & Liou, 2019; Bonilla-Silva, 2006; Dobbin & Kalev, 2016; Emerick, 2021; Flores & Rosa, 2015; Herdman & McMillan-Capehart, 2009; Twine, 2018). Such initiatives, often half-hearted, poorly composed and positioned to reproduce marginality or, at best, maintain the status quo, have yet to prevent the kinds of patterned and practised harm that they purport to target – especially in the treatment of racialised or minoritised school children. Black parents (including caregivers) have increasingly relied on customs and informal practices to counteract and protect their children from distortionist incursions on childhood and educational opportunity.

This essay briefly highlights Black Data Traditions in the USA and UK because these countries provide rich examples of anti-subordination practices focused on shielding Black children from data collection practices that can reverberate throughout a child's lifetime. This is not meant to be a typological examination or comprehensive analysis of the various kinds of dirty data that exist in school data systems about Black children. The impact of those data distortions that I argue should be understood as dirty data are illustrated in the statistics that are well known and referenced in a multitude of studies, some of which I reference here.

This essay centres Black Data Traditions by comparing two of the common elements that I have identified thus far that exist in both Black America and Black Britain: the preservation of Black childhood and educational opportunity through home education, and the supplemental education and therefore alternative data creation practices as a way to circumvent subordinating policy and/or law.

Preservation of Black childhood and educational opportunity through home education

Opting out in the USA

The COVID-19 pandemic caused a massive retreat from in-person work, educational and social life across the globe. In the USA, K-12 schools† held classes virtually in increasing numbers. A year and a half later, many parents began to complain that their children needed to return to school (Brenan, 2021; Demsas, 2021). They were concerned that their children's education, psychological wellbeing and ability to socialise was suffering as a result of virtual learning. However, there was hardly a consensus among parents, as the disaggregated data showed that the vast majority who wanted their children to return to school in-person were white. Black, Asian, Latinx and other non-white parents were not as convinced that their children would be better off returning to school in-person (*The Economist*, 2021). Black parents in particular sought to preserve what they understood as an opportunity for their children to have a better educational experience away from the formal school settings in the privacy of their own homes. In 2020, Black families had the highest increase in home-schooling rates (Eggleston & Fields, 2021).

The early reports on the reasons Black families gave for continuing to home educate their children in 2021, documented in news reports, mirrored many of the reasons that white parents sought to have their children return to school. Black parents reported that their children were benefiting psychologically, were not being subjected to disproportionate discipline, and performed better in their virtual classes. They specifically expressed relief about keeping their children home where they could monitor how they were being treated and perceived in virtual classrooms. They also reported that their children's anxiety and feelings of safety around school experiences were improved, along with their performance. Black children reported feeling safer (Fernando, 2021; Today, 2021) and able to concentrate more while attending virtual school. In sum, Black children and Black parents reported that

† K-12 in the British education system spans from nursery school through to ages 17-18 (12th Grade).

they were better off outside of the formal educational systems that have historically and currently mistreat them (Anderson, 2021; Chao, 2021; Harris, 2020; Saavedra et al., 2021; Shapiro et al., 2021; St George, 2021).

The processes that render Black children consistently over-represented in metrics related to underachievement and exclusionary discipline and under-represented in metrics such as matriculation into gifted and talented programmes and graduation rates are connected to teacher perceptions calcified in data creation and usage. Historically, Black children have been viewed and treated as less innocent, older and more aggressive and violent by teachers and school officials in formal K-12 educational settings (Goff et al., 2014; Ingraham, 2015). Furthermore, the subsequent data created in schools about Black children based on how they are perceived is connected to, if not the cause of, their being over-represented in exclusionary discipline leading to the school to prison pipeline (Heitzeg, 2009; Shedd, 2015; Tyner, 2014) as well as being under-represented or excluded in gifted and talented programmes (Blake et al., 2011; Downey & Pribesh, 2004; Grissom & Redding, 2016; Grissom et al., 2015; Skiba & Williams, 2014; Morris, 2007; Payne, 2011).

Civil rights research released in 2016 showed that Black children are 3.6 times likely to be suspended in pre-school in comparison to white children and 3.8 times likely to be suspended in K-12 (US Department of Education Office for Civil Rights, 2016). Furthermore, in terms of access to education opportunities, Black children are less likely to be recommended for gifted and talented programmes (Nicholson-Crotty et al., 2016). However, disaggregated data found that Black children were three times more likely to be recommended for these programmes if their teachers were Black instead of white. White teachers were found to be 12% less likely to predict that the same Black student would finish high school and 30% less likely to predict the student would graduate from college (Gershenson et al., 2016).

Black parents who decide to home educate their children prevent their children from navigating biased teacher perceptions that are codified as data in the education record. Furthermore, these parents create the space for their children to access educational opportunity unencumbered by racialisation. The data created about Black children who are home educated reflects their experience as learners in childhood. Black children who are home educated are not documented based on racially biased perceptions that align with historical stigma and marginality. My forthcoming research will explore child perceptions of the learning opportunities they encounter in home education as a result of not having to navigate the barriers of racialisation that the research illustrates exists in formal school environments. These practices of preserving childhood, encouraging child curiosity in a supportive learning environment, free from racialisation, which also prevents the learning child from knowing themselves as targets of racialisation, are some of the essential elements of the Black Data Tradition.

Research indicates that a larger proportion of Black families chose to home educate their children compared to white families even as the pandemic seemed to recede with the disbursement of vaccinations and boosters. While the extra burden of home education undoubtedly created challenges for Black families, these were more than likely offset by not having to deal with the complaints, distortions and other incursions that Black families experienced when their children were enrolled in formal educational environments. Still, the notion that Black families must continue to create their own workarounds to ensure their children are not subjected to disparate treatment in school is a matter of inequity that is incongruent to the claimed values of public education.

The decision to opt out of formal educational settings to ensure the preservation of educational opportunity and childhood in 2022 is not new for African Americans. The benefit of removing children from harmful educational environments creates an opportunity for self-development in childhood, where intrusions of racial stigma are known to create psychological and emotional harm - which are both barriers to academic success. Additionally, data collection practices that align with racially motivated biases cause negative distortion and reputational harm, but those practices can either be mitigated or stopped, or are an impossibility in real time for children who are home educated or attending virtual school.

For example, parents who intervene on their children's behalf when being unfairly disciplined or reprimanded have the benefit of being present in the moment. This allows them to advocate on behalf of their children promptly by de-escalating a situation, and may also prevent further harm by preventing unnecessary documentation in the child's education record.

I have written previously about how African American parents have opposed this treatment, data creation and collection practices dating back to 1787 in Massachusetts (Peters, 2019, 2022). Now, in 2022, notwithstanding the promise of *Brown vs Board of Education*,[1] Black American parents find this tradition useful for the same reasons Black American parents espoused 235 years ago. While similar research has been difficult to find among Black British families, the rise of home education among Black families in the UK illustrates similar practices and motivations. Still, at the time of writing this essay, it was not clear that UK families engaged in home education at the same rates as American families.

Opting out in the UK

Cheryl Phoenix founded The Black Child Agenda (BCA) in 2011 after she experienced the 'systematic and psychological abuse her and other Black children faced within the UK education system'[2] – to focus on the disproportionate exclusionary discipline of Black children in the British school system and address the school to prison pipeline. In June 2014, Phoenix participated in the Black Homeschool Fair Teaching Our Own, founded by Leah Salmon.[3] Other community-based organisations have since set up platforms and partnerships to develop ways to help remove their children from the harms experienced in the British school system.† Many of the same concerns, including disciplinary practices and exposure to the police while in school, form the bases for concern of Black British children (Crozier, 2005; Dodd & Quinn, 2022; Matiluko, 2020). While there is a growing consciousness and movement

† The Black Curriculum provides a variety of educational materials for both home-educating and non-home-educating families in the UK (https://theblackcurriculum.com). The Black Child Agenda website provides resources for home-educating families in the UK (https://theblackchildagenda.org/home-education).

around home education in the UK, more research is available on the indirect approaches of supplemental practices of education, data creation and data collection as a response to the official policy of miscategorising West Indian and other Black children.

Miscategorisation of student capabilities is another example of calcified teacher perception. When teacher perceptions are inaccurate but treated as accurate this disadvantages targeted individual children. When biased perceptions are baked into educational policy as was/is the case in the UK where Black children are disproportionately categorised as subnormal, social subordination is likely to be correlated to, if not a result of, that policy. Inaccurate data is another form of dirty data that also includes incomplete and/or misleading data.

Supplemental education and data creation practices as a way to circumvent subordinating law and policy

During his recent presentation with Dave Neita on 11 November 2021, posted on YouTube,† Bernard Coard highlighted that a child's educational structure of opportunities are determined by both their parents and their teachers' expectations. Coard's sobering observation about the British educational system reflects scepticism that the system will in fact change for the betterment of Black British children. Instead, he stressed the importance of supplementary education in Black British communities for school age-children, drawing on his work from over 50 years ago, when Coard documented how the children of the Windrush Generation in the UK were disproportionately forced into educationally subnormal (ESN) schools. His book made five main arguments (1971, p.5):

1. "There are very large numbers of our West Indian children in schools for the Educationally Sub-Normal – which is what ESN means.

† Dave Neita is a lawyer, poet and public speaker known for his social justice and human rights work. Bernard Coard is an educator, former Deputy Prime Minister and Minister of Finance, Trade and Industry and Planning in Grenada. Bernard is also the author of How the West Indian child is made educationally subnormal in the British school system, originally published in 1971. A record of their presentation can be accessed at: https://youtube.com/watch?v=2l4Pw7mUqlc

2. "These children have been wrongly placed there.
3. "Once placed in these schools, the vast majority never get out and return to normal schools.
4. "They suffer academically and in their job prospects for life because of being put in these schools.
5. "The authorities are doing very little to stop this scandal."

Closely held beliefs of British political officials (such as Enoch Powell†) aligned with the development of the British school system's discriminatory educational policies described by Coard. Coard's work catalysed the rise in consciousness and movement within Black communities in the UK that resulted in the rise and development of the supplemental schools that exist today. His work excavates how data creation and collection were weaponised against West Indian children in the British school system. The policy of categorising West Indian school children as 'subnormal' was based on the presumption of inferiority of people of African descent. That presumption was built into the British government's decision to aim its policy of subnormality at West Indian school children specifically. The implementation of the policy and therefore the base assumption rendered the self-fulfilling data prophecy that shows West Indian children were, in fact, over-represented in ESN schools.

On closer examination, the results were made and not found - the ESN policy of categorising West Indian children was based on dirty data (Richardson et al., 2019),‡ or data that is misleading, inaccurate and/or incomplete. What are race-based assumptions of inferiority, enacted in policies and laws, if not examples of dirty data? The broad discretion to create, collect and otherwise process dirty data in the British and American school systems are deeply connected to the motivations of Black parents in both nations who continue to

† Enoch Powell was a Member of Parliament and a British politician from 1950-87, including a stint as Minister of Health from 1960-63. He is most widely known for delivering an anti-immigration speech on 20 April 1968 that came to be known as the 'Rivers of Blood' speech.

‡ Dirty data is defined as data that is 'skewed, or systemically biased'.

place their children in alternative educational environments in place of or to supplement formal education. To counter the impact of being exposed to the environment created within British and American schools, supplemental education centres Black history and culture by teaching students about Black people's contributions, innovations and achievements.

Facing the similar onslaught of racial discrimination while pursuing educational opportunity, Black Americans also developed supplemental education in their communities. They developed a host of educational alternatives that would allow children to learn without the intrusion of anti-Black racialisation. In the 1960s and 1970s Black communities continued the tradition of home education through supplemental education with the focus of enhancing 'institutional autonomy and an Afrocentric perspective' (Kifano, 1996, p. 209). These educational centres also provided resources to formal K-12 schools in an attempt to raise the level of educational attainment and understanding of American history where its Black citizens were concerned.

Conclusion

Recent attacks on teaching American history have resulted in responses that range from violent outbreaks in school board meetings (Greenberg, 2022), legislation to protect the teaching of Black history[4] and the firing of teachers (Li, 2020; Steinberg, 2022). As was the case in the rhetoric of Enoch Powell in the UK, the rampant disinformation and misinformation around teaching American history continues to shape policy and law to the detriment of its educational system. The combination of silencing or erasure and dirty data at work in the attack on teaching American history means that remedial movements continue to be required to counteract the degradation of the American education system.

Like the data distortions weaponised as per the foregoing discussion, the integrity of our educational system must be critically analysed to understand what continues to happen to create disparities along with the overreaction to policies and actors who are trying to eliminate those disparities. The first step towards correcting the hardwired dirty data practices is to recognise that dirty data exists. While new encroachments

threaten this approach, there are other policy and legal possibilities that we might consider.

The right to personal data accuracy in education records was recognised as key to the passage of the Family Education Rights Act (FERPA) in 1974 following the Watergate scandal. (FERPA is known as the federal student privacy law in the USA.) I recently argued that FERPA should be enhanced by adding a requirement of content validation during the data collection phase, along with a right to a reasonable inference and a disclosure of law enforcement access to the education record (Peters, 2022). Unlike in the UK, the principle of accuracy is not as clear in the text of FERPA, where Article 5(1)(d) of the UK General Data Protection Regulation (GDPR) states that personal data shall be: '(d) accurate and, where necessary, kept up to date, every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay'.

While the UK GDPR does not define the term 'accurate', the Data Protection Act 2018 states that 'inaccurate' means 'incorrect or misleading as to any matter of fact'. Further, in the discussion of accuracy of opinions impacting personal data, the guidance states that 'in order to be accurate... records must make clear that it is an opinion, and, where appropriate, whose opinion it is. If it becomes clear that an opinion was based on inaccurate data, you should also record this fact in order to ensure your records are not misleading'.[5] And the following steps should be taken to ensure accuracy: 'accurately record information provided; accurately record the source of the information; take reasonable steps in the circumstances to ensure the accuracy of the information; and carefully consider any challenges to the accuracy of the information'.[6]

The USA and UK have passed laws recognising data and data privacy rights that seemingly protect the right to accurate data or personal data integrity. However, the will to enforce the laws and policies is where we see the USA and UK once again sharing the same fate, for the same problematic reasons in regard to its historically racially marginalised communities. The fate of disproportionate discrimination and access to educational opportunity will continue to be of great concern as long as systems continue to allow dirty data to be introduced into the education record without recourse or accountability. Cognitive dissonance about the afterlife† of historical marginalisation creates patterns of anti-enforcement of laws that readily align to create subordination and subjugation. This is exactly why Black communities in both the USA and UK developed and continue to develop their own data traditions to preserve the very rights that are often deprioritised and circumscribed elsewhere.

† Saidiya Hartman and Christina Sharpe, respectively in *Scenes of subjection: Terror, slavery, and self-making in nineteenth century America and In the wake: On Blackness and being*, have both written on the connections of subordination and precarity that are based in the history of enslavement of people of African descent in Western societies and are at the foundation of persistent inequity in access to educational opportunity in the USA.

Allen, R. L., & Liou, D. D. (2019). Managing whiteness: The call for educational leadership to breach the contractual expectations of white supremacy. *Urban Education, 54*(5), 677-705

Anderson, M. D. (2021). Why Black parents aren't joining the push to reopen schools. *Mother Jones*, 18 March

Blake, J. J., Butler, B. R., Lewis, C. W., & Darensbourg, A. (2011). Unmasking the inequitable discipline experiences of urban Black girls: Implications for urban educational stakeholders. *The Urban Review, 43*(1), 90-106

Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States* (2nd edn). Rowman & Littlefield

Brenan, M. (2021). Amid pandemic, 79% of K-12 parents support in-person school. *Gallup News*, 11 March

Chao, E. (2021). As schools reopen, will Black and Asian families return?, *PBS News Hour*, 5 July

Coard, B. (1971). *How the West Indian child is made educationally sub-normal in the British school system*. New Beacon for the Caribbean Education and Community Workers' Association

Crozier, G. (2005). 'There's a war against our children': Black educational underachievement. *British Journal of Sociology of Education, 26*(5), 585-598

Demsas, J. (2021). Poll: 79 percent of parents want a return to in-person schooling. *Vox*, 16 March

Dobbin, F., & Kalev, A. (2016). Why diversity programs fail. *Harvard Business Review*, July-August

Dodd, V., & Quinn, B. (2022). Black girl strip-searched by Met officers at London school tells of trauma. *The Guardian*, 16 March

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education, 77*(4), 267-282

Economist, The (2021). More non-white than white parents prefer remote learning for their children. 15 March

Eggleston, C., & Fields, J. (2021). Census Bureau's Household Pulse Survey shows significant increase in homeschooling rates in fall 2020. United States' Census Bureau, 22 March

Emerick, M. R. (2021). Diversity ideology and school leadership: Obscuring inequities for emergent bilingual students in career and technical education. *Educational Administration Quarterly*

Fernando, C. (2021). Some Black parents say remote learning during pandemic has kept students safe from racism in classroom. Associated Press, 4 May

Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review, 85*(2), 149-171

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychiatry, 106*(4), 526-545

Greenberg, L. (2022). Public, NAACP urge Volusia schools to continue Black history education and critical race theory ban. *FOX 35 Orlando*, 27 January

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student-teacher demographic match on teacher expectations. *Economics of Education Review, 52*, 209-224

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The 'representative bureaucracy' in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher, 44*(3)

Harris, B. (2020). Why Black families are choosing to keep their kids remote when schools reopen. *The Hechinger Report*, 7 August

Heitzeg, N. A. (2009). Education or incarceration: Zero tolerance policies and the school to prison pipeline. *Forum on Public Policy Online*

Herdman, A. O., & McMillan-Capehart, A. (2009). Establishing a diversity program is not enough: Exploring the determinants of diversity climate. *Journal of Business and Psychology, 25*(1), 39-53

Ingraham, C. (2015). Why white people see Black boys like Tamir Rice as older, bigger and guiltier than they really are. *The Washington Post*, 28 December

Kifano, S. (1996). Afrocentric education in supplementary schools: Paradigm and practice at the Mary McLeod Bethune Institute. *The Journal of Negro Education, 65*(2), 209-218

Li, D. K. (2020). Student-teacher in Tennessee dismissed over Black History Month assignment on slavery. *US News*, 6 February

Matiluko, S. (2020). Black parents, private school will not save your child from racism. I should know, I suffered at one. *Independent*, 16 January

Morris, E. W. (2007). 'Ladies' or 'loudies'? Perceptions and experiences of Black girls in classrooms. *Youth & Society, 38*(4)

Payne, A. (2011). *Equitable access for underrepresented students in gifted education*. The George Washington University Center for Equity & Excellence in Education, 8-9

Nicholson-Crotty, S., Grissom, J. A., Nicholson-Crotty, J., & Redding, C. (2016). Disentangling the causal mechanisms of representative bureaucracy: Evidence from assignment of students to gifted programs. *Journal of Public Administration Research and Theory, 26*(4), 745-757

Peters, N. R. (2019). The right to be and become: Black home-educators as child privacy protectors. *Michigan Journal of Race & Law*, 21

Peters, N. R. (2022). The Golem in the machine: FERPA, dirty data, and digital distortion in the education record. *Washington and Lee Law Review, 78*(5)

Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online*, 192

Saavedra, A., Rapaport, A., & Silver, D. (2021). Why some parents are sticking with remote learning - even as schools reopen. Brookings, 8 June

Shapiro, E., Green, E. L., & Kim, J. (2021). Missing in school reopening plans: Black families' trust. *The New York Times*, 1 February

Shedd, C. (2015). *Unequal city: Race, schools, and perceptions of injustice*. Russell Sage Foundation

Skiba, R. J., & Williams, N. T. (2014). *The Equity Project at Indiana University, are Black kids worse? Myths and facts about racial differences in behavior*. Supplementary Paper 1, March

St George, D. (2021). Some families of color remain wary of returning to classrooms as new school year begins. *The Washington Post*, 1 September

Steinberg, A. (2022). Lawmakers want to ban discomfort in school. But Black history isn't always comfortable. *NPR News*, 24 February

Today. (2021). 'We feel safer': Black parents say remote learning gives kids reprieve from racism. 6 May

Twine, F. W. (2018). Technology's invisible women: Black geek girls in Silicon Valley and the failure of diversity initiatives. *International Journal of Critical Diversity Studies*, 1 June.

Tyner, A. R. (2014). The emergence of the school-to-prison pipeline. American Bar Association, 1 June

US Department for Education's Office for Civil Rights. (2016). *2013–2014 civil rights data collection: A first look*

1 *Brown vs Board of Education 347* US 483 (1954).

2 https://theblackchildagenda.org/about-the-black-child-agenda

3 https://teachingourownuk.wordpress.com

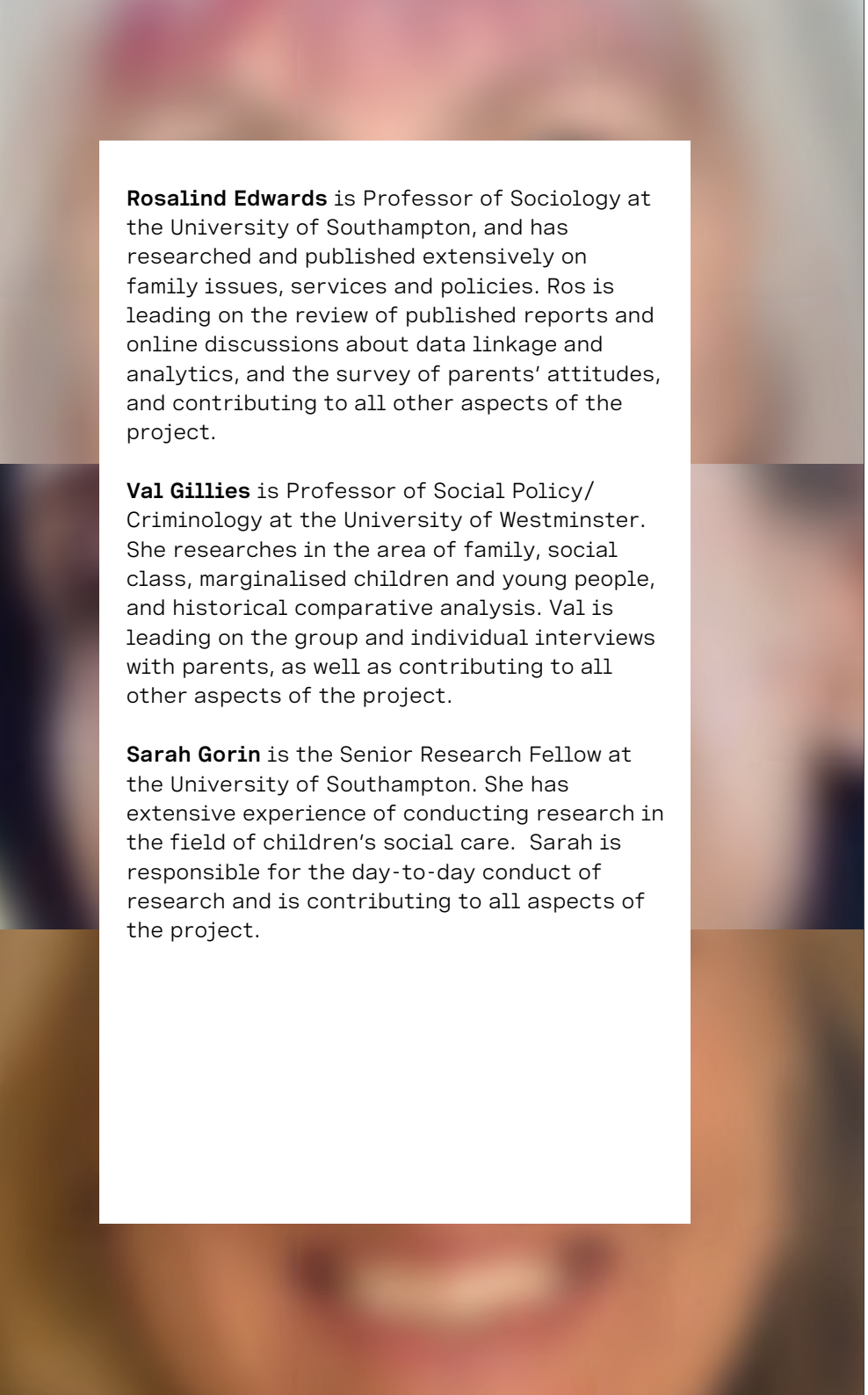
4 https://congress.gov/bill/117th-congress/house-bill/1394/text?format=txt&r=69&s=1;

https://booker.senate.gov/imo/media/doc/one-pager_for_african_american_history_act.pdf;

https://booker.senate.gov/imo/media/doc/section_by_section_summary_for_african_american_history_act.pdf

5 https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/accuracy

6 https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/accuracy



Rosalind Edwards is Professor of Sociology at the University of Southampton, and has researched and published extensively on family issues, services and policies. Ros is leading on the review of published reports and online discussions about data linkage and analytics, and the survey of parents' attitudes, and contributing to all other aspects of the project.

Val Gillies is Professor of Social Policy/ Criminology at the University of Westminster. She researches in the area of family, social class, marginalised children and young people, and historical comparative analysis. Val is leading on the group and individual interviews with parents, as well as contributing to all other aspects of the project.

Sarah Gorin is the Senior Research Fellow at the University of Southampton. She has extensive experience of conducting research in the field of children's social care. Sarah is responsible for the day-to-day conduct of research and is contributing to all aspects of the project.

Do parents trust how data about their family is linked together?

Rosalind Edwards, University of Southampton, Val Gillies, University of Westminster and Sarah Gorin, University of Southampton

Governments, including in the UK, are increasingly promoting electronic linking together and analysis of administrative records from education, health, social care and other public services (e.g., DCMS, 2020). Separate sources of information from different national and local public services, such as education, social care, health, welfare, housing, criminal justice, etc., can be shared between them, joined together, and then subject to algorithmic data analysis. These practices are championed by national and local government, and by data analytics companies, as offering powerful knowledge, timeliness and economic efficiency in public services delivery, thus improving outcomes for children (Edwards et al., 2022). Local authorities can use data linkage in an attempt to identify and predict which children are at risk of, for example, becoming NEET (not in education, employment or training) or involved in criminal behaviour. Central government initiated a Local Data Accelerator Fund for children and families (MHCLG, 2021) where local authorities bid for funding for data sharing and matching projects, with one city council combining 35 feeds of data from schools and other public services.

This across-the-board data sharing, electronic merging and analysis involves information about all children, parents

(including caregivers) and families. Yet there is no easily accessible means, such as a public register, for parents to find out what is happening to this data. Indeed, there appears to be little oversight of how data is being shared and linked between public services. Nor is there any process for obtaining parental consent to such use of their children's and family's data, which may override General Data Protection Regulation (GDPR) principles. What little public consultation there has been about sharing and merging of administrative records has usually focused on anonymised data for research purposes. This begs the question of whether or not data linkage and analytic practices are out of kilter with what parents think is acceptable and trustworthy use of information about their children and families.

Our Parental Social Licence for Data Linkage for Service Intervention project† aims to fill this gap, and gain a comprehensive understanding of parents' views. It focuses on social licence as the dynamics of social legitimacy and acceptance of practices that lie outside general norms, in part, sustained through trust. We commissioned a representative survey of parents of dependent children across the UK to gain a systemic overview of social licence consensus and parameters of trust (see Edwards et al., 2021), as well as conducting focus group discussions with subpopulations of parents to understand how social acceptance is articulated and negotiated, and individual interviews with parents about their experiences of family support or intervention services and use of their data. From a social licence perspective, the trust that parents may place in schools and other public services to electronically share and merge together sources of information about their children and families will relate to their assessment of the process as fair and legitimate and thus as acceptable, even if there are some apprehensions, or as suspect and discriminatory (Leonard, 2018).

In this essay we outline some of our findings about the extent to which parents from different social groups trust schools and other public services to share and electronically link data about their children and family, relating these to the wider social licence explanatory issues of legitimacy and suspicion, as well as the implications for government efforts to bring together and use administrative records from different sources.

† Funded by the UKRI Economic and Social Research Council under grant number ES/T001623/1: http://generic.wordpress.soton.ac.uk/parentdata/about; ethical approval for the research was granted by the University of Southampton: ERGO II 56997.

Do parents trust public services with their data?

A majority of the parents in our survey were aware that schools and other services collected and stored digital information about children and families (72%), but only half said that they knew that the various records could be linked together. There was overwhelming agreement that parents should be informed as to how data about their children and family were used (81%), and a strong view that they should be asked for permission for information about them from different sources to be joined together (60%). The view that parents need to give consent to whether and how schools and other services share and link information about children and families was even stronger among some groups, such as Black parents and lone parents (at 66%). Yet, while policy assertions about improving public trust in data linkage often focus on awareness-raising and transparency (e.g., DCMS, 2020), the idea that parents (and the wider public) should be asked for, and could withhold, consent to sharing and merging of public services' administrative records is not a current feature.

We asked parents about the use of data by local council education and other public services, such as early years services, children's social work teams, the police and immigration, and whether or not they trusted these services to electronically merge administrative records about children and families. Figure 1 shows the extent of parental trust. Levels of trust among all parents in how their information could be used by various public services hover around the halfway mark, or fall below it (between 55% and 35%). It is notable that trust in the way that school data about children could be used (47%) was lower than social work, early years and crime records. This may relate to parents' increasing experience of the way that schools monitor and collect data about their children, creating a culture of behaviour control (e.g., Manolev et al., 2018).

Once again, however, echoing an uneven pattern that is evident across the levels of trust in various services, there are

differences between social groups relating to their positioning in society (see also Helland et al., 2022; Jakesch et al., 2022). For example, in our survey there are differences in the extent to which parents in higher occupation, qualification and income categories trust local education services (although still only 48%) compared to parents from more marginalised groups having less trust, especially lone parents and Black parents. Indeed, these are the parents who are more likely to experience prejudice and various interventions in whether and how they bring up their children (e.g., Bywaters et al., 2017).

We now turn to material from our group discussions and individual interviews to help explain the uneven patterns in levels of trust, viewed through the social licence lens of legitimacy and suspicion.

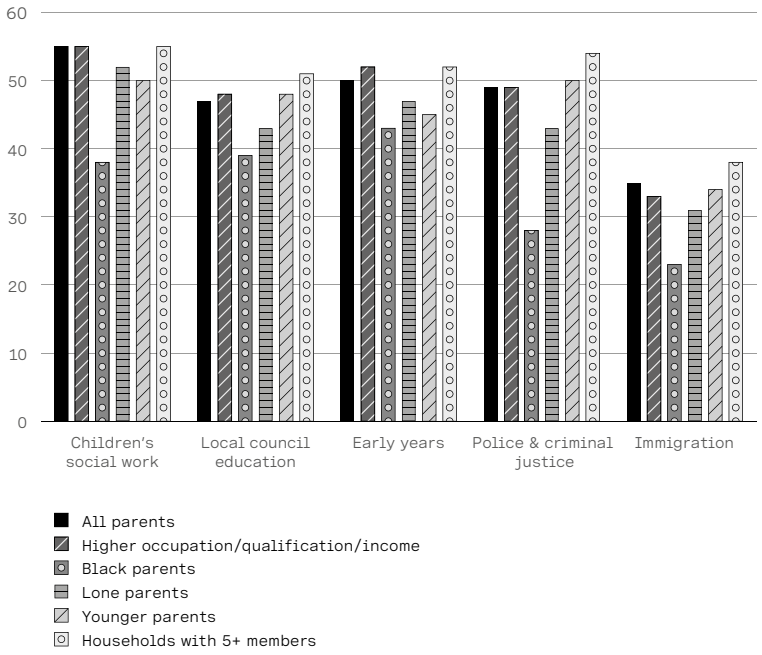


Figure 1: Parents trusting organisations to join together administrative records. Sample size: 843 parents. Source: NatCen panel survey (2020)

Legitimacy

More parents in professional occupations and from the white majority are more likely to feel that data linkage is legitimate because it will protect children in 'other' families, and to trust services undertaking the process on this basis. It is families who are in need of support or parents maltreating their children who will be the focus of electronic merging and analytics. In contrast to the families, these parents themselves have 'nothing to hide'. The interweaving of legitimate purposes and trust is evident in this exchange as part of a group discussion between parents who were working as service operation managers and coordinators in the voluntary sector, drawing on their professional experiences of working with families in need of support:

> *Manager*: We need to have more information so that people can get the right kind of services or the right support at the right time... if we [service providers] had a central place for records, then they would all be linked and that would definitely help those services...
>
> *Coordinator 1*: Yeah, because I've come from the point of view where I assumed that the system and integration was far more solidified than it actually is, you know... and there's a part of me that thinks, actually, it should've been done a long time ago and been more, sort of, coherent... on the one hand there's clear benefits for society, on the other I'm not entirely sure that we have the [data linkage] infrastructure to necessarily support it to its maximum efficacy and efficiency...
>
> *Coordinator 2*: So my background was working in schools, especially, like, the working together to safeguard children and kind of, like, everyone coming together, you know, to share information and things like that... So I think it's great in one sense, definitely, to assess and see how families could benefit in certain things. But also I think it's the other side of who can get access to the information...

Coordinator 3: None of it worries me at all. I'm kind of open-minded. I suppose I'm one of them that's the old adage of if you've got nothing to hide, what's the problem.

Coordinator 2's comment about 'who can get access to the information' also indicates how some parents in professional occupations may nonetheless have data security apprehensions, despite them judging data linkage to be a legitimate practice. These concerns coalesce around illegitimate use of their own and their family's data, and of families like them, rather than those who should be identified for intervention. Other parents also had suspicions about use of administrative records, which we now discuss.

Suspicion

Parents from marginalised social groups participating in our survey were suspicious that the information collected about children and families is not always accurate, with high levels of distrust among Black parents (79%) and lone parents (63%). The rationale underpinning such suspicion about the extent and effects of inaccurate data about children and families in school and other public service records is evident from some of our individual interviews with parents who have asked for help or received interventions for their children. Parents discovered that the data recorded about their children and families, and which would be linked into other administrative records, was incorrect. Experiences of this could run from misinformation about family structure and relationships, through lapses with potentially far-reaching consequences, to suspicions of malign intent on the part of practitioners.

One mother, for example, recounted the inaccuracies in the health-related data that the school held on her child, with serious implications for her and her family. Her child's medical condition was erroneously recorded as the mother potentially abusing her child rather than the medical diagnosis that her child had received, with the possibility that her child could be removed from her care:

Mother: I've made a subject access request... And then I was just told the other information I was asking for I was not entitled to, and they wouldn't give me a reason.

Interviewer: Oh, okay, what information was that and from whom?

Mother: It was school. And also the Council because we had TAC (Team Around the Child) meetings. It was to do with my child's medical records, what had been shared with whom. And we were never told... And I've been told by the school that they've destroyed [my child's records]... because [my child] is no longer at the primary school. But I said it's such a serious allegation, it's a safeguarding file... I think the primary school still has it... I don't know what the secondary school have on [my child]... What the Education Department at the Council are holding and primary school, with that not being accurate how can they help my child?... At the moment I'm terrified. If anybody makes another referral, they just look at what records they've got at the moment and if it's not accurate, that's what worries me.

Datafied systems are inherently subject to at least some error with false positives and false negatives, but in the case of administrative data about families and children where one service's records are linked to, merged with and used by other services, inaccurate details are compounded in their reach, with potentially serious consequences (Eubanks, 2018; Henman, 2020). Article 5(1)(d) of the UK GDPR obliges accuracy, and services could be liable to penalties or enforcement if parents pursue action over misinformation – albeit the mother in the example above has been unable to gain sight of her child's records.

Black parents, whatever their occupational category, expressed extensive distrust in how information about their children and families would be understood, judged and used, based on their knowledge of racist stereotypes and prejudice. In this exchange between Black parents working in professional

occupations, the first parent raises the issue that linking of data can lead to families being labelled retrospectively with deleterious consequences, which is then picked up by the second parent to reinforce suspicion of institutional racism in how information is understood:

> *CEO, voluntary sector*: So the issue I have is that if you have people who are then exposed to people's experiences in one agency and then another agency who's therefore supposed to be able to help them find out about certain discrepancies in the past for a family or something like that, then they could potentially make decisions about this family which could be long-lasting and impactful.
>
> *Customer services*: Yeah, I agree, especially asking about the age, the financial, the culture, background, the ethnicity. I don't know in which side they're going to look at it... My worry would be to be honest more for my children than myself. Myself I grew up in Africa and I know what my background is, I know where I came from, where my culture is. But for my kids, they're born in this country and they're raised in this country... To be honest for them the racial, ethnicity or this does not really make much more sense. But behind the closed door, that [racism] is going on. So because of actually their name, how their name is spelled, how their name is called. Even in terms of their [school] grade.

Parents also expressed other suspicions of labelling and distorted views as a result of data sharing and linking. In particular, those who were in receipt of service intervention and were interviewed individually could lack trust in schools or other public services because they felt that they and their children had been or could be judged. For example, one mother was worried that her teenage son, who had been receiving social support following a difficult divorce between his parents, might be pigeonholed by the school if his information was shared:

> I have a feeling sometimes it can paint a bit of a distorted view by sharing things with other agencies. So part of my ex-husband's family is a teacher and I remember her saying she actually has a list, a register, with all of those children who are being looked after or supported by social services. And if you're handed that list, then already it clouds a little bit of your judgement about that child. It's almost, like, 'Well is this child going to be difficult in class? Is this child going to need extra support?' So I think, yeah, that when [the support worker] was sharing as much information as she was with the school, I did think, 'I don't want [my son] to be labelled, to get a label', whatever that label would be, I didn't want [my son] to have that label.

A moratorium and meaningful dialogue

Parental trust in electronic linking of data held by public services about their children and families is bound up with considerations of information from schools and other public services being used in legitimate or suspect ways. Transparency about the merging of administrative records, and informed consent to the use of data about their children and families, is important for parents. Yet it is far from the case currently, where local authorities obfuscate and evade how they use and link data about children and families (Gillies et al., 2022).

The need for parents to provide consent is a stronger issue among some marginalised groups of parents, which raises alarms about the implications of electronic data linkage and analysis for their trust in schools and other services that their children might use. Indeed, there is strong suspicion of data linkage among marginalised social groups of parents, with some holding little trust in schools and other public services implementing data sharing. These are parents and children who are likely to be subject to labelling, stereotyping and discrimination. This lack of legitimacy and its implications should be a concern for policy prescriptions about sharing and linking children and families' administrative records, and any initiatives to mandate local authorities tracking and tracing

children across data bases through unique identifiers.†

Policymakers need to recognise that sharing and merging data about children and families, and tracking children across data systems, will be received and judged quite differently among different social groups of parents. Different social groups will see the relationship between legitimacy and trust in different ways, because they are not all positioned in the same way in society (Leonard, 2018). There are low levels of acceptability and a worrying lack of trust among marginalised groups of parents in society in linkage of data about their children and family. The issue of legitimacy seems all the more pressing when it is clear, from our discussions with parents and from other studies (e.g., Amnesty International, 2018; Vannier Ducasse, 2021), that there are errors, biases and inequalities embedded in the data sources about children and families that are being merged. This inevitably means that some parents and children's lives will be disrupted by uncalled-for scrutiny (Keddell, 2022; Leslie et al., 2020). Further, there is little evidence that data linking to identify and predict which families need intervention in order to pre-empt harm actually generates accurate knowledge (Clayton et al., 2020; Salganik et al., 2020).

At a minimum, meaningful dialogue with parents that shapes the parameters of the curation, use, sharing and linking of data from schools and other public services is required if legitimacy and trust is to be generated and actively sustained. Government and public services need to engage in greater transparency and accountability to parents, enabling them to challenge and dissent from electronic merging of their data (Redden, 2020), but again, efforts towards informing parents are likely to be received and judged quite differently among different social groups of parents (ARI Working Group 3, 2020). More fundamentally, however, a responsible question is raised for policymakers about whether or not it should be done at all. A recent United Nations Office of the High Commissioner for Human Rights' report (2021) calls for a moratorium in the use of data sharing on the basis of concerns about individual rights to privacy. Perhaps even more significant is whether or not data linkage and tracking of their children is likely to further disengage and alienate already marginalised parents, with wider implications for a cohesive and equal society.

† A recently reported suggestion from the UK Children's Commissioner for England; see Adams (2022).

Adams, R. (2022). No one knows how many children in England, says children's commissioner. *The Guardian,* 9 March

Amnesty International. (2018). *Trapped in the matrix: Secrecy, stigma and bias in the Met's Gangs database*

ARI (Areas of Interest) Working Group 3. (2020). *Rebuilding a resilient Britain: Trust in public institutions*

Bywaters, P., Kwhali, J., Brady, G., Sparks, T., & Bos, T. (2017). Out of sight, out of mind: E thnic inequalities in child protection and out-of-home care intervention rates. *British Journal of Social Work, 47*(7), 1884-1902

Clayton, V., Sanders, M., Schoenwald, E., Surkis, L., & Gibbons, D. (2020). *Machine learning in children's services: What works for children's social care*

DCMS (Department for Digital, Cultural, Media & Sport). (2020). *National data strategy*

Edwards, R., Gillies, V., & Gorin, S. (2021). Data linkage for early intervention in the UK: Parental social licence and social divisions. *Data & Policy* 3, e34

Edwards, R., Gillies, V., & Gorin, S. (2022). Problem-solving for problem-solving: Data analytics to identify families for service intervention. *Critical Social Policy, 42*(2), 265-284

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police and punish the poor*. St Martin's Press

Gillies, V., Gardner, B., Edwards, R., & Gorin, S. (2022). *Freedom of Information requests on the use of data analytics in children's services: Generating transparency*

Helland, H. S., Pedersen, S. H., & Skivenes, M. (2022). Comparing population views on state responsibility for children in vulnerable situations - The role of institutional context and socio-demographic characteristics. *Journal of Public Child Welfare*

Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration, 42*(2), 209-221

Jakesch, M., Buçina, Z., Amershi, S., & Oltenanu, A. (2022). How different groups prioritize ethical values for responsible AI. arXiv:2205.07722

Keddell, E. (2022). Mechanisms of inequality: The impact of instrumental biases in the child protection system. *Societies, 12*(3)

Leonard, P. G. (2018). *Social licence and digital trust in data-driven applications and AI: A problem statement and possible solutions*

Leslie, D., Holmes, L., Hitrova, C., & Ott, E. (2020). *Ethics review of machine learning in children's social care: Executive summary*

Manolev. J., Sullivan, A., & Slee, R. (2018). The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Leaning, Media and Technology, 44*, 36-51

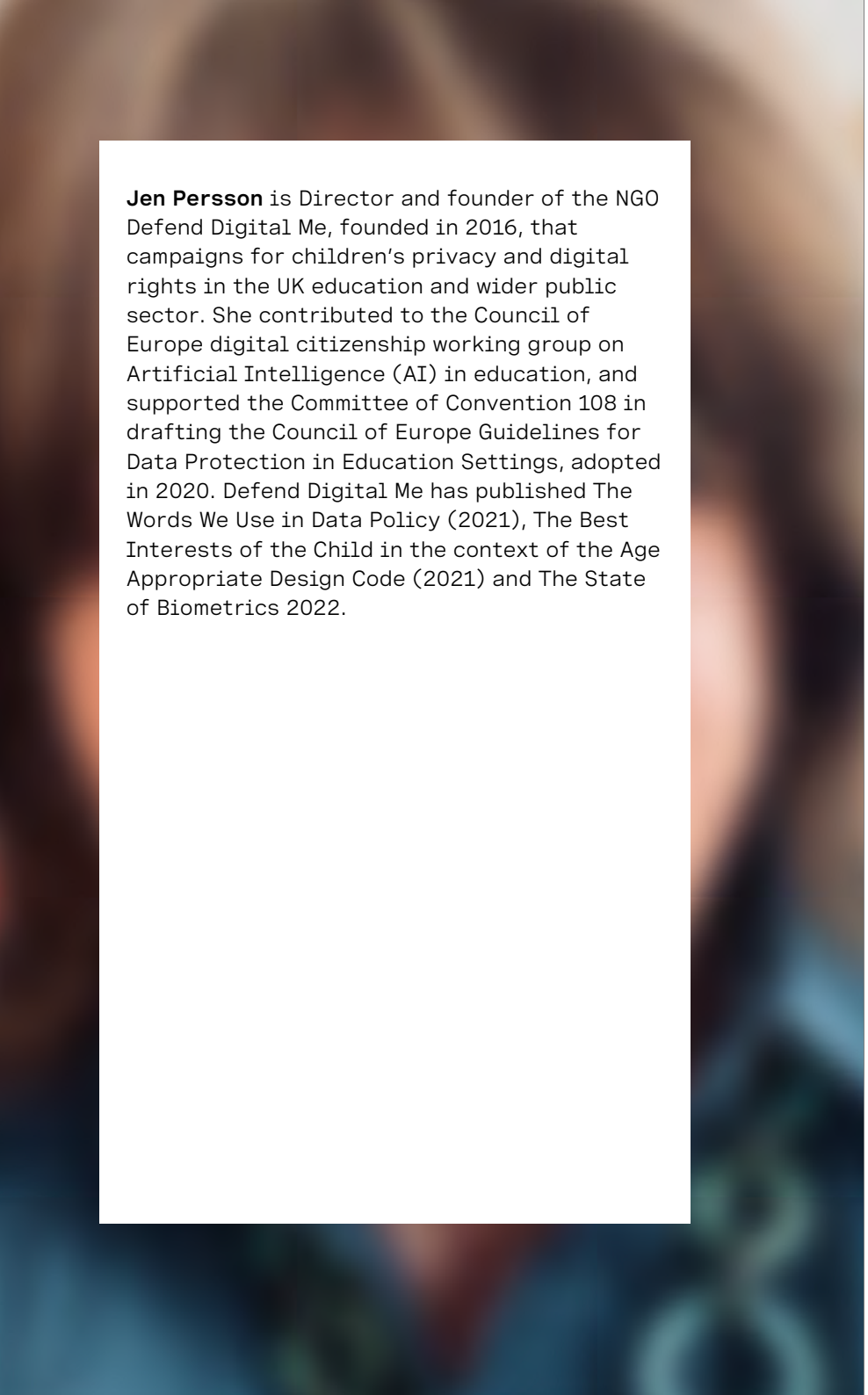
MHCLG (Ministry for Housing, Communities and Local Government). (2021). *Local Data Accelerator Fund for children and families: Prospectus*

Redden, J. (2020). Predictive analytics and child welfare: toward data justice. *Canadian Journal of Communication, 45*(1), 101-111

Salganik, M. J., Lundberg, I., Kindell, A. T., & 151 others. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences, 117*(15), 8398-8403

UN OHCHR (United Nations Office of the High Commissioner for Human Rights). (2021). *The right to privacy in the digital age*

Vannier Ducasse, H. (2021). Predictive risk modelling and the mistaken equation of socio-economic disadvantage with risk of maltreatment. *British Journal of Social Work, 51*(8), 3153-3171



Jen Persson is Director and founder of the NGO Defend Digital Me, founded in 2016, that campaigns for children's privacy and digital rights in the UK education and wider public sector. She contributed to the Council of Europe digital citizenship working group on Artificial Intelligence (AI) in education, and supported the Committee of Convention 108 in drafting the Council of Europe Guidelines for Data Protection in Education Settings, adopted in 2020. Defend Digital Me has published The Words We Use in Data Policy (2021), The Best Interests of the Child in the context of the Age Appropriate Design Code (2021) and The State of Biometrics 2022.

Building a rights-respecting environment in state education

Jen Persson, Defend Digital Me

In a world on fire, it's easy to ask why anything else matters that's not burning. Until we solve the current crises of COVID-19, conflict and climate (Beasley, 2022), we may well ask if policymakers should address anything in the education environment. But now, more than ever, at Defend Digital Me we believe that to create a better future for children needs a sense of urgency. When it comes to state education in the UK, our vision is of a rights-respecting† environment for every child without discrimination, with a high-quality standards framework for an open digital infrastructure that addresses the lack of equity, access and inclusion at local level, and with qualified due diligence in procurement processes assessing companies' integrity, technical and ethical terms. And this means change.

First, we need a common understanding of what rights we are talking about that should be respected in policy and practice before we consider what the educational environment looks like today. This will allow us to explore which rights are involved where, with whom, and how they would be actionable

† Children and young people have the same general human rights as adults and also specific rights that recognise their special needs (OHCHR, 1989).

and respected. Only then can we put in place the mechanisms to realise them.

As people's interactions with the state become increasingly automated, the UK Department for Education's (DfE) approach to digital rights matters, not only for treating children with dignity and respect to their human rights within educational settings, but also, as duty bearers, educators must teach in ways that encourage learners to become digitally literate, to flourish in society *living* with technology, as well as to manage their own use of technology while *learning* with it.

Digital literacy, and specifically algorithmic literacy, is becoming increasingly important in digital citizenship (Selwyn, 2022) for both learners and teacher training. The questions of pedagogy, design and ethics around the use of artificial intelligence (AI) (Miao et al., 2021) - and the problem of ethics washing as pseudo-legislation (Ulnicane, 2021) - apply broadly to technology in education or educational technology (EdTech) and must be addressed by all devolved UK governments. For every year that policymakers fail to get this right, more children leave school without understanding their digital footprint left behind, and with insufficient awareness of the role of data in our economy, society and daily lives (de Terwangne, 2022).

The right to education

There is a broad worldwide consensus on the importance of the right to access education (UNESCO, 2016). In the UK we hear criticism of other countries that prevent access for girls, for example, but we are yet to solve our own problems of racism in education[1] and barriers to inclusion for children with additional needs or disabilities. The DfE 2022 SEND (*Special Educational Needs and Disabilities*) review has identified significant inconsistency in how needs are met, but in all the mentions of the 'right support, in the right place and at the right time', there is no mention of the human rights of the child. This is a problem where changes are focused through the lens of impact for institutions, such as schools, rather than family or child themselves.

First and foremost, the right of every child to education is enshrined in three parts in Article 26 of the Universal Declaration of Human Rights (UDHR) going back to 1948 (UN, 1948). Later laws build on this foundation. Readers may believe they understand what is meant by the right to education, but what about its wider aims? And is it any different in the digital context? Sustainable Development Goal 4 emphasises the commitment to ensure inclusive and equitable quality education, and lifelong learning for all. States have obligations to provide free public education for children of primary age and access to education into adulthood (UN DESA, 2015).

The European Convention on Human Rights, drafted after the Second World War by the Council of Europe, added Article 2 of Protocol No. 1 to the Convention shortly afterwards. In its interpretation in cases, the European Court of Human Rights (ECtHR) (2021) has relied on the UDHR as well as other international instruments, including the Convention against Discrimination in Education (UNESCO, 1960), the *International Covenant on Civil and Political Rights* (ICCPR) (OHCHR, 1966a), the *International Covenant on Economic, Social and Cultural Rights* (ICESCR) (OHCHR 1966b) and, much later, the *UN Convention on the Rights of the Child* (UNCRC, 1989).

Article 29 of UNCRC General Comment No. 1 sets out to, 'promote, support and protect the core value of the [UNCRC] Convention: the human dignity innate in every child and his or her equal and inalienable rights' (OHCHR, 2001). This is about the holistic development of the full potential of the child, including development of respect for human rights, an enhanced sense of identity and affiliation, socialisation and interaction with others and with the environment.

There is no debate in these conventions and articles on whether the aims of education are to deliver knowledge or skills or prepare children for employment for example. The right to education is widely recognised as both a human right in itself and indispensable in realising other human rights (OHCHR, 1999). It is the same universal right, regardless of environment.

What other rights are we talking about in education?

When it comes to which rights are involved in the digital environment, discussion can quickly become reduced to exclusively 'data' or 'child protection'. Instead, the full range of human rights is relevant, and what is needed to respect them in state education must be explored in more depth.

The *right to non-discrimination, freedom of expression, freedom of thought, the right to protection for reputation and from arbitrary or unlawful interference with privacy, family, or correspondence* are all challenged in the digital environment in education, often in unseen ways, designed to influence a child's mood, cognitive or personal development (Alegre, 2017; Taylor & Rooney, 2017; ViewSonic, 2022). This demands 'accurate, nuanced and comprehensive knowledge derived from rigorous and independent research' (Council of Europe, 2022). And while there are views on how children's right to be protected from commercial exploitation (UNCRC, Article 32) could apply in the digital environment (van der Hof & Lievens, 2019), the implications for children's behavioural tracking in education and its use for-profit - such as profiling children's in-app achievement combined with adTech (targeted advertising tools) to email parents (including caregivers) marketing for products to meet the child's 'gaps' identified by the same company - is yet to be fully understood by the teaching sector or families.

It is noteworthy that it was the UK Human Rights Act (HRA), not legislation about education, that embedded Article 2 of Protocol No. 1 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, ECHR) (Council of Europe, 1950) into UK domestic law in 1998. Sadly, we see threats emerging in planned government reform of the UK HRA 1998[2] that will undermine protections for children in law. These threats make it harder to access justice through additional permissions stages and narrowing the right to respect for private and family life.

When talking about the rights of the child in education many are quick to reach for the UNCRC, which can mean that the fundamental principles of the human rights of the child (FRA, 2022) are forgotten, or 'the best interests of the child' becomes the be-all and end-all of debate. The UK HRA is arguably more important to children in England while the UNCRC remains unincorporated into domestic law, and cannot be relied on in law, compared with Wales or Scotland. However, there is no hierarchy of rights in either the UNCRC or human rights law, and no single right should be considered in isolation.

Human rights are 'universal, inherent to every individual without discrimination; inalienable, meaning that no one can take them away; indivisible and interrelated, with all rights having equal status and being necessary to protect human dignity' (Balsera, 2019). No one right trumps another and there is no pecking order of importance:

> Where, after all, do universal human rights begin? In small places, close to home - so close and so small that they cannot be seen on any maps of the world. Yet they are the world of the individual person; the neighborhood he lives in; the school or college he attends... Unless these rights have meaning there, they have little meaning anywhere. Without concerned citizen action to uphold them close to home, we shall look in vain for progress in the larger world. (Roosevelt, 1958)

The educational environment

The delivery of state education is no longer necessarily physically constrained by geography, or differences between the mainstream and alternative, or home settings. It is, in fact, the loss of boundaries between school and home in both place and time that 'digital' introduces that is perhaps the biggest change from digital adoption into the education environment. There is no longer a clear start and end to the school day or school setting that a teacher or child can leave and switch off, or where the school no longer has oversight of behaviour now that the school staff and/or EdTech company can see if and when a child is online.

Article 13 of the UN ICESCR (OHCR, 1966b) concerns itself with the wider environment and systems of education, recognising *both the importance of enabling access to education and also protecting the fundamental freedom from imposition of its form of delivery*. Article 24 of the 2006 *Convention on the Rights of Persons with Disabilities* (UN DESA, 2006), dedicated to education, recognises the importance of a *suitable* environment, the most appropriate languages, modes and means of communication, maximising academic and social development.

The education environment, including the elements of *accessibility, acceptability and adaptability* - common to education in all its forms - is set within the broader environment of a child's development, rest and play, health and the living, economic, social and political conditions in which a child grows up.

What does the digital environment look like?

Exploring international governance frameworks for definitions of what the 'digital environment in education' is reveals that the digital environment in education is interdependent on the non-digital environment. UN General Comment No. 25 (2021) on children's rights in relation to the digital environment points out that the delivery of its aims first depends on action outside the digital environment. It requires States parties to:

> ... coordinate policies, guidelines and programmes relating to children's rights among central government departments and the various levels of government... engage with schools and the information and communications technology sector and cooperate with businesses, civil society, academia and organizations to realize children's rights in relation to the digital environment at the cross-sectoral, national, regional and local levels.

Thinking of the digital environment as a single homogeneous place may be misleading. The provision of physical infrastructures is controlled by different companies and might not be owned by the educational setting, from home-school messaging communications or cloud-based data storage of information management systems handling millions of pupil records to virtual learning environments. Hardware used might be mobile phones, Chromebooks, iPads owned by schools or families, or servers with access for the local authority, police, research or national databases (defenddigitalme, 2020). What the digital environment looks like from providers' point of view is increasingly intrusive, with more and more companies accessing or monitoring pupils' personal devices and using biometrics or bodily data in 'trait and gait analysis' (defenddigitalme, 2022).

While mapping common data flows into, across and out of state education in England in *The state of data* (2020), we imagined England's school system as a giant organisational chart. How do organisations relate to one another? Which institutions does a child physically attend? Add to that the digital world children cannot see. Now imagine that 24/7, 365 days a year, every year of a child's education and long after leaving age (defenddigitalme, 2020). Unsurprisingly, parents believe they have inadequate control of their child's digital footprint in school (Survation, 2018).

Understanding those interpersonal, institutional and commercial contexts (Livingstone et al., 2019) is key to the next step in understanding what the digital environment looks like. It involves first identifying who does what in each educational activity: the owner of the digital infrastructure, who controls the decisions made in it, and whether a child, family or school staff are actively or passively involved or can access it themselves. Only then can a step in each process be inserted that is the action point for rights. Why someone interacts with the process (i.e., in admissions, daily admin tasks or academic research projects) will determine the infrastructure needed there to exercise relevant rights (i.e., receive information, make an opt-in choice or object, submit an online form).

We see three common features across what we might think of as different territories in the digital environment that overlap to varying degrees across the spectrum of data processing:

1. The child is actively present or involved in the environment, creates and may view their own data (i.e., directly in EdTech apps).

2. The child's offline activity and characteristics are digitised, edited, accessed, discussed or distributed by others (i.e., the administration of attendance, teaching and learning, assessment and attainment, behaviour management, digital safeguarding, CCTV, school census collections by the local authority and DfE).

3. Pupil or staff data is processed exclusively by 'others' (i.e., commercial pupil data analytics, data brokers selling staff details for direct marketing emails, with sending times based on recipient opening patterns of behaviour).

The same information collected during a child's learning may simultaneously be part of all three. For example, EdTech apps may first process data when a school registers a child's user profile, the pupil creates data during its use, and then the company uses the behavioural data to show advertising to parents and caregivers (Williamson & Rutherford, 2017).

In *The state of data* (2020) we identified a vast and growing number of actors involved in education, and an imbalance of power with lack of accountability no longer between only the child, family, school staff and the State, but across this unlimited number of 'others' who process both pupil and staff data (see point 3 above).

The exercise of rights

Different digital territories require different features in how to exercise rights. Some will need to hear views or act on input from the learner or parent, and others require transparency to ensure understanding of the accuracy of third party tools, that is, those used in safeguarding that claim to infer radicalisation and create child profiles.

Where a child or parent is actively present in the environment, their data rights that need to be able to be exercised, such as consent obligations or the right to object (Nottingham et al., 2022) or to request a copy of their own data (subject access request), could be made through the same user interface.

Where a child's offline activity and characteristics are digitised out of sight, to be fairly processed the data controllers must nearly always ensure the child knows it is happening. New mechanisms may be needed to provide information to the child and/or family, receive decisions from them and demonstrate they have been acted on, such as processing in the National Pupil Database.[3]

Where the learners' or school staff data is created or processed exclusively by others, that is, data brokers unbeknown to the educational setting or learners, it will be exceptional that such processing is lawful and meets the obligations of fair and transparent processing. Better enforcement or new safeguards may be needed to ensure the breach of rights is restored.

From our research, schools are at best inconsistent in enabling rights (defenddigitalme, 2020). While new standards, especially with regards to equality laws, may have been expected after the General Data Protection Regulation (GDPR) and UK Data Protection Act 2018 introduction of the Age Appropriate Design Code (AADC), it has not yet created visible change in the education sector. School data protection officers focus on the protection of the institution and less on supporting pupils and their families. Whereas NHS organisations and local authorities that provide social services must have a Caldicott Guardian† to advocate for patient rights, this is not extended to education, despite processing similarly highly sensitive data about physical and mental health, or children-at-risk or in-need. While accredited academic researchers may follow recognised and peer-reviewed ethical practice for research in the public interest (BERA, 2018), commercial companies may not.

Despite Article 12 of the UNCRC's right of the child to express their views, the Committee on the Rights of the Child recognised in 2013 that children are often politically voiceless in decisions that affect them. Children are not adequately heard on data policy about their lives (defenddigitalme, 2021) or represented in the data used for policymaking (Office for Statistics Regulation, 2022). Educational settings can restrict children's agency and autonomy in non-consensual settings, with an imbalance of power between the authority and child, but this does not remove duty bearers' obligations to uphold children's rights.

Children's capacity must be recognised when considering their rights. It must also take the rights and duties of parents into consideration, as set out in Articles 3, 5 and 18 in the

† A senior person responsible for protecting the confidentiality of people's health and care information and making sure it is used properly: www.ukcgc.uk/caldicott-guardian-role

UNCRC, especially since they have a prior right in the UDHR and ECHR to choose the kind of education that shall be given to their children, grafted onto the child's right to education. This intersection of the rights of the child and parents needs particular consideration when it comes to designing the mechanisms for accessing information, automated decisions to be explained or challenged and choices exercised, such as opt-in to commercial reuse of identifying pupil records or objections.†

How pupil data rights are made actionable is already a gap in today's practice in respecting UK data protection law. Even specific guidelines for data protection in educational settings (Council of Europe, 2020), without enforcement, are not enough to make that change.

The UK government direction of travel in reform of the Data Protection Act 2018[4] is to diminish rather than design for children's rights. Despite this, 5Rights and the Digital Futures Commission (2021) proposals for 'immediate steps to solve the education data governance vacuum in the digital environment' must be considered both necessary and urgent.

Conclusion

Building a rights-respecting digital environment in education means one founded on human rights, which are indivisible and interdependent, underpinned by the aims of education and principles of accessibility, acceptability, availability and adaptability. It must be made safe and transparent, with genuine choice and agency (Stoilova et al., 2020) to exercise rights as protected in data protection and human rights law and seek redress.

How can it be achieved? First, the DfE must proactively focus on mapping the universal processes in the delivery of state education, identifying roles and responsibilities. Next, it must identify (a) the arrangements needed in national, local or corporate infrastructure; (b) State obligations regarding the procurement impact of the business sector on children's rights (UNCRC, 2013); and (c) enable the role of parents and learners in realising their human rights in the right place at the right times through physical mechanisms to meet duty holders' obligations.

Looking at human rights in education through the lens of data protection must not mean the wider aims of education are forgotten. The second clause of Article 26 of the UDHR, the right to education, is rarely articulated in full, but is as apt today as ever:

> Education shall be directed to the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms. It shall promote understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace.

There should be no doubt that the time to build a rights-respecting environment in education is now.

† The DfE recognised the right to object to marketing in cloud-based apps in 2014, but offers no mechanisms to enable schools to exercise it or realise it in any areas of their own pupil data processing (DfE, 2014, p. 8).

Alegre, S. (2017). Rethinking freedom of thought for the 21st century. *European Human Rights Law Review, 3*(13), 221-33

Balsera, M. R. (2019). *Human rights: Universal, inalienable and indivisible*. ActionAid

Beasley, D. (2022). 'You're asking me to choose which children live and which children die', says WFP head. Channel 4 News, 14 March

BERA (British Educational Research Association). (2018). *Ethical guidelines for educational research* (Fourth edn)

Council of Europe. (1950). *European Convention on Human Rights*

Council of Europe. (2020). *Convention 108+. Guidelines for the protection of personal data in educational settings*

Council of Europe. (2022). Recommendation CM/Rec(2022)13 of the Committee of Ministers to Member States on the impacts of digital technologies on freedom of expression

Day, E. (2021). The education data governance vacuum: Why it matters and what to do about it. Digital Futures Commission

defenddigitalme. (2020). *The state of data 2020: Mapping a child's digital footprint in the state education landscape in England*

defenddigitalme. (2021). *The words we use in data policy: Putting people back in the picture*

defenddigitalme. (2022). *The state of biometrics 2022*

DfE (Department for Education). (2014). *Cloud (educational apps) software services and the Data Protection Act: Departmental advice for local authorities, school leaders, school staff and governing bodies*

DfE. (2022). *SEND review: Right support, right place, right time*

de Terwangne, C. (2022). Privacy and data protection in Europe: Convention 108 and the EU GDPR. In G. González, R. van Brakel & P. de Hert (eds) *Research Handbook on Privacy and Data Protection Law* (Chapter 1). Edward Elgar

ECtHR (European Court of Human Rights). (2021). *Guide on Article 2 of Protocol No. 1 – Right to education*

FRA (European Union Agency for Fundamental Rights). (2022) *Handbook on European law relating to the rights of the child*

Livingstone, S., Stoilova, M., & Nandagiri, R. (2019). *Children's data and privacy online: Growing up in a digital age – An evidence review*. London School of Economics and Political Science

Miao, F., Holmes, W., Ronghuai, H., & Hui, Z. (2021). *AI and education: Guidance for policy-makers*. UNESCO Digital Library

Nottingham, E., Stockman, C., & Burke, M. (2022). Education in a datafied world: Balancing children's rights and school's responsibilities in the age of Covid. *Computer Law and Security Review, 45*, 105664

Office for Statistics Regulation. (2022). *Visibility, vulnerability and voice: The importance of including children and young people in official statistics*

OHCHR (Office of the United Nations High Commissioner for Human Rights). (1966a). *International Covenant on Civil and Political Rights*

OHCHR. (1966b). *International Covenant on Economic, Social and Cultural Rights*

OHCHR. (1989). *UN Convention on the Rights of the Child*. UNICEF UK

OHCHR. (1999). *General Comment No. 13: The right to education (Article 13)* (1999). Adopted at the Twenty-first Session of the Committee on Economic, Social and Cultural Rights, on 8 December 1999 (contained in Document E/C.12/1999/10).

OHCHR. (2001). *General Comment No. 1: The aims of education (Article 29)* (2001)

OHCHR. (2021). *General Comment No. 25 (2021) on children's rights in relation to the digital environment*

Roosevelt, E. (1958). 'Close to home' – The Universal Declaration of Human Rights

Selwyn, N. (2022). What should 'digital literacy' look like in an age of algorithms and AI? *DigiGen*, 1 April

Stoilova, M., Livingstone, S., & Nandagiri, R. (2020). Digital by default: Children's capacity to understand and manage online data and privacy. *Media and Communication, 8*(4), 197-207

Survation. (2018). Poll of parents of children in school in England and Wales commissioned by Defend Digital Me

Taylor, E., & Rooney, T. (2017). *Surveillance futures: Social and ethical implications of new technologies for children and young people*. Routledge and CRC Press.

Ulnicane, I. (2021). *Artificial intelligence in the European Union: Policy, ethics and regulation*. Routledge

UNESCO. (1960). *Convention against discrimination in education*

UNESCO. (2016). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*

UN (United Nations). (1948). *Universal Declaration of Human Rights*

UNCRC (United Nations Committee on the Rights of the Child). (2013). *UNCRC General Comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights*

UN DESA (United Nations Department of Economic and Social Affairs). (2006). *Convention on the Rights of Persons with Disabilities* (CRPD), Article 24.

UN DESA. (2015). *Goal 4 Sustainable Development Goals*

van der Hof, S., & Lievens, E. (2019). The child right to protection against economic exploitation in the digital world.

ViewSonic. (2022). ViewSonic's myViewBoard Sens brings UK's first AI-powered classroom to Smestow Academy. 21 March.

Williamson, B., & Rutherford, A. (2017). ClassDojo poses data protection concerns for parents. LSE Blog, 4 January

1 Child Q and the law on strip search' (2022): https://commonslibrary.parliament.uk/child-q-and-the-law-on-strip-search

2 Human Rights Act Reform: A Modern Bill of Rights - Consultation (March 2022): https://gov.uk/government/consultations/human-rights-act-reform-a-modern-bill-of-rights/human-rights-act-reform-a-modern-bill-of-rights-consultation

3 https://find-npd-data.education.gov.uk/categories

4 https://gov.uk/government/consultations/data-a-new-directioninformation

THE VALUE OF BETTER REGULATION

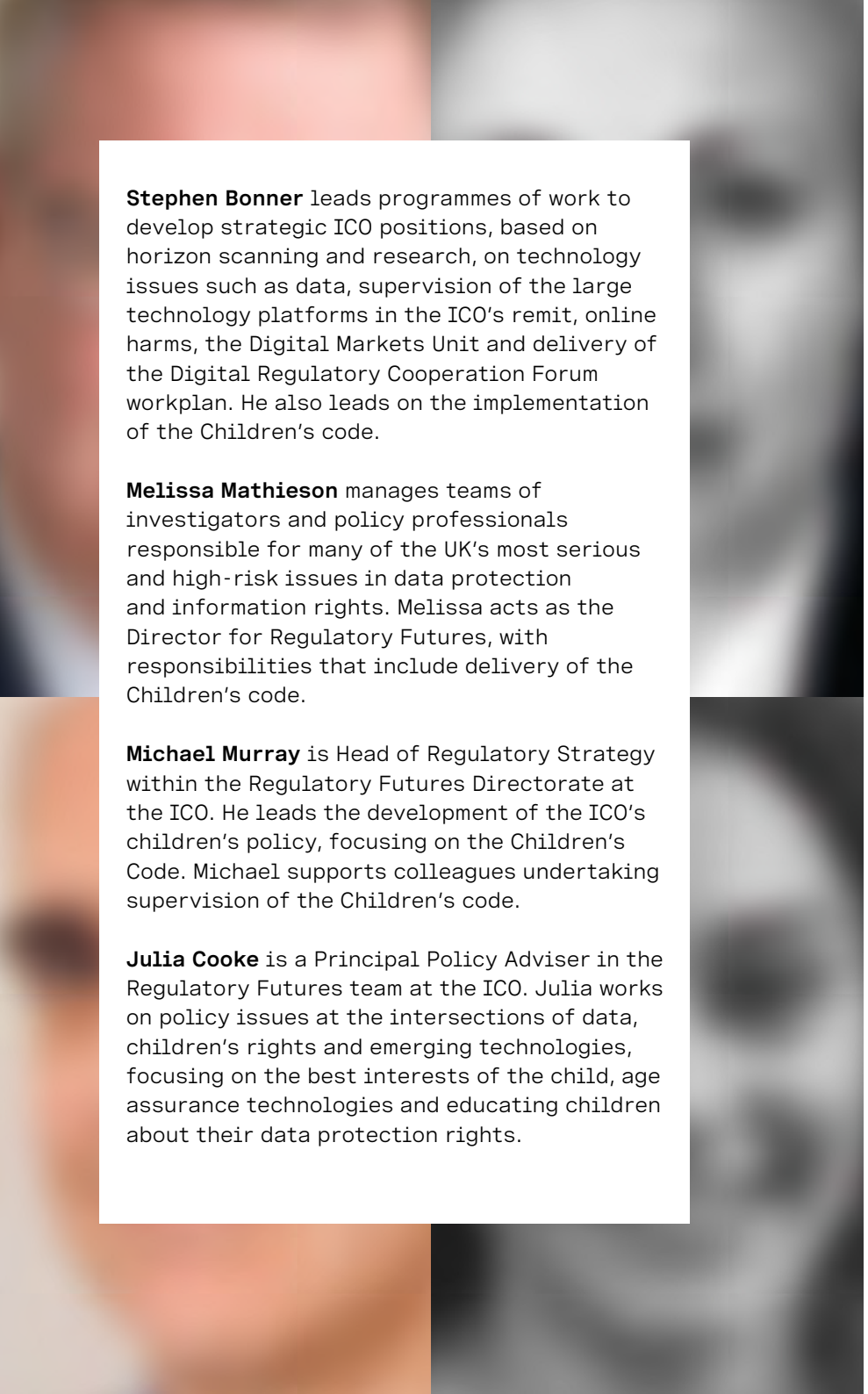
9 in 10 children consider it unacceptable for apps they use at school **to share information about you and your classmates with other companies**

ClassDojo keeps track of my points and how well I am doing and lets my mum know (Boy, 10)

I don't like when we had classes on Zoom so I don't think it helped my learning (Girl, 12)

Only 1 in 5 say their school has talked to them about what data is processed by EdTech

Digital Futures Commission (2022) EdTech survey



Stephen Bonner leads programmes of work to develop strategic ICO positions, based on horizon scanning and research, on technology issues such as data, supervision of the large technology platforms in the ICO's remit, online harms, the Digital Markets Unit and delivery of the Digital Regulatory Cooperation Forum workplan. He also leads on the implementation of the Children's code.

Melissa Mathieson manages teams of investigators and policy professionals responsible for many of the UK's most serious and high-risk issues in data protection and information rights. Melissa acts as the Director for Regulatory Futures, with responsibilities that include delivery of the Children's code.

Michael Murray is Head of Regulatory Strategy within the Regulatory Futures Directorate at the ICO. He leads the development of the ICO's children's policy, focusing on the Children's Code. Michael supports colleagues undertaking supervision of the Children's code.

Julia Cooke is a Principal Policy Adviser in the Regulatory Futures team at the ICO. Julia works on policy issues at the intersections of data, children's rights and emerging technologies, focusing on the best interests of the child, age assurance technologies and educating children about their data protection rights.

Data protection - a framework for sharing children's data in their best interests

Stephen Bonner, Melissa Mathieson, Michael Murray and Julia Cooke, Information Commissioner's Office

The Information Commissioner's Office (ICO) has long advocated data sharing that supports children's best interests and the benefits that timely data sharing can bring.† As the UK's regulator for data protection, the ICO is uniquely placed to provide insights gathered through supervision of the legislation and policy engagement.‡ This is our opportunity to demonstrate why data sharing is important and how to do it well, but also to discuss the improvements needed to ensure data sharing supports children's best interests. We advocate a framework for sharing children's data that balances the risks of sharing data (such as excessive, inappropriate sharing) with the risks of not sharing data (such as not being able to make informed decisions or effectively act on crucial information). Such a framework would enable children to experience benefits including improved access to services, targeted help and support, and better projected and actual life outcomes (ICO, 2020a, b; National Infrastructure Commission, 2017).

† The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

‡ The ICO's policy engagement spans all sectors and industries that process children's data, from health, social services, education, central and local government to recreational and online services. It has undertaken research to inform these recommendations, including with parents, caregivers and children (ICO, n.d., a, b). It also funded research led by Professor Sonia Livingstone on children's data and privacy online (ICO, n.d., c).

Step-change in children's digital footprints

Personal data powers innovative technologies and online commerce. With many of us now leading the majority of our daily lives online, personal data quantifies our behaviour, our interests, our spending patterns, our loves and likes, our beliefs, our health, sometimes even our DNA – the very blueprints that make us who we are. However, any economic and societal benefits from sharing data are only sustainable if people have confidence and trust in how their data is used. This is never more so than when considering the interests of children and vulnerable people.

In 2003, only half of UK homes were connected to the internet (Ofcom, 2021). By 2021, Ofcom (2021) found that 97 per cent of all children aged 5–15 went online. This step-change in digital usage has accelerated concerns about protecting personal data, especially when children are creating digital footprints and sharing data from a young age that follow them into adulthood on a scale previously unseen (Lupton & Williamson, 2017). The rise in online learning, including automated decision-making and artificial intelligence (AI)-powered teaching aids, as a consequence of the COVID-19 pandemic, has further exacerbated these concerns.

The internet was not designed for children's use or with their best interests in mind. Children can be unaware of the impact sharing their data has, as seen with prominent celebrities facing repercussions for comments they made as children (Ritschel, 2019; Watson, 2021). As well as having less understanding of how their data is used and what their rights are, the unequal power differential between organisations and data subjects means children are often less empowered to complain about misuse of their personal data, and frontline, child-focused services are often unaware of how data collected on children by digital service partners can be shared with third parties. It is imperative that we protect and educate children within the digital world, and ensure all stakeholders are aware that data protection is fundamental to supporting children's rights and making the internet, and data sharing, better for children.

The legislative framework

The UK's data protection regime, comprised of the UK General Data Protection Regulation (GDPR) and the Data Protection Act 2018, creates a framework that enables child-centric, fair, necessary and proportionate data sharing to take place in a way that safeguards children and supports their best interests.† The ICO has developed two statutory codes of practice to support organisations in sharing children's data lawfully: the Data Sharing Code of Practice (ICO, 2020b), which includes a section specifically focused on sharing children's data (ICO, 2020d), and the world-leading Children's Code (Age Appropriate Design Code; ICO, 2020e). The success of the Children's Code has caught the attention of other countries, which are now looking at changes to their legislative frameworks, as seen in Ireland with the Fundamentals, in the Netherlands with the Code for Children's Rights and most recently, in the USA with the California Age Appropriate Design Code Act.

The Data Sharing Code of Practice helps organisations balance the benefits and risks in order to implement successful data sharing. It demonstrates that the legal framework is an enabler to responsible data sharing, and busts some of the myths that currently exist. Organisations can use it to guide them through each step of the data sharing process (ICO, 2020c). For example, the ICO's Innovation Hub provided advice and guidance to an organisation designing solutions that used Open Banking to divert a small, adjustable portion of income into a hidden account, in order to increase women's financial independence and empowerment.‡ This data sharing focused

† The UK Government is reviewing the data protection regime to ensure it is fit for purpose and underpins the trustworthy use of data (see DCMS, 2021). The ICO welcomes this opportunity to review the state of play three years on from the introduction of the GDPR to ensure the framework continues to support data sharing in children's best interests. Organisations are encouraged to engage with this review, and the ICO has published its response (ICO, 2021a). This demonstrates the ICO's support for proportionate ways organisations can demonstrate their accountability for how they collect, store, use and share data. Organisations must ensure data is safe and not used in ways that might cause harm, that all people, including children, are able to exercise rights over their personal data.

‡ Open Banking (www.openbanking.org.uk) enables customers to allow organisations to receive data directly from their bank with their explicit consent (FCA, n.d.). Research has previously found that 60 per cent of women in refuges had children with them, so using Open Banking to share data in this instance has the potential to support children in households experiencing domestic abuse by financially enabling them to flee violence and the known negative impacts of living in a home with domestic violence (NSPCC, 2021; Women's Aid, n.d., 2022).

on enabling customers in controlling or abusive relationships to retain some financial independence, as finance is a key barrier cited by victims that prevents them from leaving (Butt, 2020; Women's Aid, 2022).

What does 'good' data sharing look like?

Too often, harm and detriment are caused to children where data is not disclosed for fear of breaching data protection, is shared too slowly or without due consideration of the risks, potentially rendering the sharing ineffective, or the data biased or inaccurate, leading to flawed decision-making.[†] Data protection helps organisations to confidently share data correctly, efficiently, safely and in support of children's best interests.

Successful data sharing:

- Is underpinned by a data protection impact assessment (DPIA)
- Takes a child-centric, holistic approach
- Builds on existing best practice
- Involves collaboration with other parties.

DPIAs

Organisations can use DPIAs to identify and minimise the data protection risks of any processing operation. Article 35 of the UK GDPR specifies several circumstances where it is necessary for organisations to complete DPIAs, including where there is large-scale processing of special category data, which includes children's data (ICO, n.d., d). Standard two of the Children's Code requires organisations to complete DPIAs in order to process children's data.[‡] It is a living document rather than a one-off process, and should be regularly reviewed and updated. For example, having an emergency plan in place that considers data sharing can help prevent any delays in a crisis and get children the emergency support they need (ICO, 2021b).[§]

† Such harm and detriment can range from not receiving care and support, emotional distress, unwarranted intrusion on families or discrimination to loss of life in extreme cases; see DfE (2016).

‡ See the ICO's DPIA template (ICO, 2020h).

§ Schools should also have plans in place for emergencies, as outlined in DfE (2018).

Child-centric holistic approach

The United Nations Convention on the Rights of the Child (UNCRC) recognises that children need special safeguards and care in all aspects of their life, and that these should be guaranteed by appropriate legal protections (UN OHCHR, 1989). In the UK, the Children's Code ensures domestic data protection laws truly transform the way children are safeguarded when they access online services. This means that the best interests of the child are a primary consideration when designing and offering services to children.

The ICO has developed a framework to assist industry to apply the principles of the UNCRC, so they can demonstrate their decision and justification for processing personal data, their consideration of risks and measures to mitigate any risks identified (ICO, n.d., e). If these considerations are not undertaken, it's likely their processing will not be in compliance with data protection and will not be considering the best interests of the child. This framework can be integrated into DPIAs and organisations empowered to substantively and holistically centre children in their considerations. Open sharing of children's data that is child-centric and takes a holistic approach that enables early intervention (ICO, n.d., f) highlights the need for a culture that supports and facilitates appropriate data sharing for early intervention and how to practically implement this, such as clear, designated points of contact for sharing.

Promoting the best interests of the child aligns with schools' educational role. They must comply with data protection legislation, and the Children's Code sets out what good practice compliance looks like in the areas it covers. All organisations should be encouraged to meet the Code's standards as a matter of general good practice. Doing this will ensure the schools' and their digital services providers' processing of personal data centres on the child's best interests and supports their learning.

Build on existing best practice

When creating data-sharing systems, organisations should build on existing best practice examples. This will enable data sharing that is safe, secure and timely, in line with the highest

possible standards. For example, the Welsh Government-funded Wales Accord for Sharing Personal Information (WASPI) provides a toolkit for data sharing to support delivery of frontline services by the public sector and other partners.[1] One of the principles of WASPI is that quality-assured information-sharing agreements are published so other partnerships can refer to them when developing similar proposals. Data sharing through WASPI also supports coordinated delivery of services for children and teenagers with special needs, and for those who may be at risk of neglect, abuse, exploitation by criminals or radicalisation, and those who have gone missing.

Multistakeholder approach

A multistakeholder approach has several benefits for children that can lead to a joined-up offer that better supports different aspects of a child's life. In a child welfare context, this allows specialist services to investigate potential harm and put in place tailored support for children that is appropriate to their individual needs. It can also reduce gaps in knowledge that could lead to a risk of harm to a child, and increasing the efficiency of sharing data.

The ICO engaged with the Northern Ireland Department of Justice on new Regulations to safely facilitate the sharing of personal data in relation to incidents of domestic violence.† These Regulations, which came into force on 1 April 2022, enable early intervention safeguarding to support children and young people experiencing domestic abuse. They enable schools to provide immediate support to impacted pupils and support their best interests. This initiative includes multiple stakeholders, including the Education Authority, the Police Service of Northern Ireland, the Safeguarding Board of Northern Ireland and a number of nurseries and schools, empowering them to share data to support children and minimise negative impacts on them.

Organisations need to demonstrate effective accountability and transparency, ensure the accuracy of data and work to increase confidence and trust in how they share data.

Accountability

The UK GDPR accountability principle means that organisations must be able to demonstrate how they comply with the law, including by demonstrating how their data sharing is proportionate to the risks to children associated with the data sharing (ICO, n.d., g). For example, organisations must:

- Complete a DPIA for sharing children's data
- Have contracts or agreements in place that define the responsibilities of the various organisations
- Detail the lawful basis for processing and sharing data
- Provide privacy information to children about how their data is used

In 2020, the ICO audited the Department for Education (DfE), focusing on their use of data compiled into the National Pupil Database (NPD).† This audit did not find any instances where data protection legislation impeded data sharing or placed barriers on the use of data in the public interest, although it did identify risks arising from data sharing without sufficient controls.‡ There was limited oversight and consistency around how data was shared externally, with no formal, consistent assessments carried out about the purpose, legal basis and risks of sharing the data.§ For example, only 12 instances of data sharing were rejected out of 400 applications, largely because the data-sharing process was designed to find a legal gateway to 'fit' the application, rather than a holistic assessment of the application against a set of robust measures designed to provide assurance and accountability that the sharing was lawful and in line with statutory requirements.

In a similar but separate instance, a lack of controls was a key concern with how the police and local councils processed

† The Domestic Abuse Information-sharing with Schools etc. Regulations (Northern Ireland) 2022 (NI) (UK). See also Campbell (2022). These are modelled on existing regulations that have been in place in England for over 10 years, and demonstrate the value of adopting best practice.

† This is not a database in its own right, but is made up of links to various other databases or collections of data, including the school census, the Early Years Foundation Stage Profile (EYFSP) and Children In Need census (CIN). See DfE (2022) and ICO (2020f).

‡ These risks included data sharing that was not in compliance with data protection legislation, which is a risk to the data subjects, who, in this instance, are children, and also the data controller.

§ For further information, see ICO (2020g).

and shared information on young people suspected of being involved in gang violence. Data was inappropriately shared with several organisations, resulting in the withdrawal of services and opportunities. Many individuals were under 18 and not all were accused of any crime, but rather, some were victims of crime. This case illustrates the need for an organisation to demonstrate accountability when it shares data, and to have sufficient policies and protocols in place to enable proportionate, secure and accurate data sharing in the public interest (ICO, 2020b).

Data minimisation

Organisations must apply data minimisation to their processing, including data sharing. This means data must be adequate, relevant and limited to what is necessary. UK GDPR requires organisations to:

- Be clear about the purposes for which they collect personal data
- Only collect the minimum amount of personal data needed for those purposes
- Only store that data for the minimum amount of time required

Data minimisation is a central concern around children's data, as seen with moves to use biometric data in schools to facilitate the provision of services, in particular the use of facial recognition technology to enable contactless payments for school lunches (BBC News, 2021). To comply with data minimisation, organisations should use the least intrusive measure available to achieve the processing goal, and should consider whether any use is necessary and proportionate prior to processing.

Better transparency and education

Transparency is fundamental to people's trust and confidence in how their data is used; it is a crucial foundation for successful data sharing. However, transparency is often absent, especially in schools, leaving people in the dark about how children's data is being shared. Coupled with a lack of education around the use of their data, this can leave children or their parents or caregivers disempowered, unaware of and unable to assert their data protection rights.

A 2018 review of the 1200 highest ranked apps targeted at children from the Google and Apple app stores found the average reading age for privacy policies was 13 - four years above the average reading age for an adult in the UK of nine (Das et al., 2018). Lack of transparency is therefore of particular concern when it comes to processing children's data. Children can be less aware of the risks involved in their data being processed and shared, and this can impact them in ways they don't expect (Wang et al., 2019). The ICO has published school resources for teachers to use when educating children about personal data and how it is used.[2] These empower children to know their rights and how to assert them.

The ICO recently called for transparency champions to find good practice in providing accessible, easy to understand transparency information to children (ICO, 2021c). This produced five recommendations:

1. Be creative with format - but avoid style over substance
2. Put children's needs and views at the heart of the design process
3. Meet children and parents where they are
4. Unbundle privacy information for engagement and understanding
5. Create space for meaningful parent-child conversations

Organisations should embed these recommendations into their approach to sharing children's data, and tell children about this in order to improve trust and confidence in their data sharing.

Increased confidence in sharing data

It is crucial that organisations develop a culture whereby staff feel empowered to share data safely, but to do this, staff need to have received appropriate training.

An ICO audit of Multi Academy Trusts (MATs) (ICO, n.d., h) found that 70% did not include training for all staff on key areas such as data protection, data sharing or requests for personal data. Forty per cent either hadn't allocated adequate resources to deliver such training or staff had not received appropriate training in order to train other staff (ICO, n.d., h). One organisation directly provided training to all staff, including temporary and contract staff such as supply teachers. This ensured all staff had the same level of training and awareness in order to successfully share data in ways that complied with data protection and supported children's best interests.

The 2011 Munro review of child protection highlighted that to ensure the sharing of data to support children's best interests, we need to move towards a child protection system with 'greater trust in, and responsibility on, skilled practitioners at the frontline' (Munro, 2011). The UK Government's 2018 information sharing advice for practitioners reaffirms this, highlighting that for data sharing to be successful, practitioners should be confident about the processing conditions. It starkly notes that 'poor or non-existent information sharing is a factor repeatedly identified as an issue in Serious Case Reviews (SCRs) carried out following the death of or serious injury to, a child. In some situations, sharing information can be the difference between life and death' (HM Government, 2018). The ICO's Data Sharing Code of Practice emphasises that 'sometimes, it can be more harmful not to share data' and includes case studies where data sharing is needed, particularly in relation to vulnerable children or safeguarding purposes.

Standard 9 of the Children's Code notes that organisations should not share children's data unless they can demonstrate a compelling reason to do so, taking account of the best interests of the child. As there is some confusion about what might constitute a compelling reason to share data, the ICO has explicitly highlighted examples in the Data Sharing Code of Practice. For example, data protection law does not prevent data sharing that is necessary to prevent serious harm to a person, prevent the loss of human life or the effective safeguarding of children.

Accuracy and rights related to automated decision-making

Finally, any decision or action taken is only as good as the data it is based on, so data must be accurate.† This means that data is not only up to date, but also impartial, with any bias sufficiently mitigated (ICO, 2020i). Research highlights the long-lasting impact that follows children into adulthood when personal data is incorrectly, or insensitively, recorded in health and social care files, so it is imperative that data is recorded accurately, particularly when sensitive or relating to subjective judgements (Antcliffe, 2021).

Accuracy of data is particularly important when shared or used as part of automated decision-making, as any impacts from inaccuracy are compounded. For example, a study on children's social care found that automated models missed four out of every five children at risk, and when identifying children at risk, were wrong on six out of ten occasions, meaning that such an inaccurate identification could be added to a child's social care file (What Works for Children's Social Care, 2020). This demonstrates the compelling need for these systems to be built with data protection by design and default,‡ and the need for parents, caregivers and children to be empowered to assert their rights in relation to automated decision-making (Edwards et al., 2021; Redden, 2020; UN OHCHR, 2021).§

Conclusion

To successfully share children's data in their best interests, organisations should carry out a DPIA; take a child-centric, holistic approach; build on existing best practice; and collaborate with other stakeholders. Organisations must ensure effective accountability is in place; minimise the amount of data being processed; be transparent and educate children; ensure the accuracy of data; and create a culture where staff are empowered to safely share data.

Data protection legislation provides a child-centric, proportionate, flexible and risk-based approach to sharing data that supports and empowers organisations to decide for

† Article 5(1)(d) and Article 16 of the UK GDPR require data to be accurate.

‡ Article 25 of the UK GDPR.

§ Article 22 of the UK GDPR provides these rights.

themselves how, what and when to share data. Crucially, it enables an approach that balances the risks of sharing data with the risks of not sharing data, centring a child's best interests in these decisions. The ICO is committed to supporting organisations sharing personal data in children's best interests and highlighting good practice. It recently published additional data-sharing case studies following engagement with Ofsted, and worked with a design company to redesign best interests guidance for organisations (ICO, n.d., e, f).

The ICO's Data Sharing Hub and Children's Code resources house a suite of non-statutory resources for stakeholders including toolkits, case studies, frequently asked questions and guidance. These all demonstrate what 'good' looks like, and organisations are encouraged to use these hubs as the launchpad for any data-sharing initiatives they may want to undertake.

Antcliffe, M. (Host) (2021). Reflections on accessing care records and supporting good recording. Research in Practice podcast, 12 March

BBC News (2021). Schools pause facial recognition lunch plans. 25 October

Butt, E. (2020). *Know economic abuse: 2020 report*. The Co-operative Bank and Refuge

Campbell, N. (2022). Domestic abuse alert scheme extended to 77 more schools in NI. *Belfast Telegraph*, 2 February

Das, G., Cheung, C., Nebeker, C., Bietz, M., & Bloss, C. (2018). Privacy policies for apps targeted toward youth: Descriptive analysis of readability. *JMIR mHealth and uHealth*, 6(1), e3

DCMS (Department for Digital, Culture, Media & Sport). (2021). *Data: A new direction*. 10 September

DfE (Department for Education). (2016). *Information sharing to protect vulnerable children and families: A report from the Centre of Excellence for Information Sharing*. July

DfE. (2018). *Data protection: A toolkit for schools*. July

DfE. (2022). *Complete the school census: Data items 2021 to 2022*. 14 January

Edwards, R., Gillies, V., & Gorin, S. (2021). Data linkage for early intervention in the UK: Parental social license and social divisions. *Data & Policy, 3*, E34. doi:10.1017/dap.2021.34

FCA (Financial Conduct Authority). (no date). Women's economic empowerment. TechSprint 2021

HM Government. (2018). *Information sharing: Advice for practitioners providing safeguarding services to children, young people, parents and carers*. July

ICO (Information Commissioner's Office). (2020a). The benefits of sharing personal data - What can we learn from Open Banking? Blog, 6 January

ICO. (2020b). *Data sharing: A code of practice*

ICO. (2020c). Navigating the data sharing code. In *Data sharing: A code of practice*

ICO. (2020d). Data sharing and children. In *Data sharing: A code of practice*

ICO. (2020e). *Age appropriate design: A code of practice for online services*

ICO. (2020f). Annex D: DPIA template. In Age appropriate design: A code of practice for online services. September

ICO. (2020g). Department for Education (DfE) - *Data protection audit report*. February

ICO. (2020h). Statement on the outcome of the ICO's compulsory audit of the Department for Education. 7 October

ICO. (2020i). Guidance on AI and data protection. 30 July

ICO. (2021a). *Response to DCMS consultation 'Data: A new direction'*. 6 October

ICO. (2021b). Sharing personal data in an emergency - A guide for universities and colleges. Blog, 14 September

ICO. (2021c). *Designing data transparency for children: Insights from the children's code transparency champions open call*. June

ICO. (no date, a). Background to the children's code | About the ICO

ICO. (no date, b). Research and reports | About the ICO

ICO. (no date, c). London School of Economics and Political Sciences (LSE) | Grants programme

ICO. (no date, d). Examples of processing 'likely to result in high risk' | Guide to the GDPR

ICO. (no date, e). Best interests of the child self-assessment | Children's Code Hub

ICO. (no date, f). Case studies and examples | Data Sharing Information Hub

ICO. (no date, g). Accountability and governance | Guide to the GDPR

ICO. (no date, h). *Findings from the ICO's consensual audits of 11 multi academy trusts, September 2018 to October 2019*

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society, 19*(5), 780-794

Munro, E. (2011). *The Munro review of child protection: Final report – A child-centred system*. May. Department for Education

National Infrastructure Commission. (2017). *Data for the public good*

NSPCC. (2021). Protecting children from domestic abuse

Ofcom. (2021). *Online nation, 2021 report*. 9 June

Redden, J. (2020). Predictive analytics and child welfare: Toward data justice. *Canadian Journal of Communication, 45*, 101-111

Ritschel, C. (2019). Justin Bieber acknowledges using racial slur as a teen while asking fans to 'stand against racism'. *Independent*, 4 December

UN OHCHR (United Office of the High Commissioner for Human Rights). (1989). *Convention on the Rights of the Child*. 20 November

UN OHCHR. (2021). *The right to privacy in the digital age: Report (2021)*

Wang, G., Zhao, J., & Shadbolt, N. (2019). Are children fully aware of online privacy risks and how can we improve their coping ability? *ArXiv*. abs/1902.02635

Watson, P. J. (2021). England cricketer gets international ban for edgy tweets he posted when he was a teenager. *Summit News*, 7 June

What Works for Children's Social Care. (2020). *Machine learning in children's services: Summary report*. September

Women's Aid. (2022). *The domestic abuse report 2022: The annual audit*

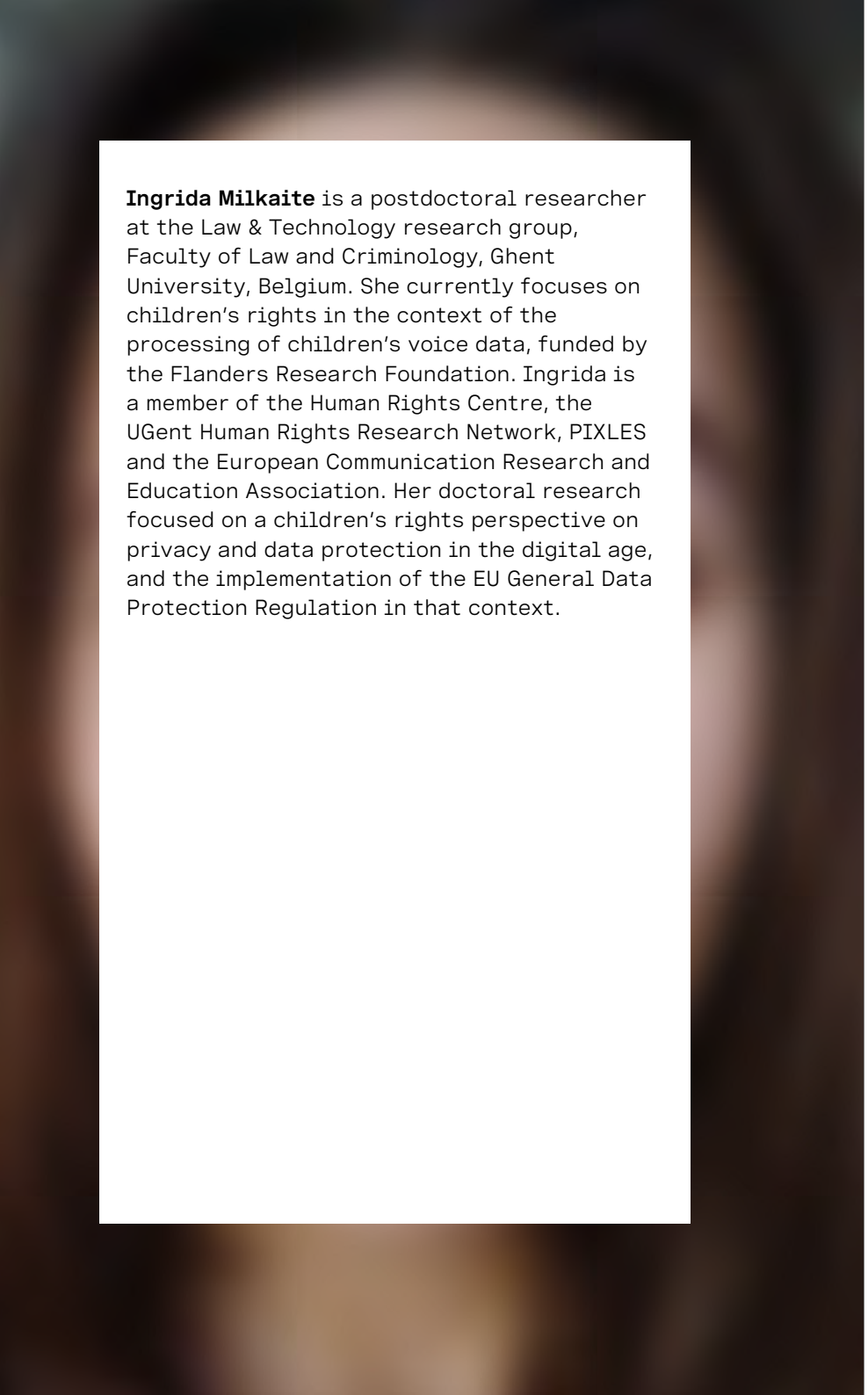
Women's Aid. (no date). What is domestic abuse? Impact on children and young people

Legislation

Domestic Abuse Information-sharing with Schools etc. Regulations (Northern Ireland) 2022 (NI) (UK)

1 http://waspi.org/information-sharing-protocols

2 https://ico.org.uk/for-organisations/posters-stickers-and-e-learning/school-resources



Ingrida Milkaite is a postdoctoral researcher at the Law & Technology research group, Faculty of Law and Criminology, Ghent University, Belgium. She currently focuses on children's rights in the context of the processing of children's voice data, funded by the Flanders Research Foundation. Ingrida is a member of the Human Rights Centre, the UGent Human Rights Research Network, PIXLES and the European Communication Research and Education Association. Her doctoral research focused on a children's rights perspective on privacy and data protection in the digital age, and the implementation of the EU General Data Protection Regulation in that context.

An international perspective on data protection for children's education data

Ingrida Milkaite, Ghent University

Today's children are 'datafied' as soon as they are born or undergo their first medical scans in utero (Lupton & Williamson, 2017). Their families further develop their digital traces, resulting in 80% of children younger than two having a digital footprint in Western countries (UN General Assembly, 2021, para. 86). As they grow older, their digital records continue to expand exponentially - in the home, social and school environments.

In recent years (and especially during the COVID-19 pandemic), we have witnessed a growing reliance on educational technologies, or EdTech. Many believe that the use of EdTech can benefit teachers and students (UN Committee on the Rights of the Child, 2021), but it is crucial to remember that the use of most digital technologies is intrinsically linked with the processing† of children's personal data by the various actors providing them.

In the words of the UN Special Rapporteur on the right to privacy, children's 'immersion in the ever-expanding range of

† 'Processing' can be defined as 'any operation or set of operations performed on personal data, such as but not only the collection, storage, preservation, alteration, retrieval, disclosure, making available, erasure or destruction of, or the carrying out of logical and/or arithmetical operations on such data' (CoE, 2020a, p. 7).

digital technologies produces an ongoing stream of data, collected and enhanced by artificial intelligence, machine-learning applications and facial and speech recognition technologies' (UN General Assembly, 2021). This context calls for the vigorous implementation of data protection and children's rights law to guide the development and use of EdTech.

Processing children's personal data in the educational environment

In November 2020, the Council of Europe (CoE) adopted specific guidelines on children's data protection in an education setting (CoE, 2020a). It specifically called for recognition of the 'breadth of personal data that may be processed, its wide uses including in support of learning and non-learning aims, for administration, behavioural management and teaching purposes, its sensitivity, and the lifelong risks to privacy that may arise from processing both non-digitised and digitised records in an educational setting' (CoE, 2020a, p. 5). This is important as children's education data is now not only provided by children themselves, their parents, caregivers and teachers, but is also deduced from 'data that is created as a by-product of user engagement or data that is inferred (for instance on the basis of profiling)' (CoE, 2020a, p. 11).

The processing of children's personal data in the educational setting† used to focus on 'routine' monitoring of the security and physical movement of the pupils (Lupton & Williamson, 2017). Now, in addition to relying on cameras to keep an eye on students and (unwelcome) visitors to schools, biometric tracking technologies are also increasingly employed, for example, facial or voice recognition, and iris, fingerprint or palm vein scanning (Alba, 2020; Leaton Gray, 2018; Steeves et al., 2018). The Information Commissioner's Office (ICO) – the UK data protection authority (DPA) – recently decided to intervene and investigate concerns about the use of facial recognition technology on pupils queuing for lunch in the canteens of nine schools (Weale, 2021). The use of such technologies in schools has already led to fines for unlawful processing of children's personal data in Sweden, and has been banned altogether in France (IAPP, 2019; Lee, 2019).

Today's education data landscape is ever-expanding and has generally been shifting towards the routine collection and analysis of children's increasingly sensitive personal data (Lupton & Williamson, 2017; Taylor, 2013). This change is associated with processes such as data or learning analytics,† e-learning platforms,‡ behaviour monitoring programmes and ever-growing educational databases. Consequently, children's learning data may now include 'thinking characteristics, learning trajectory, engagement score, response times, pages read, and videos viewed' (UN General Assembly, 2021, para. 107).

In addition to monitoring students' academic progress, some use 'emotional learning analytics' that can 'make extensive use of psychometrics, sentiment analysis and natural language processing [and] employ other data sources such as face cams, video, eye tracking, skin temperature and conductivity to enable the automatic detection, assessment, analysis and prediction of the emotional state of learners' (Lupton & Williamson, 2017, p. 785). Such 'learning analytics' constitute one of the most significant forms of child tracking in the contemporary educational setting since these technologies 'mine data about learners as they go about educational tasks and activities in real time and provide automated predictions of future progress that can then be used as the basis for intervention and pre-emption' (Lupton & Williamson, 2017, p. 785).

Another related issue concerns children's profiling in the educational environment. Profiling can be understood as:

> any form of automated processing of personal...

† Here, 'educational setting' or 'educational context' refers to schools attended by children under the age of 18. However, the discussed issues, concerns and recommendations are also very relevant and, in many cases, applicable in the context of other educational institutions, such as universities and colleges.

† 'Data analytics' 'refers to personal data used in the computational technologies that analyse large amounts of data to uncover hidden patterns, trends and correlations, and refers to the whole data management lifecycle of collecting, organising and analysing data to discover patterns, to infer situations or states, to predict and to understand behaviours', while 'learning analytics' 'can be described as the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimising learning and the environments in which it occurs' (CoE, 2020a, pp. 6, 7).

‡ The term 'e-learning' 'may broadly include learning with the support of information and communication technologies (ICT), especially for delivery or accessing of content, distance learning or web-based learning (including tools used in online and offline modes)' (CoE, 2020a, p. 7).

> ... data including use of machine learning systems consisting of the use of personal or non-personal data to evaluate certain personal aspects relating to an individual, in particular to analyse or predict aspects concerning that person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements. (CoE, 2020a, p. 7)

While profiling may be used for evaluating and personalising education, it is also associated with data - that children can hardly challenge - being hardcoded into their profiles and potentially following them throughout their educational and professional paths for the rest of their lives (Livingstone et al., 2021). Given that a profile 'refers to a set of characteristics attributed to an individual, characterising a category of individuals or intended to be applied to an individual' (CoE, 2020a, p. 7), profiling children may lead to 'sorting them into boxes'. Due to an algorithm's categorisation of a specific child, children may be provided with limited information and education opportunities, negatively affecting their rights to non-discrimination, development, identity, education and information. According to the UN Special Rapporteur on the right to privacy, profiling children 'limits their potential self-development in childhood, adolescence and possibly adulthood, as behavioural predictions and nudging techniques can predetermine options and choices' (UN General Assembly, 2021, para. 92). Therefore, EdTech needs 'to be assessed against children's rights and best interests' (ibid).

While the educational environment is associated with the processing of children's personal data on an increasingly large scale and over long periods of time (CoE, 2020a, p. 17), the exact impact of data-fuelled EdTech learning on children's education, development and future lives is neither clear, nor foreseeable. Few voices are questioning the actual benefits these technologies may bring for children since many of the products and services seem very appealing as they promise enhanced learning and personalised education for individual children. However, current automated processes, decisions and predictions on children's educational trajectories are very opaque and difficult to understand for children, parents, caregivers, teachers and schools. Despite their appeal and anticipated potential, the exact learning outcomes for children using EdTech are currently contentious and unproven (Defend Digital Me, 2020; Livingstone et al., 2021).

'Datafication' of children for commercial goals

Complex revenue models operate behind the various fun activities accessible to children online - including through EdTech - 'creating value for companies by feeding children's data into algorithms and self-learning models to profile them and offer personalised advertising or by nudging children to buy or try to win in-app items' (van der Hof et al., 2020, p. 833). Notably, the national DPAs in the UK and Ireland have recently expressed serious doubts as to whether commercial interests can be reconciled with the best interests of the child in the digital environment (ICO, 2020; Irish Data Protection Commission, 2020).

The selection of different EdTech is also influenced by the fact that some are offered to schools for free (with such software now being referred to as 'freeware') (CoE, 2020a, p. 3). In this context, schools may struggle to make informed risk-benefit decisions. Financial considerations may outweigh potential privacy and data protection issues, even though it has already been shown that many EdTech applications collect excessive amounts of children's personal data, including their device identifiers and location data, which, in many cases, may then be shared with third parties and advertisers (Kelly, 2019; Ng, 2020; UN General Assembly, 2021, para. 108; Wodinsky, 2021). As children's data increasingly 'fuel the business of the digital world' (UN General Assembly 2021, para. 90; see also Zuboff, 2019), the question as to whether these processes actually benefit children and their best interests remains contentious.

Meaningful implementation of data protection and children's rights law

Any digital service that processes children's personal data needs to comply with data protection law requirements. These include service providers' accountability, the requirement for

the lawful ground for data processing, such as meaningful consent or public interest task, compliance with the principles of fairness, transparency, purpose limitation, data minimisation, data protection by default and by design. The EU also requires specific protection of children's personal data and imposes stricter requirements for such processing (e.g., Recital 38 of the General Data Protection Regulation, GDPR).

Specific protection for children includes the requirement for data-processing information to be tailored particularly to them (Article 12); special vigilance regarding child profiling (Recital 71); a reinforced right to be forgotten (Recital 65); the child's right not to be subject to automated individual decision-making (Article 22) and the requirement for data protection impact assessments (DPIAs) when new technologies are used and the data processing is likely to result in a high risk to children's rights.

The principles of service providers' accountability, fairness and the requirement for specific protection of children's personal data (Articles 24 and 5(1)(a) GDPR, Recital 38) closely relate to the requirement to consider children's best interests as a primary consideration in any action affecting them. This stems from Article 3 of the UN Convention on the Rights of the Child (UNCRC). Whereas its provisions are primarily directed at States, the UN Committee on the Rights of the Child has emphasised that the Convention's provisions should also be respected by private businesses as the business sector affects children's rights in the provision of digital services (UN Committee on the Rights of the Child, 2013, 2021).

A precautionary approach

Despite the promises of personalisation, enhanced learning and improved education results, the actual positive impact of EdTech on children's learning remains unclear (Defend Digital Me, 2020, p. 45; Livingstone et al., 2021). Recent research indicates a clear need for evidence-based and child-centric risk assessment of technologies used by children, including EdTech (UN General Assembly, 2021, para. 82). Therefore, a crucial point in this contribution relates to the fact that many – if not most – of the anticipated benefits that EdTech may bring are not yet proven, and may instead lead to negative consequences for children's rights and future lives.

Generally, the precautionary principle 'compels society to act cautiously if there are certain – but not necessarily absolute – scientific indications of a potential danger and if not acting upon these indications could inflict harm', and it has 'traditionally been accepted that it is justified to err on the side of caution when it comes to the protection of vulnerable beings against potential harm' (Lievens, 2010, pp. 38, 42; 2021). This principle has been endorsed by both the CoE and UN (CoE, 2018; UN Committee on the Rights of the Child, 2021). The CoE specifically noted the need for a 'precautionary approach and a strengthened protection towards sensitive, special categories of data, including genetic and biometric data, and ethnic origin, or relating to sexual orientation, or offences, recognising children's additional vulnerability' (CoE, 2020a, p. 8). It also explicitly relied on the precautionary principle with regard to processing children's biometric data in the educational setting (CoE, 2020a, p. 20).†

It is currently very difficult to assess and predict the impact that extensive processing of children's (sensitive) data, profiling, personalisation, learning analytics and behavioural monitoring programmes developed by commercial actors will have on children('s rights) in the long term. Aside from a potential substantial impact on children's rights to privacy and data protection, there may be direct or collateral impact on their rights to development, identity, non-discrimination, freedom of thought, expression and association, as well as the right to protection from commercial exploitation. Bearing in mind that doubts exist as to whether data-processing practices in the educational environment are in some ways harmful to children, it would be in line with the best interests of the child to conduct fundamental, empirical and longitudinal evidence-based research on the matter first (Lievens, 2020, 2021).

The basic idea underpinning the precautionary principle relates to the adoption of risk mitigation measures in situations of inconclusive or incomplete evidence in terms of risks

† Specifically, '… processing characteristics about voice, eye movement, and gait; social emotional and mental health, and mood; and reactions to neurostimulation, for the purposes of influencing or monitoring a child's behaviour should be done on the basis of a precautionary principle and treated as biometric data' (CoE, 2020a, p. 20).

(Gellert, 2016). One such practical measure is a data protection impact assessment (DPIA), which is in certain cases required by the GDPR (Article 35). Many national DPAs have classified the processing of children's personal data for certain purposes as high-risk activities, and some specifically refer to children's data processing in the educational environment as high risk (Milkaite, 2021). Whether specifically required or not, carrying out DPIAs provides an opportunity for EdTech providers to take children's rights and best interests into account when their educational data is processed.

In addition to concerns surrounding children's data and privacy, it is also crucial to acknowledge that 'it is not only the child's right to data protection that is affected when it comes to education and digital technologies and that the right to privacy and data protection are enabling rights to the protection of further rights of the child' (CoE, 2020a, p. 11). Consequently, the CoE has also noted service providers' responsibility to conduct DPIAs, and stressed that these 'should have regard for the specific impact on children's rights and should demonstrate that the outcomes of algorithmic applications are in the best interests of the child and ensure that a child's development is not unduly influenced in opaque ways' (CoE, 2020a, p. 19; UN Committee on the Rights of the Child, 2013, paras 77-81).

In order to evaluate and implement children's best interests meaningfully, DPIAs should draw from children's rights impact assessments (CRIAs) so that EdTech providers can assess and take various rights of the child into account when they consider processing children's education data.† In line with a children's rights-based perspective and best interests, both CRIAs and DPIAs should be undertaken every time EdTech processes children's data. In light of children's right to be heard, these assessments should also actively involve children, and draw from their opinions and views associated with the EdTech they may be using every day (CoE, 2020a, p. 16).

† Whereas the 'rationale for conducting CRIA was originally formulated for States as the primary duty-bearers in public-decision making, … the same rationale is now also being extended to businesses.' The same can be said about human rights impact assessments and human rights due diligence requirements. These tools were also initially addressed at states but are now also directed at industry actors (Mukherjee et al., 2021, pp. 6, 11, 12). See also: CoE (2020); UN Committee on the Rights of the Child (2013).

Recommendations for national DPAs

The CoE has made a powerful acknowledgement that children 'cannot see or understand how large their digital footprint has become or how far it travels to thousands of third parties across or beyond the education landscape, throughout their lifetime' (CoE, 2020a, p. 4). At the same time, it also stresses that 'children's agency is vital and they must be better informed of how their own personal data are collected and processed' (CoE, 2020a, p. 4). In this regard it is essential to note that a consensus exists that 'children [and their parents and caregivers] cannot be expected to understand a very complex online environment and to take on its responsibilities alone' (CoE, 2020a, p. 4). In my view, however, the issue of extensive processing of children's (sensitive) data in the educational environment cannot be fully and meaningfully addressed through calls for child empowerment, resilience and data literacy.†

To a large extent, the limit of child empowerment is rooted in a power imbalance between children, parents and caregivers, schools and service providers.‡ States as the primary duty bearer for realising children's rights have obligations to enable children, parents and schools to exercise their agency. Therefore, policymakers and national DPAs have the responsibility to 'develop evidence-based standards and guidance for schools and other bodies responsible for procuring and using educational technologies and materials to ensure these deliver proven educational benefits and uphold the full range of children's rights' (CoE, 2020a, p. 6). Most importantly, States are responsible for holding service providers to account.

The accountability of EdTech providers in terms of the existing data protection and children's rights law requirements

† This statement is made without prejudice to the requirements that 'States should ensure that easily accessible, meaningful, child-friendly and age-appropriate information about privacy tools, settings and remedies is made available to children. Children and/or their parents or carers or legal representatives should be informed by a data controller how their personal data is being processed. This should include information for instance on how data is collected, stored, used and disclosed, on their rights to access their data, to rectify or erase this data or object to its processing, and how to exercise their rights' (CoE, 2018, para. 33).

‡ The increase in the use of EdTech 'amplified existing power imbalances between education technology companies and children' (UN General Assembly, 2021, para. 106), and 'most children and parents do not have the capacity to challenge educational technology companies' privacy arrangements or to refuse to provide data, as education is compulsory' (UN General Assembly, 2021, para. 107)

should be better ensured and enforced by DPAs. In addition to adopting specific guidance, codes of (best) practice and certification schemes on children's educational data, they should also require that EdTech providers adopt, rely on and publicly disclose their CRIAs and DPIAs (CoE, 2020a, p. 16; 2020b, para. 5.3), which should be developed in direct cooperation and consultation with children. Any such guidance and codes provided by DPAs should also be regularly reviewed in line with the rapidly evolving digital developments in the EdTech sector, and also be based on consultations with children.

Generally, it appears that 'there is a mindset of collecting [all possible data] now and thinking about what to do with it later' (Livingstone et al., 2021, p. 8). Such an attitude to children's education data is contrary to the principles of data minimisation and purpose limitation, as well as the best interests of the child and their right to privacy, as the processing of data must not involve more data than necessary to achieve the legitimate purpose for which it is collected. In this context, and in line with the precautionary principle, policymakers and DPAs should 'require the refusal of certain systems when their deployment leads to high risks of irreversible damage or when, due to their opacity, human control and oversight become impractical' (CoE, 2020b, p. 6). Indeed, national DPAs should be eager to enforce the existing data protection law requirements and consider the potential of imposing certain limits on children's education data processing.

The CoE has proposed a number of such limitations - for instance, that biometric data should not be routinely processed in educational settings, and that children's educational data 'should not be processed to serve or target behavioural advertisements' (CoE, 2020a, paras 7.7.1, 8.3.7). Both the CoE and the UN Committee on the Rights of the Child advocate for the prohibition of profiling with regard to children, unless scientific evidence shows that this can be done in the best interests of children and that appropriate safeguards are provided (CoE, 2020a, para. 7.6.2; UN Committee on the Rights of the Child, 2021). The CoE also maintains that using children's education data for data analytics and product development

cannot be considered 'legitimate compatible use for further processing [of children's education data] that override a child's best interests or rights' (CoE, 2020a, para. 7.1.11). In the same vein, EdTech providers should not be allowed to 'give away children's personal data collected in the course of their education, for others to monetise, or reprocess it for the purposes of selling anonymised or de-identified data, for example to data brokers' (CoE, 2020a, para. 7.1.12). In line with the purpose limitation and data minimisation principles, as well as children's best interests, only the minimum necessary amount of identifying data should be retained at the time when children leave education (CoE, 2020a, para. 7.4.1).

Finally, aside from implementing these and other recommendations for processing children's education data, national DPAs should also ensure that the rights and values of the UN Convention on the Rights of the Child (concerning in particular non-discrimination, development and privacy) clearly underpin their policies and decisions as 'children do not lose their human rights by virtue of passing through the school gates' (UN Committee on the Rights of the Child, 2001, para. 8). We need to ensure that this is also the case when children use EdTech.

Alba, D. (2020). Facial recognition moves into a new front: Schools. *The New York Times*, 6 February

CoE (Council of Europe). (2018). *Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment*

CoE. (2020a). *Children's data protection in an education setting*. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal data, Convention 108. Guidelines

CoE. (2020b). *Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems*

defenddigitalme. (2020). *The state of data 2020*

Gellert, R. (2016). We have always managed risks in data protection law: Understanding the similarities and differences between the rights-based and the risk-based approaches to data protection. *European Data Protection Law Review*, 2(4), 481-492

Kelly, H. (2019). School apps track students from classroom to bathroom, and parents are struggling to keep up. *The Washington Post*, 29 October

IAPP (International Association of Privacy Professionals). (2019). CNIL bans high schools' facial-recognition programs. 29 October

ICO (Information Commissioner's Office). (2020). *Age appropriate design: A code of practice for online services (the final version)*. 2 September

Irish Data Protection Commission. (2020). *Children front and centre. Fundamentals for a child-oriented approach to data processing (the fundamentals)*. Draft version for public consultation

Leaton Gray, S. (2018). Biometrics in schools. In J. Deakin, E. Taylor, & A. Kupchik (Eds.), *The Palgrave international handbook of school discipline, surveillance, and social control* (pp. 405-424). Springer International Publishing

Lee, D. (2019). School's facial recognition checks lead to fine. BBC News, 27 August

Lievens, E. (2010). *Protecting children in the digital era: The use of alternative regulatory instruments*. Brill Nijhoff

Lievens, E. (2020). *The rights of the child in the digital environment: From empowerment to re-responsibilisation*. In Essay collection on Freedom, Security, Privacy and the Future of Childhood in the Digital World, 17 June. 5Rights Foundation

Lievens, E. (2021). Growing up with digital technologies: How the precautionary principle might contribute to addressing potential serious harm to children's rights. *Nordic Journal of Human Rights*, 39(2), 128-145

Livingstone, S., Atabey, A., & Pothong, K. (2021). *Addressing the problems and realising the benefits of processing children's education data. Report on an expert roundtable*. Digital Futures Commission, 5Rights Foundation

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, *19*(5), 780-794

Milkaite, I. (2021). *A children's rights perspective on privacy and data protection in the digital age: A critical and forward-looking analysis of the EU General Data Protection Regulation and its implementation with respect to children and youth*. Dissertation, Ghent University

Mukherjee, S., Pothong, K., & Livingstone, S. (2021). *Child rights impact assessment. A tool to realise children's rights in the digital environment*. 5Rights Foundation, Digital Futures Commission

Ng, A. (2020). Education apps are sending your location data and personal info to advertisers. CNET, 1 September

Steeves, V., Regan, P., & Shade, L. R. (2018). Digital surveillance in the networked classroom. In J. Deakin, E. Taylor, & A. Kupchik (Eds.), *The Palgrave international handbook of school discipline, surveillance, and social control* (pp. 445-466). Springer International Publishing

Taylor, E. (2013). Surveillance schools: A new era in education. In E. Taylor (Ed.), *Surveillance schools: Security, discipline and control in contemporary education* (pp. 15-39). Palgrave Macmillan UK

UN Committee on the Rights of the Child. (2001). *General Comment No. 1 (2001) Article 29 (1): The aims of education* (CRC/GC/2001/1a

UN Committee on the Rights of the Child. (2013). *General Comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights*

UN Committee on the Rights of the Child. (2021). *General Comment No. 25 (2021) on children's rights in relation to the digital environment*

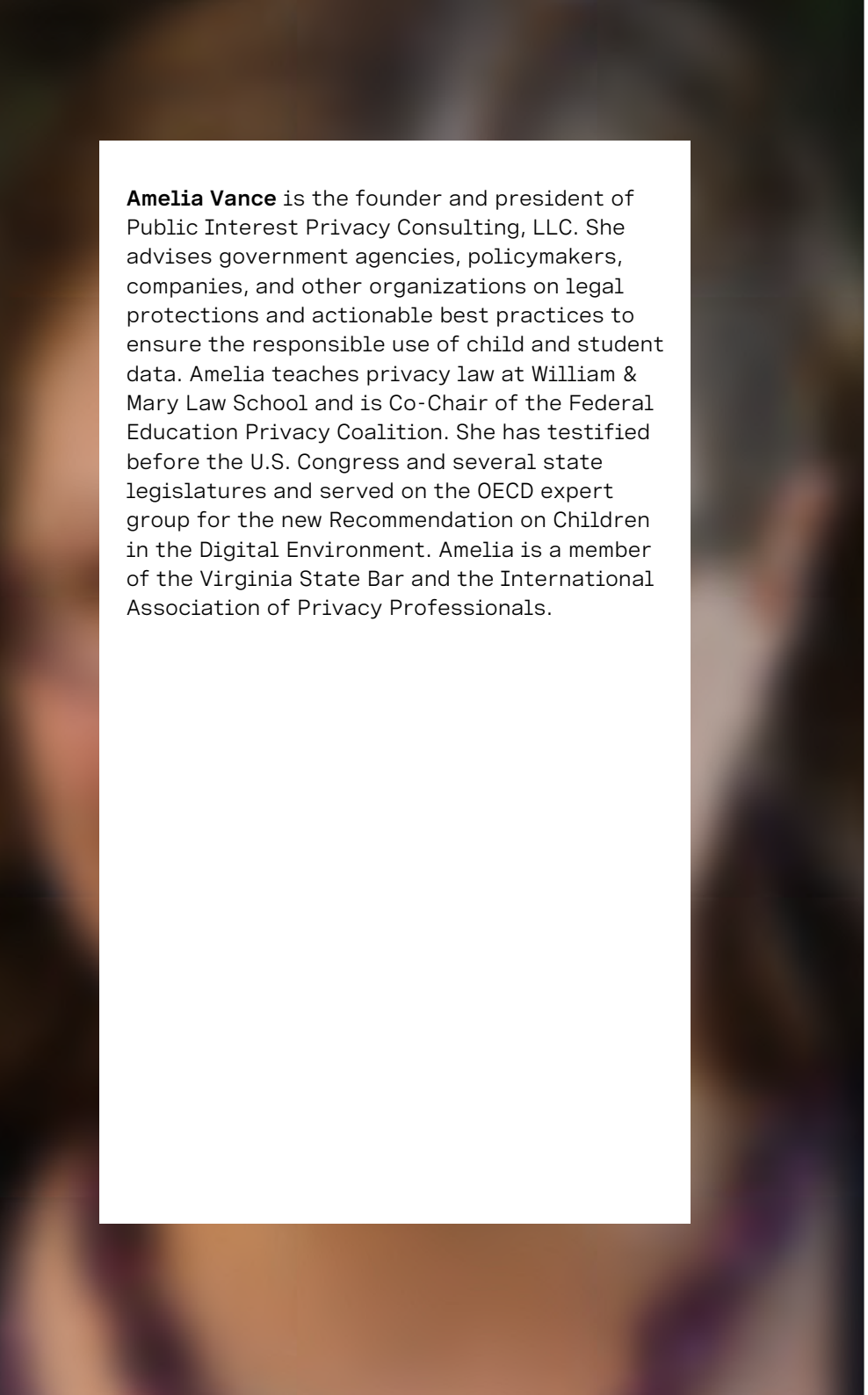
UN General Assembly. (2021) *Artificial intelligence and privacy, and children's privacy. Report of the Special Rapporteur on the right to privacy*. Joseph A. Cannataci. A/HRC/46/37. Human Rights Council

van der Hof, S., Lievens, E., Milkaite, I., Verdoodt, V., Hannema, T., & Liefaard, T. (2020). The child's right to protection against economic exploitation in the digital world. *The International Journal of Children's Rights*, 28(4), 833-859

Weale, S. (2021). ICO to step in after schools use facial recognition to speed up lunch queue. *The Guardian*, 18 October

Wodinsky, S. (2021). 60% of school apps are sharing your kids' data with third parties. *Gizmodo*, 5 April

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs



Amelia Vance is the founder and president of Public Interest Privacy Consulting, LLC. She advises government agencies, policymakers, companies, and other organizations on legal protections and actionable best practices to ensure the responsible use of child and student data. Amelia teaches privacy law at William & Mary Law School and is Co-Chair of the Federal Education Privacy Coalition. She has testified before the U.S. Congress and several state legislatures and served on the OECD expert group for the new Recommendation on Children in the Digital Environment. Amelia is a member of the Virginia State Bar and the International Association of Privacy Professionals.

Lessons learned from the Family Educational Rights and Privacy Act

Amelia Vance, Public Interest Privacy Consulting†

For decades, the Family Educational Rights and Privacy Act (FERPA) was one of the only laws in the world created to protect student privacy. Passed in 1974, it was one of the earliest privacy laws passed in the USA. This essay offers an overview of FERPA, including why it was developed, its provisions, its application to K-12‡ schools, its strengths and limitations, how it works in practice, and how it can be improved. The goal is to learn lessons from FERPA and provide best practices for other countries considering new student privacy protections.

The advent of FERPA: a necessary law amid government scandal

FERPA came shortly after the 1972 Watergate scandal in the USA, in which President Richard M. Nixon's administration was caught wiretapping and stealing documents from the Democratic National Committee's offices and subsequently attempted to cover it up. According to the bill's sponsor,

† The author thanks Stephen Hardy, Ashleigh Imus, Katherine Sledge and Elana Zeide for their assistance on this essay.

‡ The US equivalent of primary and secondary education in the UK.

Senator James Buckley, the Watergate scandal 'underscored the dangers of Government data gathering and the abuse of personal files, and ha[s] generated increased public demand for the control and elimination of such activities and abuses' (120 Cong. Rec. 14580, 1974). In addition, around this time, states and schools had begun to adopt computerised record systems (120 Cong. Rec. 13953-13954, 1974). Few school systems had policies on the use or disclosure of student records by school personnel (Wheeler, 1976, p. 49) or policies about access to records by third parties (p. 56). Student records were broadly shared with local, state and federal law enforcement, but parents (including caregivers) and students were more likely to be denied access to their records than any other stakeholder (Wheeler, 1976, p. 56). To combat these harms and abuses, Senator Buckley introduced FERPA to 'restore parental rights and to protect privacy'.

The lack of student data policies in US public schools indicated the need for a student privacy law, but the fraught political context and rapid passage of FERPA led to unintended consequences, necessitating several amendments to the law early on. Thus, although FERPA outlines key protections for students' data, gaps have always existed in its provisions.

FERPA provisions

As the law is written, FERPA guarantees parents or guardians and eligible students (generally defined as students over 18) access to their education record and the right to challenge information in those records as inaccurate or no longer relevant. It also intends to prevent unauthorised disclosure of education records without consent, with a few exceptions. The law requires certain safeguards in the absence of parental consent, such as the responsibility for schools acting as data controllers to oversee and have substantive control of their data processors, strict limitations on further processing of data beyond the original purpose, and legal contracting requirements.

Some student privacy advocates claim that FERPA's exceptions undercut the law's protections, because schools frequently use these exceptions to process student data instead of obtaining parental consent (e.g., Electronic Privacy Information Center, 2011, pp. 7, 13; Reidenberg et al., 2013, pp. 61). Advocates also argue that FERPA's exceptions do not have sufficient privacy governance requirements and protections (e.g., Electronic Privacy Information Center, 2011, pp. 7, 13; Reidenberg et al., 2013, pp. 61). These perceptions are not entirely accurate - FERPA exceptions to consent require documentation and safeguards. However, these requirements do not always translate well into practice. For example, the audit and evaluation exception - often described by advocates as being overly broad and not inclusive of necessary privacy protections and data-sharing restrictions (Electronic Privacy Information Center, 2011, pp. 10-13) - has incredibly detailed data governance documentation requirements, more than are required under any of FERPA's other exceptions (US Department of Education, 2015). In practice, FERPA is confusing, poorly understood and almost impossible for schools to follow, especially at a time when schools routinely share information with third party companies.

The next section explains the value of a law dedicated to student privacy and the complexities and problems that have arisen in FERPA's implementation.

FERPA: best practice and lessons learned

The value of a student privacy-specific law

The USA has traditionally approached privacy from a sectoral perspective, with laws that govern particular types of information. Legislators singled out student privacy for standalone legislation for three primary reasons:

1. **Students are required to attend school**. Parents are required to send their children to school and students are required to participate. Most school activities generate data. Because parents and students usually have no choice about having their sensitive data processed by schools, additional privacy protections are appropriate.

2. **Data collected by schools is generally about and from children**. Children are uniquely vulnerable to privacy harms and need additional protections.

3. **Data processing is an integral part of school responsibilities**. FERPA's exceptions exist because schools cannot operate without processing a significant amount of data about and from their students, so parental consent or opt-out is often not feasible. For example, obtaining parental consent every time teachers record attendance is unnecessary and cumbersome. To effectively educate students, schools must evaluate and track what students know and 'how quickly [they] are able to grasp new ideas or acquire new skills' (Wheeler, 1976, p. 29). It is also vital for schools to track student allergies, grades, test scores, parental contact information and custody status, as well as other data to audit whether students and families from marginalised communities are treated equitably to identify areas for improvement. For example, the US Department of Education (2016) collected and analysed such data and found that students of colour and students with disabilities endure higher rates of discipline in public schools compared to their white counterparts. The collection of this data has led to greater awareness and new initiatives to correct disproportionate rates of punishment (e.g., Amos & Manley, 2019).

Having a standalone federal student privacy law raises awareness of the sensitive nature of student information. It codifies the fundamental rights that parents and students should have when they cannot consent to data processing, although those fundamental rights may not adequately protect student privacy. Because the USA does not have an underlying data protection law, data not covered by FERPA may lack any legal protections. For example, when data is independently collected in schools by a law enforcement officer and not shared with school staff, that data is generally not covered by FERPA and could be broadly shared and not subject to access requests and other rights (US Department of Education, 2019, pp. 14-15).

Problems with reactive laws

As noted, however, FERPA also arose in part as a reaction to the Watergate scandal and to growing public concern about schools' unregulated collection of student data. For example, one widely shared magazine article, 'How secret school records can hurt your child', described how a black father discovered 'five pages of notes about his and his wife's "political activity"' in his child's record, and another case where parents were told their child would not be able to attend graduation ceremonies because her record showed she was a 'bad citizen' and were then refused access to the record explaining why (Divoky, 1974, as cited in 120 Cong. Rec. 13953-13954, 1974). Reacting to these concerns, legislators rapidly debated and passed the law, despite concerns raised during the short debate about the potential unintended consequences of ambiguous language in the law (120 Cong. Rec. 14579-14597, 1974).

For example, Senator Alan Cranston, of California, described the law's language as 'breathtaking in its sweeping generalities', arguing that the law 'could undermine attendance laws by allowing parents to refuse to have their child attend a class' whose content parents found objectionable (120 Cong. Rec. 14595, 1974). Other legislators pointed to the bill's 'strict limitations on sharing personal data, such as requiring a court order prior to sharing student information with law enforcement, and confusing regarding disclosing information to postsecondary institutions for financial aid' (Vance & Waughn, 2019, p. 523). Most of these and other concerns remained unaddressed before the law's passage (p. 524).

Thus, partly as a result of this context, framing and legislative process, FERPA is primarily a records management law, and not necessarily a privacy or data protection law; it was reactive and too focused on parental rights and consent as the primary mechanism for disclosure.

Consent as the cornerstone of FERPA

Consent is a key element of FERPA, the primary way that information can be disclosed. However, nearly 50 years after the law's passage, it is clear that parental consent as it currently exists is inadequate and ill suited to protect education data. Even when student information is disclosed

with parental consent, there are still serious privacy implications because consent removes all FERPA protections, including requirements for the use, minimisation, and sharing of data. Schools need to process data, and it is valuable, and sometimes essential, for them to use technology to do so. Most parents do not have time to investigate the privacy policies and practices of every educational technology (EdTech) product used by their children in school. When parents receive a consent form from the school to use EdTech, they likely lack the time and expertise to understand the rights and protections they are signing away, and they may assume that the school has already vetted the product. FERPA would be improved if there were underlying, unwaivable protections, such as those in the General Data Protection Regulation (GDPR) - the sale of student data and targeted advertising to children should generally not be waivable via consent, for example.

Confusion and misinterpretation of FERPA

Unfortunately, despite numerous minor amendments to FERPA, there continues to be confusion about when schools can disclose information without consent. FERPA is a complicated law, and the answer to most questions about FERPA is, 'it depends'. For example, data collected by law enforcement in school is sometimes unprotected under FERPA and sometimes absolutely protected, depending on who is collecting the data, the capacity they are acting in, how the data is collected and who it is shared with (US Department of Education, 2019, pp. 14-15). With that many factors to analyse, it is unsurprising that many public school districts do not know about, misunderstand or fail to adhere to their privacy obligations, in part because they cannot afford legal counsel with expertise in privacy law. Due to the general lack of legal requirements for data collection in the USA, many companies in the education market are also unaware of, or misunderstand, student privacy requirements.

Moreover, judicial interpretations of FERPA over the years have further muddied the waters. For example, in 2002, the US Supreme Court found that peer grading was allowed under FERPA, stating that a homework assignment was not part 'education record' until it was turned in to the teacher to grade

(*Owasso Independent School District No. I-011 v. Falvo*). That decision was more practical than strictly adhering to the law: no one wanted to ban peer grading, and a school should not be expected to protect data before it is in their control, such as a paper on a student's computer. However, in many modern applications there is no clear distinction between in-progress and completed assignments, since student work is often performed in cloud-based software owned and accessible by the school at any point during the process. This Supreme Court ruling creates confusion about whether these in-progress assignments are protected by FERPA. This decision also allowed for a level of plausible deniability that student information was protected by FERPA until it was provided to the school.

Direct governance of third parties

As originally passed, FERPA foresaw the growth of digital records and, to an extent, regulated data sharing with third parties. However, the original drafters did not anticipate the extent to which private companies handle student data from schools. As EdTech use grew in the USA, there was confusion and ambiguity about which data was protected under FERPA, and whether there was any direct liability for companies mishandling it.

Under FERPA's school official exception, schools can share information with companies that:

- Do something that a school would otherwise use employees to do
- Are under schools' 'direct control'
- Do not use student information for additional purposes or share it further (34 CFR 99.31(a)(1)(i)(B)).

However, legal ambiguity and practical implementation challenges keep these FERPA protections from adequately protecting student information. For example, prior to 2014, many companies placed the onus of FERPA compliance on schools, despite a little-known potential punishment in FERPA: the US Department of Education can impose a five-year ban on third parties that violate the law (34 CFR § 99.67). However,

few, if any, companies were aware of this potential penalty (especially since the penalty has never been imposed). State policymakers found passing all legal obligations on to schools to be unacceptable, and began to pass laws with specific requirements and restrictions that companies must adhere to. These laws generally prohibited targeted advertising to students and the creation of student profiles for non-educational purposes; an acknowledgement that the school is the sole data controller (to use GDPR terminology); and limits on redisclosure of student data. These laws helped to lessen the power discrepancy between schools and companies, creating better privacy protections overall, and made it clear that companies must also be proactive regarding student privacy responsibilities.

Regulation of school use

FERPA limits how schools can use and share data internally without consent in daily educational encounters: the school must 'use reasonable methods to ensure that school officials obtain access to only those education records in which they have legitimate educational interests' (34 CFR 99.31(a)); therefore, FERPA requires some level of access control. Schools must also explain, in an annual notice to parents, how they define 'legitimate educational interests'. However, the definition adopted by most schools is a catch-all that does not limit school discretion (Zeide, 2017, p. 515).

As privacy scholar Elana Zeide discusses, this limitation may not sufficiently serve the privacy and best interests of students since 'education purpose limitations equate educational functions with acceptable use' (2017, p. 515). Institutional interests could differ from student interests when it comes to the amount of information collected and retained in the first place. Schools - and the EdTech companies they partner with - can sometimes make better and more informed decisions when they have as much data as possible about a student's educational and non-educational experiences; assuming that all parties objectively act with a student's best interests in mind, having as much information as possible can allow more accurate and informed tailoring of curricular material, counselling and mentoring of students, and the overall wellbeing of students. But children may also become less willing to learn if they know that everything they do will be watched and retained (Zeide, 2017, p. 517).

Over-collection of data can also cut 'against the norms that early mistakes should not foreclose future opportunities', and as children mature, could limit their future opportunities (Zeide, 2017, p. 520). What school personnel may consider the best interests of the student body could actually be contrary to the best interests of an individual student: 'the wellbeing of the majority of students - the students who use the fewest resources and need the fewest interventions - may be prioritised over students with disabilities and students of lower socioeconomic status, who may need more resources and attention' (Selinger & Vance, 2020, pp. 42-43). School personnel may have biases related to students from marginalised populations or based on inaccurate beliefs about what a student's prior behaviour means about their future. Student privacy laws, such as FERPA, must include better guardrails to protect students when institutional interests may conflict with students' best interests.

Policymakers should also consider whether there is some collection or use of data that schools should not undertake at all because of the potential for privacy risks, inequities or abuse. For example, there are significant concerns in the USA regarding monitoring student use of the internet or activity on school devices for self-harm. While preventing self-harm is vital, the efficacy of these services is questionable; these services 'could exacerbate feelings of stigma and shame and could ultimately make students less likely to ask for help' and 'undercut the trust of students not only in their school generally but in their teacher [and] counselors' (Keierleber, 2021). The surveillance has also been criticised as it could prime students 'to accept surveillance as an inevitable reality', causing them to give up 'the ability to explore new ideas and learn from mistakes' (Keierleber, 2021).

In some cases, identification of a student's mental health crisis - whether accurate or not - can cause more harm than help (see The Southern Poverty Law Center, 2021; Vance et al., 2021). When extremely sensitive data is collected and used for purposes that could have a significant impact on a child's

life, wellbeing and future opportunities, additional privacy protections and restrictions should be incorporated into law to mitigate potential harms.

Training requirements are essential

Why should schools adhere to privacy protections in the first place? Many US schools reflect an overarching lack of understanding about why student data requires significant protection, with the exception of obviously sensitive information like medical data, special education services or a parent's financial data. Without proper training on the value of student data protection, school personnel cannot make informed decisions about data processing or the adoption of EdTech tools.

This is particularly important when teachers adopt new EdTech. After all, most apps may seem to only collect a student's name and email and their activity in the app, so teachers may ask why this is a privacy problem. If the only answer that school districts can provide is 'this is legally required', many teachers will choose to do what they think is best for their students' learning regardless of legal privacy protections. In many cases, this risk analysis may be accurate. However, teachers may not be aware of several factors that raise the risk level; for example, companies may sell data about student activity, which could lead to a student who is bad at maths receiving an ad in the future encouraging them to take out an exploitative loan.

Even seemingly innocuous information poses a threat to students: for example, releasing the name of a student who is involved in a domestic violence situation could alert an abusive parent to the student's location. And, of course, the information collected by EdTech can be far broader than just a name, email, and app activity; teachers might connect the app with the school's electronic student record system, and the app could then receive some or all of the record - including sensitive data, such as disabilities and disciplinary records - even though the app does not require that information.

Unfortunately, FERPA does not include a training requirement, and the federal and state governments provide little-to-no voluntary training. A survey from the advocacy group Common Sense Media found that 'only 25 percent of teachers who received professional development to support their use of educational technology were trained to understand student data privacy requirements and strategies' (Mandinach & Cotto, 2021, summarising Common Sense Media, 2019). Not only do most school personnel not know about the legal requirements; they also do not know why they should care about privacy protection in the first place. Similarly, companies may also not understand privacy risks and how the information they collect could be harmful. This lack of understanding leads them to deprioritise privacy, especially when they believe that the service they provide will be a net good in helping students.

Enforcement issues related to transparency

FERPA is often considered toothless. The US Department of Education has never imposed the law's ultimate penalty on any school - complete removal of all federal funds (of course, no one wants to take away education funds used to serve children). This is largely because FERPA requires the Department to work with schools before withdrawing funds, and schools understandably comply with the agency's conditions.

The US Department of Education should be more transparent about its FERPA enforcement since most FERPA complaints are not resolved publicly. This lack of transparency creates the impression that FERPA does not adequately protect student privacy. Publishing aggregate information, such as the number of in-process complaints, how long it takes to process them and which issues frequently arise, would promote public trust.

Conclusion

Other jurisdictions crafting their own student privacy laws can find value in considering lessons learned from FERPA. A standalone student privacy law allows policymakers to consider education's unique facets, such as parents' and students' lack of ability to consent. However, unlike FERPA, new laws should be created proactively, with thorough consideration of the relevant privacy problems and consultation of diverse stakeholders such as educators, parents and students

themselves. A new student privacy law should be clearly written and mitigate privacy harms without unduly burdening school systems. It should include data minimisation, training for educators and transparent enforcement methods that put the onus of protecting student data on third parties, as well as work in conjunction with more general privacy laws.

In addition to looking at FERPA, international policymakers considering these laws could benefit from examining student privacy laws passed in US states over the past decade (Vance, 2016). These were largely passed due to perceived FERPA weaknesses - for example, by adding direct regulation of third parties that receive student information - and could therefore serve as a better template for new student privacy laws.

Even with these state laws supplementing FERPA, the USA's student privacy protections still need improvement. Schools should consider the best interests of each student, and weigh the risk of certain data processing against potential benefits. Legal requirements should prevent over-surveillance and data hoarding. Enforcement processes need to be more robust and transparent.

Regulating student privacy is difficult. There are great benefits and needs met by processing student data, but also many risks. A nuanced approach, built with feedback from stakeholders, is necessary to ensure effective student privacy protections.

120 Cong. Rec. 9371 (1974).

120 Cong. Rec. 13953-13954 (1974).

120 Cong. Rec 14579-14597 (1974).

34 CFR 99.31(a).

34 CFR § 99.67.

Amos, L., & Manley, M. (2019) Using data to identify and address inequities in school discipline. Mathematica Blog, 22 October

Common Sense Media. (2019). *The Common Sense census: Inside the 21st-century classroom* (pp. 45-46)

Divoky, D. (1974). How secret school records can hurt your child. Parade, 31 March. As cited in 120 Cong. Rec. 13953-13954 (1974)

Electronic Privacy Information Center. (2011). Comments to the Department of Education, 'Notice of Proposed Rulemaking', RIN 1880-AA86. 23 May

Keierleber, M. (2021). An inside look at the spy tech that followed kids home for remote learning - and now won't leave. The 74, 14 September

Mandinach, E. B., & Cotto, J. (2021). The case for including data privacy and data ethics in educator preparation programs. *Future of Privacy Forum*, 5 October

Reidenberg, J., Russell, N. C., Kovnot, J., Norton, T. B., Cloutier, R., & Alvarado, D. (2013). *Privacy and cloud computing in public schools*. Fordham Center on Law and Information Policy

Selinger, E., & Vance, A. (2020). Teaching privacy and ethical guardrails for the AI imperative in education. *Future EDge, NSW Department of Education, 3*, 30-53

Southern Poverty Law Center, The (2021). *Costly and cruel: Thousands of Florida children suffer the harm and indignity of involuntary, and often illegal, commitment to psychiatric facilities*

Vance, A., & Waughn, C. (2019). Student privacy's history of unintended consequences. *Seton Hall Legislative Journal, 44*(3), 515-557

Vance, A., Collins, S., Park, J., Reddy, A., & Sharifi, Y. (2021). The privacy and equity implications of using self-harm monitoring technologies. Future of Privacy Forum, 27 September

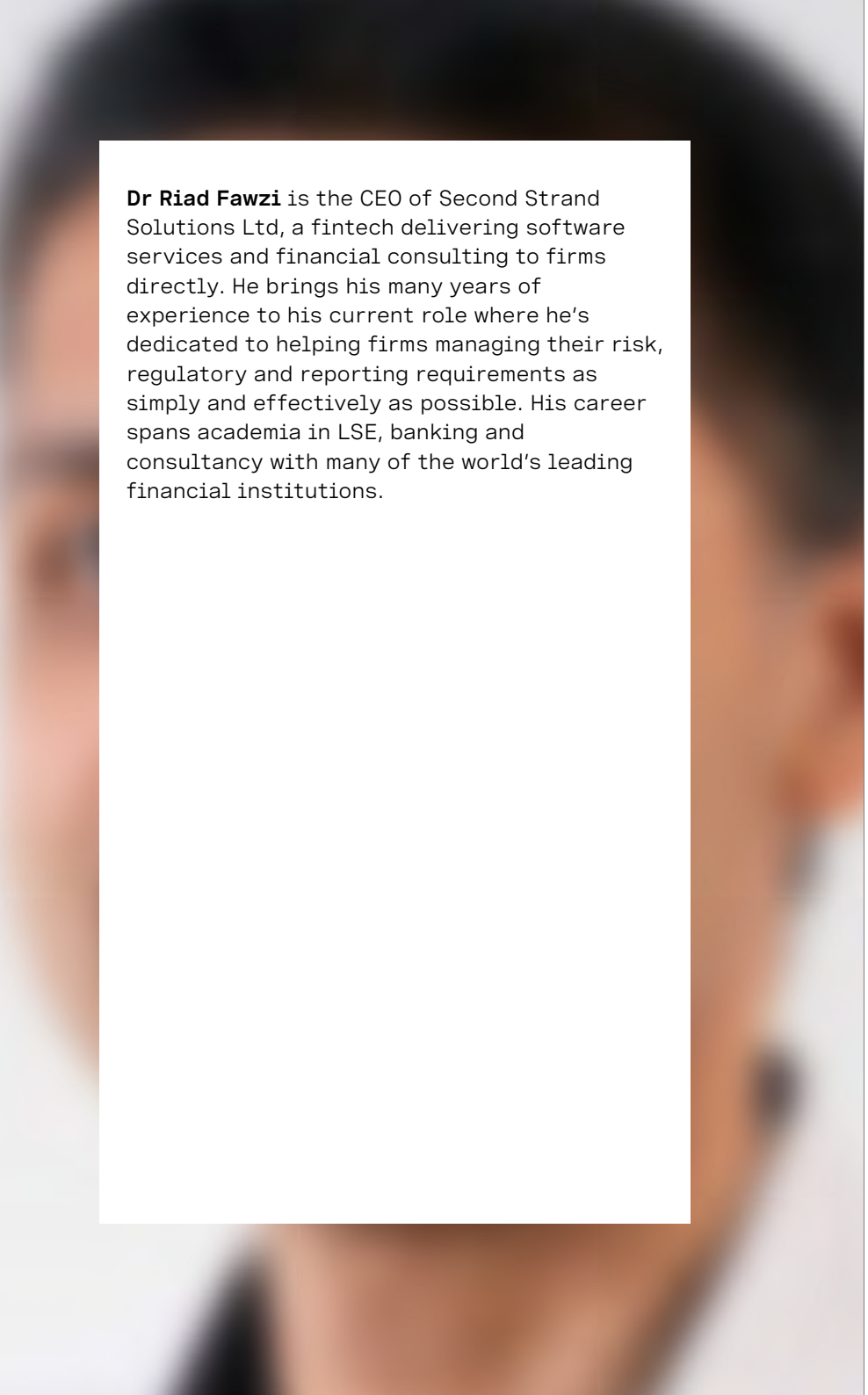
US (United States) Department of Education. (2015). *The Family Educational Rights and Privacy Act: Guidance for reasonable methods and written agreements*

US Department of Education. (2016). *School climate and discipline: Know the data*

US Department of Education. (2019). *School resource officers, school law enforcement units, and the Family Educational Rights and Privacy Act (FERPA)*

Wheeler, S. (1976). *On record: Files and dossiers in American life*. Transaction Books.

Zeide, E. (2017). The limits of education purpose limitations. *University of Miami Law Review, 71*(2), 494-515



Dr Riad Fawzi is the CEO of Second Strand Solutions Ltd, a fintech delivering software services and financial consulting to firms directly. He brings his many years of experience to his current role where he's dedicated to helping firms managing their risk, regulatory and reporting requirements as simply and effectively as possible. His career spans academia in LSE, banking and consultancy with many of the world's leading financial institutions.

Building trust in EdTech: Lessons from FinTech

Riad Fawzi, Second Strand Solutions Ltd

Financial technology (FinTech) in the UK has now matured to a level where it is seen both as a credible threat to established financial services and also to the future of financial services. It offers lower rates and fees, a wider range of functionality and features, more convenience and more interesting user experiences. For investors, it offers potential material returns for less initial outlay. For regulators, it has the potential to provide more financial services to more people. And for established financial services firms, it offers new markets and a new set of tools to attract and manage consumers (Browning & Evans, 2021). Total investment in British FinTech jumped more than 217% to US$11.6 billion in 2021, second only to US firms globally, which saw US$46 billion of investment (Innovate Finance, 2022); US$24.3 billion was invested across Europe, with the UK attracting nearly half of total investment.

FinTech uses technology to improve financial services delivery through lower barriers to entry, lower running costs, increased access to more consumers and increased convenience for both the supplier and consumer. It encompasses technology-driven and technology-enabled businesses, from established financial services software vendors to emerging digital service providers, and also discrete

FinTech sectors such as Insurtech, RegTech and PayTech, which focus on specific areas of the financial services value chain. Although it originated back in the 19th century, ongoing developments, such as electronic fund transfers through FedEx and Morse code, credit cards, ATMs, Telex, SWIFT, Nasdaq and, of course, the personal computer (PC), have demonstrated the close relationship between technological innovations and financial services.

And yet, for all that growth and expenditure, trust in financial services in the UK is still generally low (47% of consumers trust any financial services, 45% trust only traditional financial services and 28% trust only FinTech; Edelman, 2021). Developing trust in digital industries is hindered by their reliance on complex technical innovations that have had negative media coverage, such as artificial intelligence (AI), machine learning, distributed ledger technology, blockchain, peer-to-peer platforms and crowd-funding (Browning & Evans, 2021). Less trust in an industry is ultimately the biggest growth obstacle. Having a better product, service, reach, cost, support, etc. are key contributing factors to success, but unless people trust the firm, they won't trust it can deliver these factors.

Educational technology (EdTech) and FinTech share common concepts. They both grew out of long-established industries and use modern technology to deliver their services with certain cost and convenience savings. They both offer fundamental and additional services. Both offer business-to-business (B2B) solutions across the incumbent value chain (e.g., Stripe in FinTech and Sprint-Education in EdTech), as well as business-to-consumer (B2C) offerings (e.g., business-to-adult consumers - Revolut in FinTech and Codeacademy in EdTech, and business-to-child consumers - GoHenry in FinTech and Kuato studios in EdTech). EdTech, like FinTech, however, has a trust problem (EdTech Impact, 2021).

There are three key approaches to understanding trust drivers in FinTech (and EdTech): (a) trust in the rules, regulations and supporting infrastructure underpinning the market; (b) trust in the protection of the tools and techniques to look after the consumer;† and (c) trust in the approach and focus of the firms delivering the services and products.

These three approaches, or layers, legitimise the industry and increase consumer trust. So what can be learned by EdTech from FinTech's experience? We will now explore some of the key lessons learned and see what value FinTech can bring to EdTech.

Improve trust in the rules

Industry support through government oversight and self-regulation has gone a long way to legitimising FinTech as trustworthy. Although there is no single UK or international regulatory body responsible for FinTech, there are multiple stakeholders and existing or proposed legal frameworks that provide a regulatory roadmap for the sector. These include the Information Commissioner's Office (ICO), Financial Services Compensation Scheme (FSCS), Financial Ombudsman Service, Prudential Regulation Authority (PRA)/Bank of England and Financial Conduct Authority (FCA). Within these there are now dedicated programmes for FinTech, such as the FCA's Innovation Hub and Regulatory Sandbox, Payments Systems Regulator (2015), Bank of England FinTech Accelerator (2016) and the Cryptoassets Taskforce (2018).

In addition, while some aspects of the FinTech environment are specific to that sector, other critical aspects - such as data and AI - are cross-cutting issues that are also relevant to other economic sectors, with their own rules and regulations (Insurtech, etc.). Regulators are also working to clarify which financial products and services fall within their remit (e.g., the FCA's 2020 *Perimeter report*) sets out several issues relating to its scope regarding oversight for FinTech, including cryptoassets and social media companies and online retailers providing financial services).

For self-regulation, the UK has bodies such as Innovate Finance and FinTech Alliance, as well as existing incumbent financial services bodies that promote their firms' interests (British Bankers' Association, UK Finance, etc.). The UK Government has also invested both in FinTech and in organisations working to develop start-ups and scale-ups, and

† We use 'consumer' here to mean the end user of the FinTech/EdTech product, whether these are schools, adults or even children.

has convened various stakeholder groups and developed international partnerships to promote the sector.† The UK Government has undertaken studies that legitimise FinTech, such as the Kalifa Review of UK FinTech (Kalifa, 2021), the UK *FinTech State of the nation* report (Department for International Trade, Innovate Finance, 2019), the Tech Nation report (Tech Nation, 2021) and the House of Commons report *FinTech: A guide to financial technology* (Evans & Browning, 2021).

As any sector becomes more prominent and complexity increases, so, too, does the level of scrutiny applied to it, and the societal and business impact of scandals and failures pertaining to it (Plaid.com, 2021). Eighty-nine per cent of UK adults lose trust in a company following an IT failure or technical problem, and 51% would consider switching banks following issues with their IT systems (ProBrand, 2020). What impacts consumers' trust in these circumstances, however, is the reaction of the government and self-regulating bodies to these scandals and losses. Since 2008, financial services have received their fair share of scandals and failures, although many of these currently relate to IT and cyber failures (enough to have their own government report; see House of Commons Treasury Committee, 2019).

Strong FinTech growth has resulted in some failures along the way, some of which relate to lack of funds (37coins), bad business models (Flowtab), management oversight/governance (N26) and fraud (Wirecard). We've also seen that the fast growth of FinTech along with market pressures has resulted in some (allegedly) questionable practices such as the recent investigations into money-laundering charges for Monzo and Revolut, and the FCA's interest in payment institutions with foreign ownership (Griffin, 2022).

What EdTech can learn from FinTech is that while having a mix of regulations, self-regulation and oversight does work to protect consumers to a large extent, it can become overly

† For example, British Business Bank programmes, Department for Digital, Culture, Media & Sport (DCMS) and HM Treasury funding for Tech Nation, Inclusive Economy Partnership (IEP), AI Sector Deal, 'FinTech bridges' with Hong Kong, South Korea, Singapore, China and Australia, as well as the UK-Africa FinTech Partnership, Insurtech Board, Asset Manager Authorisation Hub and government backing to the Investment Association to develop a FinTech accelerator for the asset management industry called Engine.

complex and difficult to manage. It is more important to have effective industry protection in EdTech (such as dedicated forums, industry bodies, etc.) given the more vulnerable nature of the end users and the different motivators between EdTech firms and end users (profit versus education), otherwise this could lead to scandals, failures and monopolistic and monopsonistic† behaviours impacting its development (Noula, 2021).

Regulation is typically not as agile as self-regulation, especially when mixing new regulations with amended old ones. Government oversight must be carefully balanced as to be a suitable deterrent for some firms without stifling others. It is only when all the parts are working together that we can expect to see a reasonable level of protection. Both FinTech and EdTech would then benefit from a chance to create a robust yet workable regulatory framework that includes self-regulation and considers the various industry nuances.

Improve trust in the protection of the tools and regulations

The approach and steps taken by firms in how they use consumer data and how they protect it from others is key in building trust in digital firms. This is especially true in a world with third-party interconnected systems and different systems by different developers working together, leading to compatibility problems. In 1973 Richard Serra and Carlota Fay Schoolman broadcast a short video titled 'Television delivers people', in which they paraphrase the quote '[if] you're not the consumer, you're the product'. The way a firm treats its consumers is reflected in the way consumers trust the firm. For digital firms especially, consumers care about how their information is managed. Where the processes or technology are complex (e.g., encryption levels, cloud-based computing, etc.) and consumer understanding is low, consumers will look to the firm and industry to protect their data.

In its annual report on data breaching, IBM found that for firms globally the average total cost of a data breach was US$4.24 million, of which US$1.59 million (38%) came from

† Monopolistic behaviour is when there is only one seller of the supply of or trade in a product or service.

lost business costs (IBM Security, 2021). Security, privacy and digital identity (SPD) are thus key concepts for improving trust in a service or product. Consumers have to trust that their information and identities are being safeguarded, and that they can entrust the firms with sensitive information. SPD concerns are also relevant in EdTech, where personal data is collected, stored and processed using multiple systems and third parties (26% of further education colleges are [cyber]attacked at least weekly - as are 6% of primary schools and 15% of secondary schools; DCMS).

So how can FinTech and EdTech legitimise themselves and allay consumers' fears? Consumers' confidence can be boosted by demonstrating good SPD practices, which includes considering SPD implications at every stage of developing a new product or service, being upfront about how they're processing consumer data (in line with the General Data Protection Regulation [GDPR] and other educational and financial services regulations on protecting consumer data requirements, for instance) and promoting the extra steps they're taking to keep personal information safe.

From a technical point of view firms should have robust security and privacy measures in place, including (but definitely not limited to) regular vulnerability scanning, scalable architecture, biometrics, passwords, one-time passcodes and more than one verification gateway based on different principles and technologies. However, SPD compliance is not static. Firms must have both the desire to protect consumers' information and the technical knowledge/infrastructure/support to cope with current and future attacks. They must keep innovating their technical defences and protection as attackers will do the same. But digital firms should also take a proactive response by investing in regular training for their staff -84% of UK data breaches are down to human error (Egress, 2021).

The increased pressure on FinTech and EdTech to deliver strong SPD measures, however, must be counterbalanced against increased inconvenience - not alienating consumers while simultaneously considering user control and consent. There is significant pressure on firms to compress the due diligence process into a short amount of time, leading to

significant resistance from financial institutions to the type and extent of diligence that regulators require. In fact, in 2022 the UK Government announced legislation to increase trust in digital identities by making firms obtain a Trustmark to show they could handle people's identity data in a safe and consistent way, and a new Office for Digital Identities and Attributes is to be established to oversee strong security and privacy standards for digital ID (DCMS, 2022). There have also been various European (Open Access Government, 2022) and private self-regulatory initiatives (Hancock, 2020). Linked to this is similar research into the 'digital divide', where consumers are unable to access financial services because they lack internet access. How FinTech is actively dealing with digital illiteracy, or consistent access to the internet, is an excellent lesson to pass on to EdTech where the most deprived children are at real risk of being disadvantaged (Saka et al., 2021; Vissenberg et al., 2022).

Building trust through good customer service

Following the 2008 global financial crisis, the housing crash and the subsequent international recession, regulators focused on restricting financial services activities with an increased emphasis on risk and regulatory compliance, which led, in turn, to an increase in technology spending. However, consumers also wanted better consumer experiences (i.e., more convenience, more support or a more entertaining consumer journey). Consumers change firms when they trust that the new firm can offer them a product or service that is cheaper, more convenient, gives them more choice or whose approach is better suited to their needs (PwC, 2021).

What has worked for FinTech companies is knowing who their consumers are and what drives them to use their product rather than traditional financial services. FinTech has learned to react quickly to consumer feedback and adjust its services accordingly. It epitomises agile development wherein it focuses on short sprints with specific goals along with regular consumer feedback so that consumers can quickly adapt to any changes (Munteanu & Dragos, 2021). By adding several support options like WhatsApp messaging or enabling 24/7 consumer service, FinTech increases its ability to interact

with consumers. Apple and Goldman Sachs offer a joint credit card in the USA that has comparable rates to other financial service providers, but is integrated into the Apple infrastructure, and so is more convenient and cheaper for many consumers (Apple.com, 2021). A better consumer experience thus has to be considered from the start to attract users and keep them interested.

As a corollary, FinTech, with lower operating costs and more focused business models, can take on more disadvantaged consumers (i.e., with lower credit scores) as they can have less conservative risk appetites (IMF, 2022), which means more people have access to financial services (which is a good thing).

So for EdTech the lesson is to build in good consumer service from the start. EdTech consumers include government agencies, schools, teachers and the end consumer (the child or adult being educated). Teachers are key, as they are both the source of the information as well as key users. Whereas FinTech has a high adoption rate, with 86% of UK consumers now using some sort of FinTech (Plaid.com, 2021), the uptake of EdTech is limited as many teachers lack the training and awareness of what it can offer (55% of teachers are unsatisfied with EdTech training; Brogan, 2021).

To improve the consumer experience for all stakeholders, especially teachers and students, EdTech firms have to make sure that their solution is aligned with consumers' expectations and test their application early and often. They should also ask for feedback from users on the ease of use and offerings of the service or product so they can make any significant changes sooner rather than later. Using AI, neural networks, etc., both FinTech and EdTech firms can also potentially satisfy their consumers' current needs and predict their future needs as well as improving their experience to build trust and develop loyalty.

Good consumer service can also mean having the right balance between digital and human interaction. Like a good game, design should funnel the user to where they need to go, raise awareness of key features and give users rewards to motivate them to use the service. The inclusion of voice assistants and chatbots can also add to the consumer's experience, through, for example, sending users notifications about bill payments that are due, warning them of a low balance, and also offering extensive proactive financial guidance. Building in human interaction, gamification and chatbots, FinTech has learned to develop trust through better consumer experiences (Hill & Brunvand, 2020).

If consumers and regulators cannot trust that the firm delivering the product or service will continue to function, how can we expect them to trust the firm at all? Recent research into FinTech business models has shown that organisations need to assess their FinTech initiatives to realise gains and business value, focusing on success factors and the unique value-added from being FinTechs (Jinasena et al., 2020). Revolut is the UK's biggest FinTech. It was founded in 2015, and as of June 2021, had a US$33 billion valuation, 15 million consumers, 2200 employees and revenue of £222 million in 2020, up from £166 million previously, although it had losses of £207.9 million in 2020, higher than the £107.7 million it lost in 2019. In June 2021 it had its next round of funding, receiving US$800 million. Monzo, another UK FinTech founded in 2015, is currently valued at US$4.5 billion, with 5 million consumers and 2100 staff. As of February 2021, its losses increased to £130 million, up from £114 million previously. Monzo also secured another round of funding.

The situation is similar internationally. On the other hand, some neobanks are profitable, such as Starling Bank, as well as many B2Bs, such as PayPal and Stripe. FinTech success is rarely measured in profitability, however, as it is seen as a growth market. FinTechs generate funding through measures such as bootstrapping, venture capitalists, angel investors, hedge funds, crowdfunding, or through more traditional sources such as bank loans. Yet for firms that purport to replace incumbent financial services that have been around for hundreds of years, many are not (yet) a success in terms of making enough profit to remain a going concern.

There are several reasons for FinTech failures, including having a bad business model, underfunding, compliance issues, technical limitations, poor management, wrong partnerships, economic downturn and competition (Boyd, 2021) (although these are also the reasons any firm will fail). The questions for FinTech and EdTech (and anyone else) are: how will they make

money and how will they ensure they continue to exist? EdTech firms must understand who the target audience is and what their needs are, and how they will deliver these needs - while making a return. If they need help in considering these questions, they may need suitable partners.

Financial services comprise many interwoven strands, covering payments, trade, fundraising, asset management, insurance, etc. There are also dependent strands such as risk management, compliance, audit and governance, to name a few. Given the level of complexity involved in financial services, FinTechs cannot progress without solid partnerships. Having the right support, whether in terms of technical advice, business or organisational support, will help ensure FinTechs can develop a sustainable business model. Having the right partnerships in place will also reduce the risk that businesses might become dependent on a single provider of critical infrastructure.

EdTech already has some strong partnerships in place, such as the School Rebuilding Programme (DfE, 2022) and the UK Government's Connect the Classroom (DfE, 2021). A key message from the Department for Education's digital strategy (Harrison, 2019) is that 'We need the future of technology in education to be driven by collaboration.' The focus should be on having partnerships to deliver and enhance the educational offerings (LEANLAB Education Editorial Team, 2021), not having education being used as a medium to sell the firm's products (Kingsley, 2020; Meyer et al., 2019). The educational agenda must be aligned to the interests of consumers, the government and firms (assuming, of course, that the government's wants are aligned with the consumers' and firms' interests).

Conclusion

So can EdTech learn from FinTech? Technology can make education more interactive and enhance the learning experience through the use of digital products. But, as we've seen here, there are many lessons EdTech can learn from FinTech, such as focusing on the consumer and their needs, not on the needs of the firm or the technology. EdTech can't succeed on its own; it must continue to develop relationships and support from stakeholders (teachers, technicians, regulators and consumers). It must work with stakeholders, especially governments and industry bodies, to deliver a strong governance and oversight model to reassure consumers. EdTech is not a separate industry from education; it is another service offering. And finally, remember that the key value that EdTech firms are selling is trust, not more convenience, better prices or more access. Once consumers trust that EdTech can provide their educational needs, the rest will follow.

Apple.com. (2021, 09 21). *Apple Card customers ranked Goldman Sachs and Apple Card No. 1 in customer satisfaction among Midsized Credit Card segment, according to J.D. Power*

Boyd, A. (2021, May). *10 fintechs that failed (and why)*

Brogan, T. (2021, December 8). *55% of teachers unsatisfied with edtech training, survey says*

Browning, J. E. (2021). Fintech: a guide to financial technology. *House of Commons Briefing Paper, 1950* (26 April)

Browning, S., & Evans, J. (2021, April). *Fintech: a guide to Financial Technology*. Briefing Paper, House of Commons

Dakshitha N. Jinasena, K. S. (2020). Success and Failure Retrospectives of FinTech Projects: A Case Study Approach. *Information Systems Frontiers*

David R. Hill, S. B. (2020). Gamification. In R. K. Anne Ottenbreit-Leftwich, *The K-12 Educational Technology Handbook* (p. Section 2.7). EdTech Books.

Department for Digital, C. M. (2022, March 10). *New legislation set to make digital identities more trustworthy and secure*

Department for Digital, Culture, Media and Sport. (2021). *Cyber Security Breaches Survey 2021 Education Annex*. Department for Digital, Culture, Media & Sport.

Department for International Trade, Innovate Finance. (2019). *UK FinTech State of the Nation*. Department for International Trade.

Department for International Trade, Lord Grimstone of Boscobel Kt. (2020, December 9). *Press release: New partnership programme to help financial institutions digitise*

Department of Education. (2021). *Connect the Classroom*

Department of Education. (2022, April 1). *School Rebuilding Programme*

Edelman. (2021). *Trust Barometer 2021*

Egress. (2021). *Insider Data Breach Survey*. Egress.com

Finance, I. (2022). *uk-sees-record-year-for-fintech-investment-reaching-11-6bn-in-2021-2*

Financial Conduct Authority. (2020/21). *Perimeter Report*. Financial Conduct Authority

Griffin, D. (2022, January 4). *London's Fintech Boom Opens the Door for Dirty Money*

Hancock, A. (2020, August 31). *Digital Identification Must Be Designed for Privacy and Equity*

Harrison, B. (2019, May 08). *The ed-tech strategy: A great leap backwards?*

House of Commons Treasury Committee. (2019). *IT failures in the Financial Servcies Sector*. House of Commons

IBM Security. (2021). *Cost of a Data Breach Report*. IBM

International Monetary Fund. (2022). *Global Financial Stability Report*. International Monetary Fund

Jamie Evans, S. B. (2021). Fintech: a guide to financial technology. *House of Commons Briefing Paper*

Joyce Vissenberg, L. d. (2022). Digital Literacy and Online Resilience as Facilitators of Young People's Well-Being? *European Psychologist, 27*(2)

Kingsley, A. (2020, May 11). *Ed-tech: Making industry collaboration work for your school*

LEANLAB Education Editorial Team. (2021, December 13). *EdTech As Augmentation: Building A Strong Partnership Framework Between Schools and EdTech Entrepreneurs*

Meyer, M., Adkins, V. M., Yuan, N. M., Weeks, H. M., Chang, Y.-J. P., & Radesky, J. M. (2019). Advertising in Young Children's Apps: A Content Analysis. *Journal of Developmental & Behavioral Pediatrics, 40*(1), 32-39

Noula, I. (2021, September 22). EdTech is at a crossroads: *Why evidence is needed, standards must be set, and regulation should be implemented*

Open Access Government. (2022, February 23). *A new world of data privacy with EU digital identities*

Orkun Saka, B. E. (2021). *Epidemic Exposure, Fintech Adoption*. European Bank for Reconstruction and Development.

Plaid.com. (2021). *The Fintech Effect: Fintech's Mass Adoption Moment*. Plaid.com

ProBrand. (2020, February 18). *IT failures cause banks credibility to plummet*

PwC. (2021). *Small businesses are hurting. Our survey shows how banks can help*

Series, S. I. (2021). *Why Trust Matters in Sales* (Autumn 2021)

Sprint Planning. (n.d.)

Tech Nation. (2021). *The future UK tech built*. Tech Nation

Treasury, H. (2017). *Imposition of Monetary Penalty – TransferGo Limited*. Office of Financial Sanctions Implementation

V. P. Munteanu, P. D. (2021). The Case for Agile Methodologies against Traditional Ones in Financial Software Projects. *European Journal of Business and Management Research* (February 16 2021), 741

Wilde, A. (2020, August 25). *Why consumers just don't trust fintechs*

SEEKING DESIGN SOLUTIONS

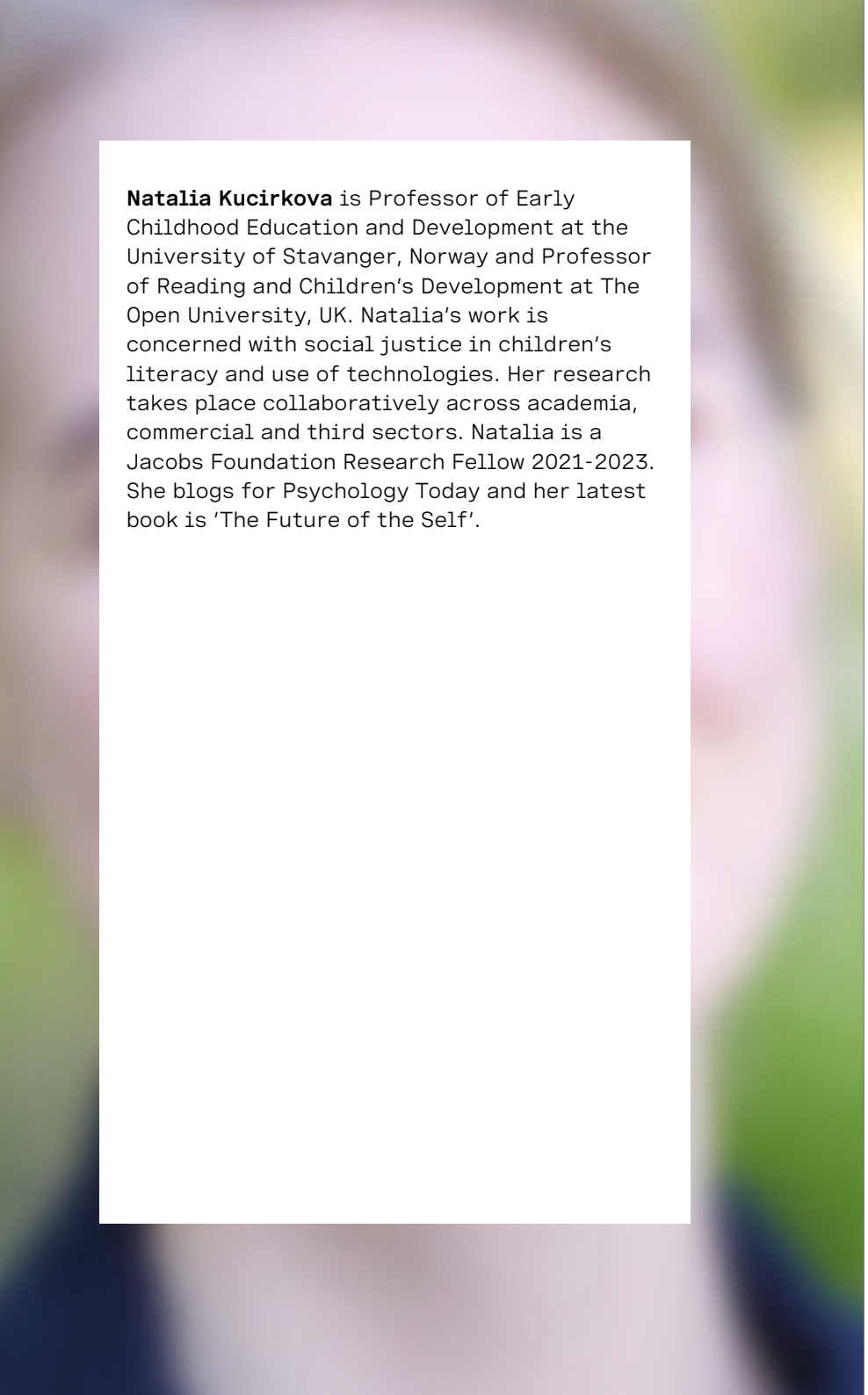
Zoom did not help at all. I could not concentrate on lessons (Girl, 16)

Not great. I don't find [Hegarty Maths] useful. Videos are too long and confusing (Boy, 15)

Show My Homework sometimes causes a distraction because I start doing my homework and then get a message or a notification and then have to check on that and I get carried away (Boy, 16)

I can't always get pages to work [on Google Classroom] (Girl, 15)

Digital Futures Commission (2022) EdTech survey



Natalia Kucirkova is Professor of Early Childhood Education and Development at the University of Stavanger, Norway and Professor of Reading and Children's Development at The Open University, UK. Natalia's work is concerned with social justice in children's literacy and use of technologies. Her research takes place collaboratively across academia, commercial and third sectors. Natalia is a Jacobs Foundation Research Fellow 2021-2023. She blogs for Psychology Today and her latest book is 'The Future of the Self'.

The promise and pitfalls of personalised learning with new EdTech

Natalia Kucirkova, University of Stavanger and The Open University

The coronavirus pandemic offered a unique opportunity to rethink the use of educational technologies (EdTech) for home and remote learning. EdTech brought the classroom home, facilitating teachers' access to communication with students and their families, collaboration among peers and management of digital information. While some EdTech platforms are designed to work offline, most of today's EdTech rely on online design solutions that process personal data. Some are designed with algorithms that can dynamically tailor the learning content according to individual students' progress and engagement. Such adaptive, data-driven EdTech is often described with the umbrella term 'digital personalised learning'. Although digital personalised learning design tends to motivate learners and streamline educators' work, personalised learning with EdTech is not without its pitfalls.

In this essay, I critically examine personalised EdTech's claimed benefits and limitations, before making some theorised, as well as tried and tested, suggestions for addressing its shortcomings. I focus on the commercially driven design logic of personalised EdTech, which must be discussed, understood and reconceptualised if EdTech is to offer learning benefits to all students.

The benefits of data-driven, adaptive EdTech

Personalised learning means adapting content to individual learners; that is, adjusting generic content to increase and make better an individual student's learning experience. The cost of such personalisation is high, however, and the possibility of outsourcing some of the personalisation of such labour to technologies helps to drive the recent interest in data-driven EdTech.

In EdTech, data are processed by algorithms designed to group similar characteristics together and categorise patterns of engagement. This is helpful for providing personalised feedback when teachers can't attend to each individual student, thus saving teaching time. When using the Duolingo app, for example, students receive automatic personalised feedback on their progress through the app, along with assignments that are tailored to them based on Duolingo's personalised learning engine.

The sophistication of individual algorithms varies. Some only collect test scores while others contain artificially intelligent tutors, with very different applications available across school subject areas. The purpose of data use in EdTech also varies widely. Some EdTech are used for monitoring school attendance (e.g., AppSheet), while others are used for monitoring learning progress (e.g., Naviance). Some EdTech have added features that allow users to exercise some control over their experience. These rely on data contributed by users themselves or on data extracted automatically by individual apps, games and platforms. With creative apps such as Scratch, for example, students can make their own designs, and with Night Zookeeper they can write their own stories.

Given the variety in how data are used and for what purpose, it is difficult to provide a simple account of the benefits and limitations of each EdTech application. What is crucial to consider when thinking about the added value of personalised EdTech is *how* the technology uses personal data and the algorithms processing that data.

The commercial design of EdTech

There is no doubt that data collection has provided a huge opportunity for the commercial sector. Commercial interest in data use is reflected in some features of EdTech that follow the logic relevant for economic but not learning gain. The commercial side of personalisation increases the benefits first and foremost for the commercial provider, and then for the user. The consequences of this have been widely reported in terms of data misuse, but the underlying design principles are less well known in the EdTech circles.

There are essentially two design principles that need to be understood here: (1) the principle of exponential data growth and the assumption that more data is always better; and (2) the like-like design principle and the assumption that recommending similar content is always beneficial. Both assumptions are rooted in economic theories about profit and psychological theories about engagement. These assumptions do not follow educational theories.

Commercial assumption 1: Exponential data growth

With data-collecting tools in almost everyone's pocket, the quantity and diverse nature of data increases every day. Experts predict that 463 exabytes of data will be circulated worldwide by 2025 (Seeds Scientific, 2021). Personalised EdTech will contribute to such exponential growth of data in the 21st century. The hunt for more and more data is driven by a commercial logic: the data economy runs with the mantra 'more data is better data'. This exponential growth in data is part of trickle-down economics where those who aggregate data profit from the data value much more than those who produce it.

In Kucirkova (2021), I describe the problem of exponential data quantity in relation to growing data complexity and its impact on children's development. With data that are being collected through multiple channels of several technologies, the portfolio of child's data becomes complex and relatively comprehensive. On the one hand, this helps with diagnostics: for example, when composing a child's reading profile, knowing how much, where and in which way (digital/analogue) the child reads, which genres and types of texts the child accessed etc., can provide a more accurate reading profile than can be afforded by data from a single e-book session.

On the other hand, the data amount and complexity creates issues with data ownership (e.g., who owns children's data when they transition from kindergarten to school and later to university?), interpretation (e.g., which criteria are used to holistically interpret data on a child's behaviour collected from school, social media and other sources?) and deployment (e.g., which subject areas or developmental goals are prioritised for applying intelligence from data?). Aggregated data require a certain level of data literacy, that is, digital and social competence for processing and interpreting numbers, trends and patterns. Children and their key caregivers, parents and educators are generally not equipped with this competence. Given the large and varied data sources, it is more manageable and convenient for schools to delegate the processing and interpretation of data to EdTech providers, which empowers them to not only collect and process data, but also to construe meanings about the data, and thus directly influence decisions about children's lives. The increasing complexity of large amounts of data and the exponential data growth may enable an unprecedented form of social control through the data it creates (Williamson, 2019).

In the hunt for more and more data, we need to ask - why do we need all this data? Personalised EdTech is being designed with the commercial goal of collecting increasing amounts of data rather than the nuanced understanding of which data are necessary for which purpose. Exponential and uncritical data collection leads to so-called 'datafied' childhoods and data-driven schools (see, for example, Lupton & Williamson, 2017), where data-driven, numeric and test-based evaluations of students' abilities carry greater weight than human assessment, gradually de-professionalising and eroding trust in caregivers' and teachers' judgements about children.

Commercial assumption 2:
The like-like recommendations

The design of a large proportion of personalised EdTech is modelled on the like-like logic of recommendation algorithms embedded in social media platforms - if you like X, the system recommends something similar (XX), and then again something similar (XXX), so that you gradually get something

that is more precisely relevant to the initial interest category. The logic works well when you look for a group sharing your niche interest, for example. The logic works less well for creating new ideas and expanding viewpoints.

The like-like logic locks users into bubbles of like-minded individuals, which carries the risk of reinforcing group views. With a steady flow of similar information presented as 'recommended' and 'just for you', the algorithms stealthily increase the feelings of shared belonging and universal truths. Homogeneity of thinking and lack of diversity are the breeding grounds for the dangerous pattern of groupthink (when a group reaches a poor decision because of similarity in ideology and background of the group members) and parochial empathy (when one feels more empathy towards those who are of similar background). When children are grouped according to similar scores, needs or preferences, the cognitive and social benefits that come with exposure to or active engagement with diversity are minimised. It is therefore essential to discuss and be aware of these design limitations so that they are avoided in EdTech design.

A like-like design in personalised EdTech is a far cry from design principles based on learning sciences. Instead of supporting collaboration and shared sustained thinking, the design promotes behaviourist learning. In such a limited model of learning, students' achievement is reduced to narrowly defined objectives where rewards are given for small task completion to extrinsically motivate students to continue with the task. Students are given badges for successful task performance, despite studies showing the ineffectiveness of such reward mechanisms for students' intrinsic motivation (Kyewski & Krämer, 2018). Each click or tap triggers a response that pushes the child towards a desired goal - as if there was only one right answer for each question. The like-like design exposes children to content that follows a linear trajectory of incremental progress, with little room for serendipitous discoveries or learning through surprise. Possibly, such a design is suited for drill learning, but not for understanding complex concepts (see Meyer et al., 2021).

The limitations of commercial design do not need to diminish EdTech's contribution to children's learning. The

learning sciences offer frameworks that educators can use to critically appraise the contribution of personalised EdTech to their classrooms. To address the exponential data growth problem, data needs to be used strategically and proportionally.

Educational assumption 1:
Strategic cut-offs for data generation

To personalise learning, data should be used to widen students' horizons and enrich their social relationships. It follows that we need to stop thinking about personalised EdTech as a panacea for post-pandemic education. Instead, developers, designers and educators need to consider *which* aspects of the educational pathway should be personalised. This needs to happen in a dynamic framework. Tetzlaff, Schmiedek and Brod (2020) developed a dynamic framework for thinking about such strategic data use. Their framework highlights intra-individual variability as a source of information for facilitating teachers' own judgement that could replace automatic loops and thus enhance the instruction support. The framework helps us see personalised EdTech in terms of its different impact on different students and different types of data for different students' needs.

Educational assumption 2:
Personalise and diversify

Acknowledging the commercial interest in the design of personalised Web 2.0, we can quickly see why the content needs to be relevant from the retailer's perspective: offering their client a recommendation for a new coat that is a complete mismatch from what they browsed and purchased recently is unlikely to result in a transaction. In the case of personalised EdTech, recommendations for content need to be based both on content units that are similar and also content units that are different from the students and their immediate surroundings (Munnich & Ranney, 2019).

Research shows that learning that 'sticks' is learning that is effortful (Brown et al., 2014). Concepts that are remembered over time are those that require deeper and longer engagement, which often runs counter to learners' preferences. While adapting content to match learners' needs might engage them it may lack the cognitive challenge required for processing the learning content.

Redesigning EdTech with educational principles

Successful education programmes need to personalise as well as diversify, and EdTech can be designed to accommodate both educational ideals. Diversification is achieved with purposefully designed content that is different from personalised content, a mechanism we refer to as 'personalised pluralisation' (Kucirkova & Littleton, 2017). The optimal model combines personalised information (relevant to an individual student) with content that is relevant to collectives (relevant to the classroom or peer cohort). It follows that personalised education not only needs to be implemented, but also co-designed, with families, teachers and communities. Such an assets-based perspective has been used in personalised trackers, success plans and navigators that show individual progress in relation to the progress of the community (e.g., individualised success plans can be transformational if they are both personalised and relationships-driven; see Sacks & Sedaca, 2021).

Redesigning EdTech with these principles implies not leaving it to commercial providers but to the communities of users. With courses and design opportunities offered by organisations like, for example, The Raspberry Pi Foundation, teachers and students can be technology co-producers. In other words, the sweet spot of learning lies in an optimal balance between the automation provided by EdTech and the teachers' and learners' own choices. The essence of this optimal balance is a combination of learners' agency with teachers' pedagogy *and* the technologies' affordances.

The '5 A's' of agency

So far in the essay, I have advocated co-design at the level of communities, and the importance of social relationships in learning. In these efforts, we need to reflect on and incorporate individual agency.

Agency, an individual's volition to make their own choices, can be thought of in terms of the '5 A's': Autonomy, Attachment, Authenticity, Aesthetics and Authorship. These '5 A's' are the

learning ingredients underpinning children's volitional choices. If they are present, EdTech can be considered to offer an educational foundation, but if they are absent, the commercial design principles, with their highly contestable assumptions, may become enshrined as strategic verities.

To elaborate, design that limits children's agency turns Authorship into consumption. Children's contributions are reduced to the providers' pre-designed templates, as with subscription programmes that furnish children with ready-made stories. In contrast, with design that invites children's agency, such as, for example, with open-ended story-making apps (e.g., Our Story), children can be the Authors of their own content. With EdTech that strips children of their Autonomy, children's agency is replaced with dependency. This happens, for example, with automated feedback loops that recommend the same content over and over. The feelings of Attachment to or ownership of a creative idea turn into dependency on a product. The Authenticity of children's own creations is reduced, and their Aesthetic sense is overpowered with adult design.

It is not just children who are stripped of their agency. With some of the bestselling digital libraries, teachers are positioned as curators and monitors of data rather than as co-readers and mentors (Kucirkova & Cremin, 2018). They are de-professionalised by having to rely on dashboards and templates that operate with a simplistic model of learning and make decisions on their behalf. Participatory design of EdTech could avoid these blind spots, but very few EdTech developers adapt a participatory research design approach. Products are presented to schools as ready-made tools, and teachers are positioned as consultants and testers of finalised designs. Disappointingly few EdTech designers think of children's involvement beyond the testing of prototypes that have been fully conceptualised and designed by adults. And despite the well-established tradition of participatory research design with children in human-computer interaction studies (e.g., Alison Druin's work on cooperative inquiry; Druin, 1999), children as co-designers of technologies are rarely involved in commercial EdTech production.

Learners can self-regulate, and learning practices that afford students agency over their learning facilitate self-regulation. This has been recently researched with the possibility of using personalised visualisations, which are external references to support learning (Molenaar et al., 2020). As learners set their own goals, evaluate their own progress and use personalised technology to visualise the process, they increase the accuracy of their performance with EdTech. This illustrates how the combination of technology-mediated and user-generated design, such as personalised visualisations, enhances self-regulation, which is known to be implicated in learning.

Conclusion

Whether data-driven personalised education lives up to its promise to educate is not yet known. However, as described in this essay, there are robust evaluation principles to guide the efforts. So that EdTech lives up to its promise of using personal data for advancing children's learning, the commercial design principles need to be replaced with educational design principles.

First, EdTech should be designed in ways that not only respect children's privacy and comply with child-inclusive policy but also minimise unnecessary data generation. Second, EdTech should be underpinned by algorithms that advance educational, ethical, moral and social goals by purposefully diversifying learning content. This is achievable as long as the personalised EdTech industry, pedagogy and policy abandon approaches inspired by commercial personalised technologies and adapt a culture of evidence and participatory co-design. EdTech developers, researchers and practitioners need to collaborate to ensure that data are used *strategically* to benefit individual *and* collective learning that advances human agency.

Brown, P., Roediger, H., McDaniel, M., & Stick, M. I. (2014). *The science of successful learning*. Harvard University Press

Druin, A. (1999). Cooperative inquiry: Developing new technologies for children with children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May, 592-599

Kucirkova, N. (2021). *The future of the self: Understanding personalization in childhood and beyond*. Emerald Group Publishing

Kucirkova, N., & Cremin, T. (2018). Personalised reading for pleasure with digital libraries: Towards a pedagogy of practice and design. *Cambridge Journal of Education*, 48(5), 571-589

Kucirkova, N., & Littleton, K. (2017). Developing personalised education for personal mobile technologies with the pluralisation agenda. *Oxford Review of Education*, 43(3), 276-288

Kyewski, E., & Krämer, N. C. (2018). To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Computers & Education*, 118, 25-37

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, 19(5), 780-794

Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R. M., & Radesky, J. (2021). How educational are "educational" apps for young children? App store content analysis using the Four Pillars of Learning framework. *Journal of Children and Media*, 15(4), 526-548

Molenaar, I., Horvers, A., Dijkstra, R., & Baker, R. S. (2020). Personalized visualizations to promote young learners' SRL: The learning path app. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, March, 330-333

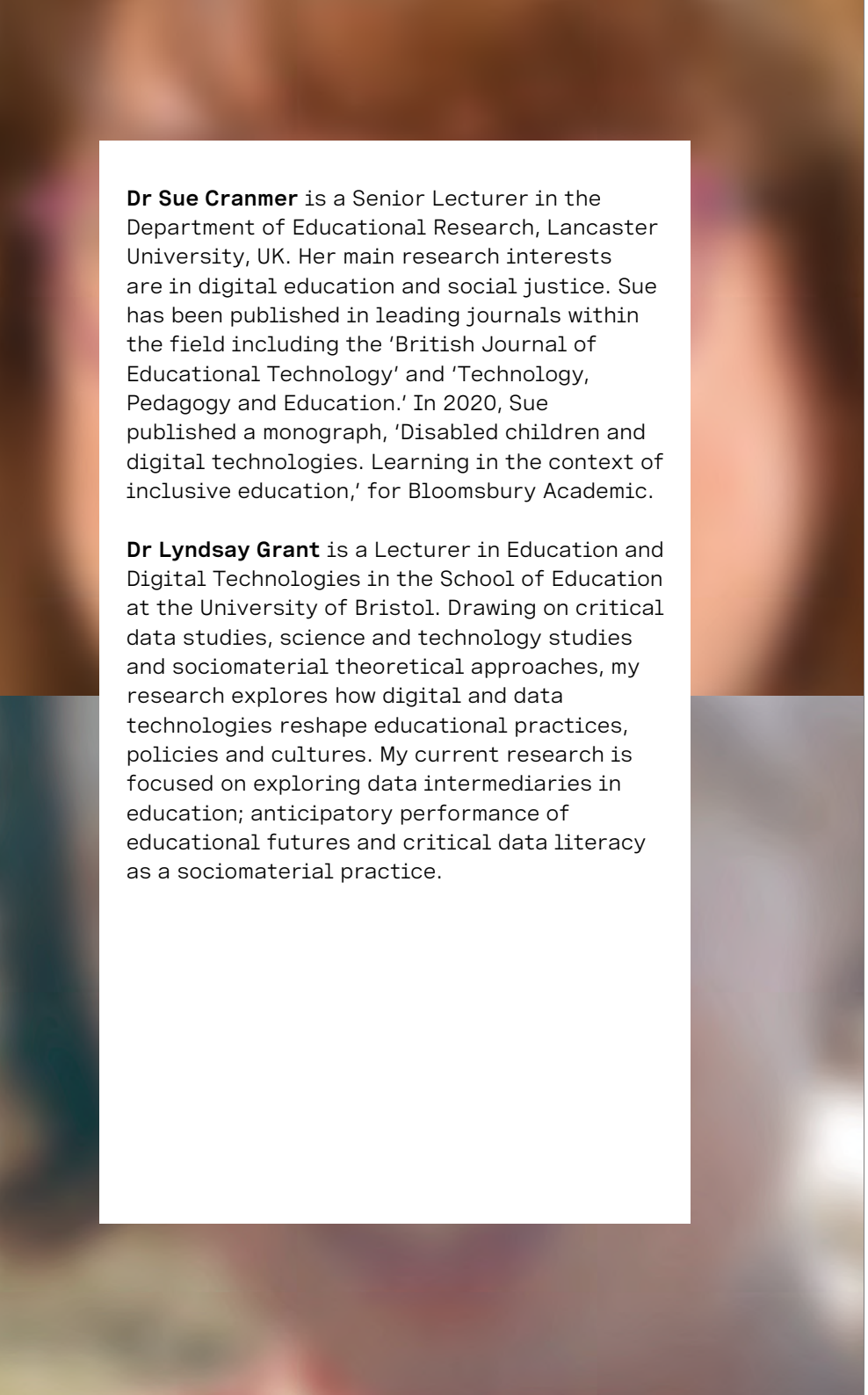
Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. *Topics in Cognitive Science*, 11(1), 164-177

Sacks, L. & Sedaca, M. (2021). Seizing the Moment for Transformative Change: A Framework for Personalized Student Success Planning. Cambridge, MA: Education Redesign Lab.

Seeds Scientific. (2021). How much data is created every day? [27 Staggering Stats]

Tetzlaff, L., Schmiedek, F., & Brod, G. (2020). Developing personalized education: A dynamic framework. *Educational Psychology Review*, 1-20

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education*, 62(3), 354-370



Dr Sue Cranmer is a Senior Lecturer in the Department of Educational Research, Lancaster University, UK. Her main research interests are in digital education and social justice. Sue has been published in leading journals within the field including the 'British Journal of Educational Technology' and 'Technology, Pedagogy and Education.' In 2020, Sue published a monograph, 'Disabled children and digital technologies. Learning in the context of inclusive education,' for Bloomsbury Academic.

Dr Lyndsay Grant is a Lecturer in Education and Digital Technologies in the School of Education at the University of Bristol. Drawing on critical data studies, science and technology studies and sociomaterial theoretical approaches, my research explores how digital and data technologies reshape educational practices, policies and cultures. My current research is focused on exploring data intermediaries in education; anticipatory performance of educational futures and critical data literacy as a sociomaterial practice.

Can disabled children benefit from education data?

Sue Cranmer, Lancaster University and
Lyndsay Grant, University of Bristol

In this essay we discuss the possible benefits for disabled children of the collection, processing and use of their education data in schools. Our conceptualisation of 'disabled children' is based on theories of childhood (McLaughlin, 2008) that argue that all children are entitled to the high 'expectations, opportunities, and aspirations afforded to the so-called typically developing children' (Goodley et al., 2016, p. 6). These approaches challenge prevalent notions of disabled children defined against typical children's development 'norms' seen to have undermined the value accorded to disabled children's 'ordinary' and 'productive childhoods' (Curran & Runswick-Cole, 2014, p. 1619). The term 'disabled children' is used here, therefore, to emphasise the social model of disability and how social, economic and political systems impact disabled children's lives. By 'education data' we follow the definition taken by the Digital Futures Commission, 'data collected about children at school and through their participation in school' (Livingstone et al., 2021, p. 3). Our viewpoint is framed by Human Rights legislation and informed by a commitment to inclusive education for disabled children (UN, 2016; UNICEF, 2017).

Trend of rising datafication in education

While the collection of education data is not new in itself (Lawn, 2013), the last 20 years has seen an intensification of the volume and scope of data collected about children at school and how it is applied to make decisions around education governance, pedagogy and practice (Grek, 2009; Ozga, 2009). Alongside this has been a corresponding trend towards the digitisation of education, under the premise that educational 'big data … can be used to [both] gain insights into the problems of education, and to find solutions at the same time' (Williamson, 2017, p. 3). The datafication of education, comprising the collection of previously unimaginable volume of data, alongside digital algorithmic and artificial intelligence (AI) processing, is now increasingly used to determine educational decisions (Grant, 2017; Jarke & Breiter, 2019; Mayer-Schönberger & Cukier, 2014; Williamson, 2017).

Limited empirical evidence for benefits of education data

Many claims have been made for the potential of education data to improve pupils' and schools' educational performance, but so far the evidence of a positive impact on learning outcomes in real-world educational settings is limited (Viberg et al., 2018; Williamson & Eynon, 2020). As for the positive impact of education data technologies specifically on disabled children's learning, there is a mixed picture. While there are useful examples of the research and development of digital technologies to support disabled learners (e.g., Metatla et al., 2020), there is limited published research that focuses on the impact of education data on disabled children's learning and outcomes (Baek et al., 2022). This could exclude them from any potential benefits their peers may gain (Zheng et al., 2019).

There is, however, an emerging body of research identifying critical questions as well as risks associated with the datafication of education for children more generally. For example, reductive approaches to teaching and learning including narrowing of the curriculum and 'teaching to the test' (e.g., Bradbury, 2019; Grant, 2017; Knight & Buckingham Shum, 2017); the reproduction and amplification of biases and inequalities in automated systems (e.g., Andrejevic & Selwyn, 2020; boyd & Crawford, 2012; Selwyn, 2015); and threats to children's wellbeing and privacy (e.g., Lupton & Williamson, 2017; Manolev et al., 2018). Research on the risks and harms of algorithmic technologies (including surveillance, discrimination and bias) for disabled children is beginning to gain increased attention, but is still in its infancy (Brown et al., 2022).

Establishing principles for beneficial data use for disabled children

While acknowledging that the use of education data in schools raises a number of areas of concern, we also need to ask whether it might be possible for it to be used in ways that are genuinely empowering for disabled children. Clearly, any potential advantages of collecting, processing and applying education data in schools must be able to reap significant benefits for disabled children's learning, inclusion and wellbeing to justify potential risks. Disabled children's digital practices must support their best interests alongside protection of their rights.

Our approach to this is underpinned by our experiences of conducting research with disabled children, digital technologies and education data in UK schools (Cranmer, 2020a, 2020b; Grant, 2022). Our understanding is informed by principles intended to foster inclusive education to ensure an equitable education for disabled children globally, enshrined in international law and founded on human rights (Pijl et al., 1997; UNICEF, 2017). However, the aims for inclusive education are often not fully realised. In practice, disabled children are integrated into schools in ways that need them to adapt to existing approaches rather than identifying and removing the barriers that prevent their inclusion. Teachers are ill prepared to support full inclusion, with disabled children often being 'referred out' or requiring adjunctive support to 'bridge' learning in the moment (Webster & Blatchford, 2017, p. 3). This potentially creates stigma, requires children to 'work around' inaccessible resources and activities, and undermines their independence.

We argue that for education data to be used in the best interests of disabled children, it should aim to support full participation in a genuinely inclusive education that challenges inequalities and deficit assumptions of disability to further

empower disabled children's agency around learning. We will consider the potential benefits and challenges of using education data in relation to inclusive education using UNICEF's framework (Figure 1), derived from Article 24 of the Convention on the Rights of Persons with Disabilities (UN, 2016; UNICEF, 2017).

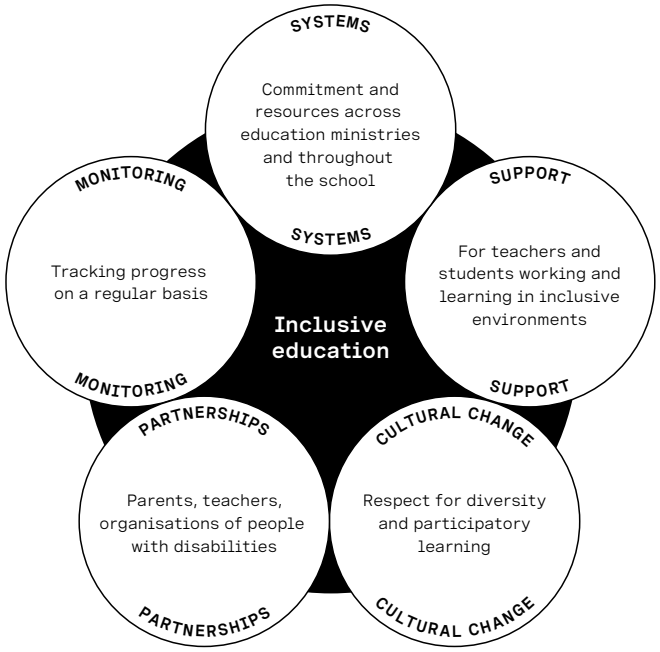


Figure 1
Source: UNICEF (2017)

Given the limited existing empirical evidence available on the benefits of education data for disabled children, in the following we draw on theoretical understandings of education data and datafication, empirical evidence from the use of education data with other groups of children as well as theoretical, empirical and policy research on disability and inclusive education to begin to articulate the *potential* benefits for disabled children's education within the five areas of the inclusive education framework introduced above. Each suggestion is a double-edged sword, however, because the possible benefits are likely to bring with them potential harms, so we also set out the associated potential challenges within each category.

Systems: Commitment and resources across education ministries and throughout the school

Potential benefits

The collection of systematic evidence about disabled children's current educational experiences and perspectives may support improvements in education policy. At a personal, institutional and national level, education data – including not just disabled children's assessed performance, but also their embodied and affective experiences of schooling – could be used to support policy-level changes including advocating for more (and more equitable access to) funding (e.g., Gallagher & Spina, 2021), specialist support and changes to pedagogical practice. Buckingham Shum (2012), for example, outlined how 'macro-level analytics' could enable cross-institutional analyses useful for evaluating and developing institutional and national improvements that foreground disabled children's experiences and voices. At school level, education data could be used as part of a process of action research and inquiry (see, for example, Armstrong & Moore, 2004), prompting investigation into disabled children's experiences and conducted with disabled children themselves, to identify where improvements are needed, and design and evaluate the impacts of change.

Challenges

The context in which education data may support inclusive and empowering system-level change for disabled children is crucial. High-stakes accountability measures such as published league tables, punitive school inspections and teacher evaluations can encourage performative approaches that prioritise improving data measures over more balanced approaches. This includes, for example, 'teaching to the test', or prioritising resources to those children on grade thresholds who are likely to make the biggest difference to school accountability measures (Bradbury et al., 2021; Grant, 2017).

Similarly, while data can provide useful evidence to support improvement, it will not in and of itself bring about necessary improvements without the political will, organisation and

funding to do so. The burden of collecting and providing data in order to access equitable and inclusive education also needs to be considered, as this potentially places further demands on disabled children and their families to 'prove' the reality of their lived experiences (D'Ignazio & Klein, 2018) or become a distraction for teachers from their core work of meeting students' needs (Gallagher & Spina, 2021).

Support: For teachers and students working and learning in inclusive environments

Potential benefits

Many claims for the benefits of education data to support learning focus on 'personalising' children's learning through assessing and monitoring progress to offer targeted interventions and next steps (Thompson & Cook, 2017). For disabled children, suggestions for data-driven personalisation include the early identification of disability-related difficulties (Jiménez-Gómez et al., 2015), identification of accessibility issues that can prevent access to learning opportunities, better targeting and allocation of resources and content, automated record keeping and feedback, identification of interventions to increase support and adaptation of resources and materials (Livingstone et al., 2021). Examples include Language ENvironmental Analysis (LENA) and Ubisense, a real-time indoor location system, to capture spatial, speech and time data work to identify physical improvements in classrooms for those at risk of communication difficulties (Sangwan et al., 2015); games-usage analytics to consider the impact of motion-based games (i.e., Kinect) on children with a range of disabilities (Kosmas et al., 2018); and video-coding software (e.g., Studiocode) automated data-coding to support learning by all students including those with disabilities (Kaczorowski & Raimondi, 2014). Even so, in both of the latter projects, results from automated data analysis were complemented by data collected by more traditional methods such as interviews and field notes.

Challenges

Challenges associated with data-driven personalisation and support may lead to reductive approaches to education, in which decisions about disabled children's learning and access to the curriculum is decided by algorithm, with neither teachers nor children aware of or involved in the decisions that concern their education (Knight & Buckingham Shum, 2017). For disabled children, highly targeted personalised content can potentially risk excluding them from the opportunities offered to all children, fuelling the 'intervention culture' in which children are removed from mainstream classrooms and activities in order to catch up with expected standards (e.g., Bradbury, 2019; Grant, 2017).

While education data may provide a useful part of the picture in terms of early diagnosis of disabilities and identification of accessibility issues, where it is focused on overcoming impairments and offering alternative opportunities, this risks perpetuating current approaches that tend to stigmatise disabled children and lead to a loss of independent learning (Cranmer, 2020a). Being diagnosed with a disability is a highly sensitive process and needs to be approached with care and caution. Automated diagnoses may be inaccurate. Not all children (or their parents and caregivers) wish to receive a formal diagnosis, and over-reliance on data-driven diagnoses risks labelling and perpetuating the current status quo whereby some individuals are identified as having specific 'needs' requiring extra support rather than ensuring that all children are provided with an equitable and inclusive education.

It is also important to consider how automated decision-making systems in many areas of life have been shown to reproduce existing social inequalities and exclusions. For example, facial recognition technology used in virtual proctoring software may fail to recognise individuals whose disabilities affect their appearance and is more likely to misgender women and individuals with darker skin (Brown et al., 2022; Buolamwini & Gebru, 2018).

Monitoring: Tracking progress on a regular basis

Potential benefits

Monitoring children's progress to support learning necessarily overlaps with the previous category. Even so, it is possible to draw out examples, whereby monitoring is the predominant

feature. These include the identification of internal and external factors that can support or hinder learning progress through automated record keeping; combining large datasets to build new insights; collection of longitudinal data; wellbeing issues such as attendance and behaviour management; and supporting administrative tasks such as performance management, resource and funding allocation (Livingstone et al., 2021). Lenz et al. (2016), for instance, have speculated about how trends in 'big data' could potentially support neurodivergent students such as those with dyslexia and dyscalculia to learn. They argue that the increased use of mobile and wearable devices, including outside of school, will enable comprehensive, long-term monitoring of behaviour to enable more appropriate support.

Challenges

In general, automated and data-driven monitoring of disabled children's education and learning may fail to recognise the specificity of disabled children's experience, for example, whether 'disability' is included as a category at all and the heterogeneity that exists among disabled children (Wald, 2021), or how children's agency governs their preferences and experiences. Some disabled children, for instance, like using mobile devices for learning while others reject them outright (Cranmer, 2020b).

Tracking individual progress towards specified performance targets is based on normative expectations of what a child 'should' achieve or how quickly they should progress (e.g., Llewellyn, 2016), how much they should attend school, or behavioural expectations that may not be appropriate for disabled children (or indeed, for many children). Performance targets are often derived from averaged data that do not reflect any individual child, let alone those who might be 'outliers' from the mean.

Disabled children may be particularly exposed to such risks of disciplinary surveillance in automated monitoring software. For example, virtual exam proctoring software is more likely to flag disabled students as 'suspicious' because of their access needs, and interpret neurodivergent behaviours and language differences as evidence of 'threat' (Brown et al., 2022).

Finally, the risks to disabled children's personal privacy and data protection are significant. Surveillance and monitoring have become normalised in schools, with children's digital activities and social media use being closely monitored even outside of school, and often without pupils or parents giving meaningful consent or adequate compliance with data protection guidance and legislation (Defend Digital Me, 2020). Furthermore, data about disability is sensitive. This means that privacy and data protection concerns are tantamount, particularly given the increasing number of cybersecurity breaches, for example the recent leak of around 820,000 New York students' personal data, including special education status, by an online platform (Elsen-Rooney, 2022).

Partnerships: Parents and caregivers, teachers, organisations of people with disabilities

Potential benefits

A suggestion that could be considered under this heading is that of multiagency hubs (Livingstone et al., 2021). Sharing disabled children's education data between different agencies, with appropriate data protection safeguards in place, may be one way to aid shared decision-making in the best interests of disabled children and involve parents and caregivers, teachers and other support personnel in understanding a child's experiences. Data may be able to make collaboration processes, such as sharing information and decision-making, more open in terms of how disabled children are supported effectively by teachers, families and other organisations to provide a foundation for further development and sharing best practice.

Challenges

The sharing of information does not, of itself, ensure that appropriate action is taken. For example, one study found that education data visualisations intended to inform school choice were largely ignored by parents, who found them difficult to locate and interpret (Fontaine & Dave, 2018). Dashboards designed for data sharing create particular expectations around student progress, imply certain roles for those involved in their education and can exclude children themselves from

interpretation and decision-making. For example, extrapolating predictions based on past performance can embed expectations that certain outcomes, such as dropping out of school, are inevitable in ways that 'reduce student agency, strengthen systemic disadvantage and foreclose the anticipation of different, unusual, unexpected futures for students' (Jarke & Macgilchrist, 2021, p. 3).

Further, while data sharing as the beginning of a multiagency conversation may be productive, in a context of high-stakes accountability targets, it can all too easily lead to more managerialist approaches to education, in which teachers become education data 'managers,' with reduced scope for more professional and contextualised decision-making (Ball, 2015; Lewis & Holloway, 2019; Selwyn, 2015).

Cultural change: Respect for diversity and participatory learning

Potential benefits

This category is challenging, but in principle, it is possible to use data to represent the diversity and variety of experience and perspectives rather than using it to define norms and averages, as in current uses of data. Data practices could also be used to challenge ableism by collecting data that questions deficit models of disability and makes more visible disabled children and people's achievements and abilities.

Challenges

In practice, education data is currently used to 'optimise' pupil performance through close monitoring towards a set of tightly defined and nationally standardised targets rather than to recognise and represent diversity (Amsler & Facer, 2017). Such narrowly defined forms of 'success' cannot account for multiple forms of achievement among groups of children with diverse skills, strengths and knowledge, including disabled children. Education data practices that truly respect diversity and participatory learning need to step away from current models that focus on individual assessment data, to account for learning as a participatory and collaborative collective endeavour.

Moving forward

Drawing on principles of inclusive education, alongside theoretical and empirical evidence from critical data studies broadly, and in education more specifically, we can begin to consider what the conditions for a genuinely empowering approach to education data for disabled children might be. This could usefully draw on recent approaches that centre the lived experiences and situated knowledges of people and groups to directly challenge power inequalities and act towards greater social justice, for example, intersectional data feminism, data activism and data justice (D'Ignazio & Klein, 2018; Dencik & Kaun, 2020; Kennedy, 2018).

Using education data to benefit disabled children means a significant shift towards disabled children themselves, their families and key support personnel, in who has the agency to make decisions about, for example, what data is collected, how that data is interpreted, and how it is used to determine decisions. This might enable disabled children and those who support them to identify and evidence issues that could be improved, and find ways of using existing data sources to show the scale of the issue and advocate for better educational opportunities. It might also explore what opportunities for disabled children's agency exist or could be developed in existing data arrangements, for example, understanding whether disabled children are able to opt in or out of data collected about them, question or refuse data-driven decisions made about them, and explore how disabled children experience and feel in relation to how their data is collected and used (Kennedy, 2018).

Centring disabled children themselves in education data practices is essential to challenge the multifaceted barriers to inclusive education in educational structures, approaches, inclusive/exclusive pedagogies and content (UN, 2016, in Slee, 2018, pp. 23-4). An example of centring children's voices and needs in data is UNICEF's Data for Children Collaborative,[1] which develops collaborative and child-centred data collection and analysis projects aimed at improving outcomes for children.

We also need to be clear about when education data becomes a solution in search of a problem. The most important

issues facing disabled children in their education are not necessarily amenable to data-driven solutions. We see some potential for more empowering uses of data that foreground the experiences of disabled children, and for the use of education data to support arguments for structural and institutional change based on increased awareness of barriers to inclusive education, increased funding (e.g., Gallagher & Spina, 2021), training and resources. However, there is a risk that an over-emphasis on education as a form of 'technological solutionism' (Morozov, 2013) can overshadow or displace the need for attention on other potential responses to support disabled children's education and inclusion, including structural reforms and political interventions that recognise and uphold disabled children's right to a genuinely inclusive education.

Amsler, S., & Facer, K. (2017). Contesting anticipatory regimes in education: Exploring alternative educational orientations to the future. *Futures, 94*, 6-14

Andrejevic, M., & Selwyn, N. (2020). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology*, 45(2), 115-128

Armstrong, F., & Moore, M. (2004). *Action research for inclusive education changing places, changing practice, changing minds*. RoutledgeFalmer

Baek, C., & Aguilar, S. J. (2022). Past, present, and future directions of learning analytics research for students with disabilities. *Journal of Research on Technology in Education*

Ball, S. J. (2015). Education, governance and the tyranny of numbers. *Journal of Education Policy, 30*(3), 299-301

Boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679

Bradbury, A. (2019). Datafied at four: The role of data in the 'schoolification' of early childhood education in England. *Learning, Media and Technology, 44*(1), 7-21

Bradbury, A., Braun, A., & Quick, L. (2021). Intervention culture, grouping and triage: High-stakes tests and practices of division in English primary schools. *British Journal of Sociology of Education, 42*(2), 147-163

Brown, L. X. Z., Shetty, R., Scherer, M. U., & Crawford, A. (2022). Ableism and disability discrimination in new surveillance technologies. Center for Democracy and Technology

Buckingham Shum, S. (2012) *Learning analytics: Policy briefing*. UNESCO Institute for Information Technologies in Education

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of machine learning research: Conference on fairness, accountability and transparency* (Vol. 81, pp. 1-15)

Cranmer, S. (2020a). *Disabled children and digital technologies: Learning in the context of inclusive education*. Bloomsbury Academic

Cranmer, S. (2020b). Disabled children's evolving digital use practices to support formal learning: A missed opportunity for inclusion. British Journal of Educational Technology, 51(2), 315-30

Curran, T., & Runswick-Cole, K. (2014). Disabled children's childhood studies: A distinct approach? Disability & Society, 29(10), 1617-1630

D'Ignazio, C., & Klein, L. (2018). *Data feminism*. MIT Press Open

defenddigitalme. (2020). *The state of data 2020: Mapping a child's digital footprint across England's state education landscape*

Dencik, L., & Kaun, A. (2020). Datafication and the welfare state. *Global Perspectives, 1*(1), 1-8

Elsen-Rooney, M. (2022). Personal data of 820,000 students exposed in NYC hack. *GovTech Today*, 28 March

Fontaine, C., & Dave, K. (2018). *Spectrum of trust in data: New York City parents navigating school choice*. Data and Society Research Institute

Gallagher, J., & Spina, N. (2021). Caught in the frontline: Examining the introduction of a new national data collection system for students with disability in Australia. *International Journal of Inclusive Education, 25*(12), 1410-1424

Goodley, D., Runswick-Cole, K., & Liddiard, K. (2016). The DisHuman child. *Discourse: Studies in the Cultural Politics of Education, 37*(5), 770-784

Grant, L. (2017). 'Don't use professional judgement, use the actual number': The production and performance of educational data practice in an English secondary school. Doctoral thesis, University of Bristol

Grant, L. (2022). Reconfiguring education through data: How data practices reconfigure teacher professionalism and curriculum. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *The ambivalences of data power: New perspectives in critical data studies*, pp.217-239. Palgrave (Springer)

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. J*ournal of Education Policy, 24*(1), 23-37

Holon IQ. (2021). $16.1B of global EdTech venture capital in 2020

Jarke, J., & Breiter, A. (2019). Editorial: The datafication of education. *Learning, Media and Technology, 44*(1), 1-6

Jarke, J., & Macgilchrist, F. (2021). Dashboard stories: How narratives told by predictive analytics reconfigure roles, risk and sociality in education. *Big Data and Society, 8*(1), 1-15

Jiménez-Gómez, M., Luna, J. M., Romero, C., & Ventura, S. (2015). Discovering clues to avoid middle school failure at early stages. *ACM International Conference Proceeding Series*, 16-20 March, 300-304

Kaczorowski, T. L., & Raimondi, S. (2014). eWorkbooks for mathematics: Mapping the independent learning experiences of elementary students with learning disabilities. *Journal of Learning Analytics, 1*(3), 179-182

Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis: Journal for Contemporary Philosophy, 1*, 17-30

Knight, S., & Buckingham Shum, S. (2017). Theory and learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 17-22). SoLAR (Society for Learning Analytics Research)

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends, 62*(6), 594-601

Lawn, M. (2013). *The rise of data in education systems: Collection, visualisation and use*. Symposium Books

Lenz, L., Pomp, A., Meisen, T., & Jeschke, S. (2016). How will the internet of things and big data analytics impact the education of learning-disabled students? A concept paper. *3rd MEC International Conference on Big Data and Smart City*

Lewis, S., & Holloway, J. (2019). Datafying the teaching 'profession': Remaking the professional teacher in the image of data. *Cambridge Journal of Education, 49*(1), 35-51

Livingstone, S., Atabey, A., & Pothong, K. (2021). *Addressing the problems and realising the benefits of processing children's education data: Report on an expert roundtable*. Digital Futures Commission/5Rights Foundation

Llewellyn, A. (2016). Problematising the pursuit of progress in mathematics education. Educational Studies in Mathematics, 92(3), 299-314

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society, 19*(5), 780-794

Manolev, J., Sullivan, A., & Slee, R. (2019). The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. Learning, *Media and Technology, 44*(1), 36-51

Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with big data: The future of education*. Houghton Mifflin Harcourt

McLaughlin, J. (2008). *Families raising disabled children: enabling care and social justice*. Palgrave Macmillan

Metatla, O., Bardot, S., Cullen, C., Serrano, M., & Jouffrais, C. (2020). Robots for inclusive play: Co-designing an educational game with visually impaired and sighted children. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery*, New York, 1-13

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs

Ozga, J. (2009). Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy, 24*(2), 149-162

Pijl, S. J., Meijer, C. J. W., & Hegarty, S. (Eds.). (1997). *Inclusive education, a global agenda*. Routledge

Sangwan, A., Hansen, J. H. L., Irvin, D. W., Crutchfield, S., & Greenwood, C. R. (2015) Studying the relationship between physical and language environments of children: Who's speaking to whom and where? *2015 IEEE Signal Processing and Signal Processing Education Workshop, SP/SPE 2015*, August, 49-54

Selwyn, N. (2015). Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology, 40*(1), 64-82

Slee, R. (2018). *Inclusive education isn't dead, it just smells funny*. Routledge

Thompson, G., & Cook, I. (2017). The logic of data-sense: Thinking through learning personalisation. Discourse: *Studies in the Cultural Politics of Education, 38*(5), 740-754

UN (United Nations). (2016). *Committee on the Rights of Persons with Disabilities. General Comment No. 4. Article 24: Right to inclusive education*

UNICEF. (2017). *Inclusive education: Understanding Article 24 of the Convention on the Rights of Persons with Disabilities*. UNICEF Regional Office for Europe and Central Asia

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98-110

Wald, M. (2021). AI data-driven personalisation and disability inclusion. *Frontiers in Artificial Intelligence, 3*

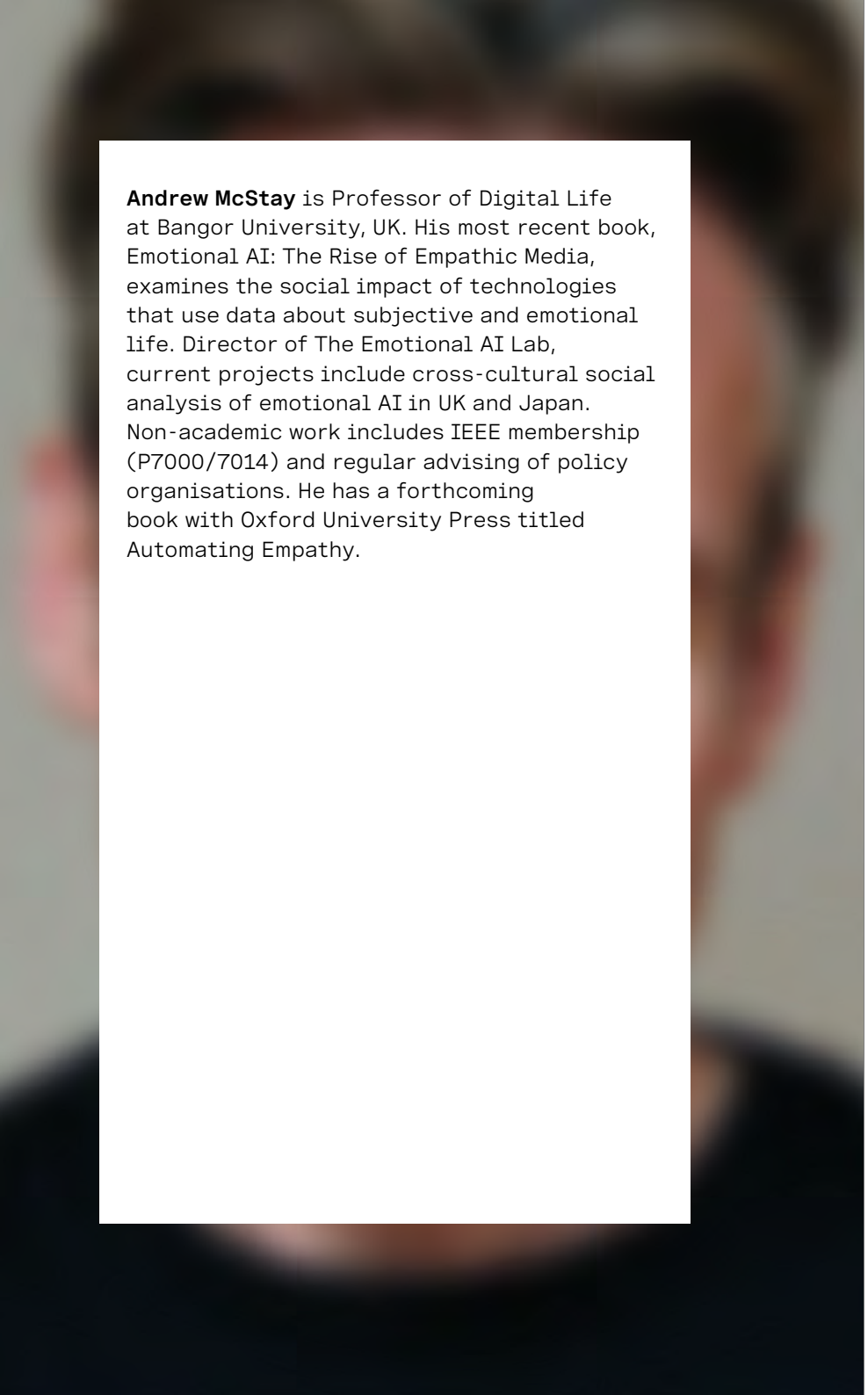
Webster, R., & Blatchford, P. (2017). *The special educational needs in secondary education (SENSE) study. Executive summary*. UCL Institute of Education

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Sage

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223-235

Zheng, G., Fancsali, S. E., Ritter, S., & Berman, S. R. (2019). Using instruction-embedded formative assessment to predict state summative test scores and achievement levels in mathematics. *Journal of Learning Analytics, 6*(2), 153-174

1 https://dataforchildrencollaborative.com/about-us



Andrew McStay is Professor of Digital Life at Bangor University, UK. His most recent book, Emotional AI: The Rise of Empathic Media, examines the social impact of technologies that use data about subjective and emotional life. Director of The Emotional AI Lab, current projects include cross-cultural social analysis of emotional AI in UK and Japan. Non-academic work includes IEEE membership (P7000/7014) and regular advising of policy organisations. He has a forthcoming book with Oxford University Press titled Automating Empathy.

Automated empathy in education: benefits, harms, debates

Andrew McStay, Bangor University†

This essay assesses what might appear a niche interest: the use of technologies to gauge emotion expressions, and whether a child is attentive and engaged. Variously labelled 'affective computing', 'emotion AI' and 'emotional AI', I use the meta-label, *automated empathy*, to cluster a variety of systems programmed to identify, quantify, judge, respond and interact with emotions, affective states, cognitive states, attention and intention (McStay, 2022: forthcoming). The essay considers claimed benefits and problems of these technologies, disputed usefulness in learning and educational development, and ethical questions about acceptability of using technologies in education this way, progressing to discuss these issues in the context of General Comment No. 25 (GC25) on the United Nations Convention on the Rights of the Child (UNCRC).

EdTech innovation in machine social and emotional learning

Educational technologies (EdTech), tutoring apps and related systems have scope to improve life chances through education, especially in regions where formal schooling is difficult (e.g.,

† This work is supported by the Economic and Social Research Council [grant number ES/T00696X/1].

because of distance from schools, violence in conflict areas, gender discrimination, cost of materials, overcrowding, poor curricula and inadequate teachers). Yet, where EdTech systems and automated empathy overlap, there are significant concerns, as this essay unpacks. As social and emotional learning balances cognitive elements in education (knowledge acquisition, analysis, reasoning and memory) with management of feelings and emotions, perseverance to achieve goals and ability to work with others, EdTech can help with the non-cognitive dimensions of learning. A stated benefit is that in the context of growing classrooms, all children receive close and recorded attention.

Companies building these systems include many start-ups and established technology companies. For example, the education branch of the global technology company Intel states that they are researching how recognition of emotion and disposition may help personalised learning. While not yet present in UK school classrooms, Intel Education's (2022) take on automated empathy involves three inputs to a classroom computer that records and predicts engagement during a class session. Inputs include *appearance*, where cameras extract facial landmarks, upper body and head movement and pose; *interaction* and how the student uses input devices such as a keyboard and mouse; and *time to action*, or how long the student is taking to complete tasks or act on a learning platform.

In addition to interest in deploying automated empathy in the physical classroom is interest in the virtual classroom, something of increased attention since the beginning of the COVID-19 pandemic. The EdTech industry saw an opportunity in home learning in that it recognised that embodied and interpersonal dynamics of in-classroom empathy were missed during the height of the COVID-19 pandemic. As a remedy, many start-ups and legacy technology firms suggested that their services could be used to gauge emotion, attention and interest. Intel (again), for example, have teamed with Classroom Technologies to develop 'Class' that runs on top of Zoom. This is claimed to detect whether students in the USA are bored, distracted or confused by assessing student facial expressions in relation to the educational content they are studying (Class, 2022).

Looking slightly further ahead to other automated empathy technologies that have realistic scope to be used in schools is metaverse-based interest in education, with Microsoft actively developing in this area. Scepticism of the metaverse is fair, but non-biometric mixed reality (through virtual reality [VR] and augmented reality [AR] systems) already has a growing presence in education, enabling students to 'feel-into' places, pasts, presents and futures, and even the constitution of objects (Microsoft Education, 2022). There is genuine value for students and teachers in this sort of non-biometric digital empathy in education, be this regarding the tangible impacts of climate change in faraway places, of biological systems, or of historical situations.

However, biometric profiling of facial expressions and student in-world interaction is looking likely through Microsoft's 'Mesh for Teams' that uses biometrics to map bodily behaviour and physical facial expressions onto in-world avatars, for novel 'Teams' meetings. Although 'Mesh for Teams' is aimed at the world of work (Roach, 2021), the potential for education is clear, through the existing presence of Microsoft Teams for online teaching. The goals of Intel, Microsoft and others are multiple, including quantification of what were qualitative phenomena, and providing an evidence base for education established on numbers, novel interactivity and performance metrics other than assignment scores and attendance.

For educators unfamiliar with biometric-automated empathy these technologies may have appeal, although lessons may be gleaned from countries that have practical actual experience with it. China's pilot tests with emotion profiling and automated empathy in the classroom are instructive. In reference to the Class Care System (CCS) from Hanwang Education, an extended report by human rights organisation ARTICLE 19 (2021) found that students feign interest and game the system to receive rewards. Self-policing may become an everyday occurrence, but 'chilling effects' (i.e., self-censorship) and being 'always-on' is not the sort of mindfulness we should be introducing. This is especially so given that students will perform for how they think the camera sees them (Andrejevic & Selwyn, 2019).

Indeed, van der Hof et al. (2020) see this through the prism of human dignity, because automated systems risk making de-individualised decisions without respecting the full and intrinsic worth of a human being. In China itself, tests with emotion-based automated empathy show it to be neither popular with students nor teachers, in part due to privacy questions, but also the lack of actionable feedback (ARTICLE 19, 2021). Does inattentiveness to part of a lesson, for example, signal boring content or boring delivery? Furthermore, the systems offer no suggestion on how to improve these. This is not to foreground practical over ethical concerns, but to provide a sense of limitations in practice.

Historical development

Use of technologies to gauge emotion and human disposition has a surprisingly long history, originating in the 1800s. Technically, this entailed pre-digital tracking of emotion by measuring temperature differentials, changes in heartbeat, blood pressure, breathing, conductivity of the skin and brain activity measures and facial coding, among other signals (Dror, 2001). How interiority was represented is also notable, as emotions were formalised into tables, charts and curves. Debates on positivism versus socially grounded understanding of emotion are beyond the limit of this essay, but emphasis on visualisation resonates with the modern educator usage of dashboards. Skipping centuries, cybernetic and computational apps to emotion have theoretical roots in the 1970s, with Manfred Clynes who argued for physical laws of emotion and its communication that could be rendered by computers. 'Sentics' for Clynes would help children 'be in touch with their emotions' and allow 'different races and backgrounds to experience their common basis in humanity' by being sensitive to the emotions of others (1977, p. xxii).

Rosalind Picard, the originator of the term and practice of 'affective computing', tried to put this into practice by building a 'computerized learning companion that facilitates the child's own efforts at learning' (Picard et al., 2001). The goal of the companion was to improve pedagogical techniques by using computer vision techniques to watch and respond to the affective states of children. By the 2010s, Sidney D'Mello's 'Affective AutoTutor' would detect and respond to learners' boredom, confusion and frustration. Through facial coding, and tracking of interaction patterns and body movement, this system sought to provide motivational feedback to students through appropriate facial expressions and voice emotion (D'Mello & Graesser, 2012). Related work was based on voice that, in addition to assessing whether verbal answers are correct, also seeks to detect learners' certainty or uncertainty (Forbes-Riley & Litman, 2011). Other work focuses on attention rather than emotion. Mention should also be made of teachers, as their teaching methods may also be subject to analysis through recording of in-classroom audio and automated methods to predict the level of discussions in these classes (D'Mello, 2017).

Pseudoscience?

Leading industry figures recognise the limitations of popular 'basic emotion' recognition technologies, with Microsoft publishing work in academic and technology journals saying so (McDuff & Czerwinski, 2018). Yet, despite Microsoft's own researchers publishing on this issue, for years this did not stop Microsoft from using this approach. Microsoft's Azure service, for example, labels 'basic emotion' facial expressions as happiness, sadness, neutral, anger, contempt, disgust, surprise and fear (Microsoft Education, 2022). Testament to the controversial nature of this approach to emotion recognition, Azure is slated to be discontinued in 2023 through publication of Microsoft's framework for building AI systems responsibly (Crampton, 2022). While this was widely interpreted to mean that Microsoft would desist from all work on emotion recognition, this is not what they said. Retirement of inference of emotional states applies only to their Azure Face services, with Microsoft adding that they 'need to carefully analyze *all* AI systems that purport to infer people's emotional states' (Crampton, 2022). This is a much weaker statement of intent than 'we have stopped all emotion recognition development'.

Despite Microsoft's retirement of emotion-based services in Azure, the method is popular. The Google Cloud Vision API, for example, also uses face landmark regions (e.g., mouth and eyebrows) to 'detect emotion' (Google Cloud, 2022).

There is a long line of scholars who will testify to this approach being a highly limited account of emotions, and that using 'reverse inference' to infer experience from expressions is questionable (Stark & Hutson, 2021). Adding to these voices, Barrett et al. (2019) observe that facial coding is especially poor with children, due to their immaturity and lack of development in emoting (also see McStay, 2019).

The reason why companies use simplistic approaches is simple: expedience. It is relatively easy to program systems to look for features (such as movement and actions of faces) and then match these arrangements to pre-given emotion expression labels. To question whether the full gamut of emotional life can be channelled through a suite of basic emotions, or whether expressions say much about experience, would add a lot of complexity for global technology firms seeking to deploy their products internationally. A universalist account of emotional life and subjectivity suits them well.

Despite highly vocal critique of claims of pseudoscience, this is not the core problem. A risk of a pseudoscience-based critique is that it invites *more* profiling and more granular labelling of brain, bodily and situational interactions. This would involve the connection of facial movements with factors connected to the personal and external contexts. For the person, it would include metabolic and historic dimensions (relating to the body and existing profiles of a person), and external factors including regional and societal norms on emoting, and specifics of the situation where the sensing is taking place (McStay & Urquhart, 2019). For example, is a child at home, in school, in virtual space, or in a mixed reality context? Who else is present? What is the situation? Who is teaching?

There are also accuracy problems - not only in psychological assumptions about the nature of emotion, but also in the curation of training datasets (regarding who does the labelling of an emotion expression and who is labelled). Overlapping with general concerns about AI bias against marginalised groups, market leading systems such as Microsoft and Chinese company Face++ have been found to label Black people with disproportionately negative types of emotion (notably, anger), especially if there is ambiguity of

what emotion label to give to a facial expression (Rhue, 2018). In work at our Emotional AI Lab, we tried to examine training datasets in terms of how they are constructed, who is doing the labelling, who is being labelled and the nature of this emotion profiling in relation to transport and usage in cars, but we found this to be an opaque and secretive practice as companies closely guard how their systems work (McStay & Urquhart, 2022). This is not to say that they are guarded because they are biased, but that industrial secrecy means that they are not open for public examination, despite social risks.

Child rights policies

The bundling of deeply questionable technologies with pro-social ambition risks lack of critical scrutiny. For example, internationally, pro-social emphasis on 'soft' abilities is something that influential bodies, such as UNESCO, see as 'fundamental to human creativity, morality, judgment, and action to address future challenges' (UNESCO, 2021, p. 68); but other key organisations see scope to instil these so-called soft abilities through questionable means, with the Organisation for Economic Co-operation and Development (OECD) seeing utility in measuring child sociality and emotion through affective computing (OECD, 2015). This illustrates the observation made by critical EdTech scholarship that rightly notes that datafication of emotion serves the overall education policymaking process around social and emotional learning, rather than children, through building of a psychometric evidence base (Williamson, 2019).

There are, of course, wider ethical and governance concerns. With an explicit focus on emotion and affect-based technologies, historically these have been under-served by tools such as the EU and UK General Data Protection Regulation (GDPR), which make no reference whatsoever to emotions. Similarly, the European proposal for the ePrivacy regulation rarely mentions emotions. Only Recitals 2 and 20 of the ePrivacy preamble mention emotions although, importantly, Recital 2 defines them as highly sensitive (McStay & Rosner, 2021). However, this lacuna is on absence of emotion profiling regulation being noted. In 2021 the United Nations Human Rights Council (UNHRC) formally adopted

the Resolution titled 'Right to privacy in the digital age' where §3 notes the need for safeguards for emotion recognition (UN General Assembly, 2021). More regionally, and with application to children, the Council of Europe (2021) likewise calls for strict limitations and bans in areas of education and the workplace. Also in 2021, the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) issued a joint statement declaring the use of AI to infer emotions of a natural person as highly undesirable, and that it should be prohibited, except for specified cases, such as for some health purposes (EDPB, 2021).

Relatedly, 2021 also saw the release of the proposed EU AI Act, a risk-based piece of legislation that classifies emotion recognition usage with children (such as in toys as well as education) as high risk (European Commission, 2021). Notably, Recital 28 of the proposed EU AI Act names the UNCRC and General Comment No. 25 that expands on rights regarding the digital environment (also see Articles 5.1b and 9 of the proposed EU AI Act). The UK itself does not have bespoke regulation on emotion profiling and children, although the Centre for Data Ethics and Innovation (part of the Department for Digital, Culture, Media & Sport) sees use of biometric data such as eye tracking, facial expressions and affective states as a way of improving understanding of levels of engagement and educational resource design (GOV.UK, 2021).

Of special interest to this essay on automated empathy and education is General Comment No. 25 (GC25) 2021 update to the UNCRC, especially because it recognises opportunities in new technologies, as well as seeking to define and defend rights. Lacklustre in name only, GC25 details how child rights in the digital environment should be interpreted and implemented by States around the world. The Emotional AI Lab responded to the call for evidence for GC25 (McStay et al., 2021). Unique among evidence provided to the call, we focused on harms associated with datafied emotion in education and toys. It appears that we were heard as GC25 contains multiple mentions of emotion analytics (§42, 62, 68), finding this to interfere with children's right to privacy, and freedom of thought and belief, also flagging the importance 'that automated systems or information filtering systems are not used to affect or influence children's behaviour or emotions or to limit their opportunities or development' (UNCRC, 2021, §62).

In relation to education itself, the digital environment is seen in the published GC25 as providing scope for 'high-quality inclusive education, including reliable resources for formal, non-formal, informal, peer-to-peer and self-directed learning' (§99), also with potential 'to strengthen engagement between the teacher and student and between learners' (§99). Surface consideration might see this as making the case for biometric-automated empathy for generalised use in the classroom, in platform-based learning and in an immersive VR/metaverse context. This would involve rendering physical facial expressions onto in-world avatars and profiling in-world interactions between students and teachers for future reference. Yet, as detailed, there are deep methodological and discriminatory problems that mitigate against this reading of GC25 in relation to emotion, immersive media and biometrics.

However, non-biometric mixed reality (through VR and AR systems that enable students to 'feel-into' places, pasts, presents, objects and imagined futures) is seen here as having pedagogic value, especially when used to deepen and enrich understanding of a topic (Daniela, 2020). In this regard, §101 is also notable, seeking to ensure that the 'use of digital technologies does not undermine in-person education and is justified for educational purposes', which points to an intrinsic belief of the value of in-person learning (and human-teacher empathy therein) and that promises of automated empathy for platform-based learning should not be allowed to undermine in-person embodied interaction.

Finally, §103 is also of keen relevance, specifying that standards for digital educational technologies should ensure that child personal data is not misused, commercially exploited or otherwise infringes their rights. Concern about datafied exploitation of children is longstanding, especially in relation to marketing and advertising (van der Hof et al., 2020), but this is extended by automated empathy in EdTech in two ways.

First, because in the context of automated empathy in education, inferences about students' emotions are used to train the neural networks owned by EdTech providers for purposes outside of education. Consequently, aggregated data

about child emotion would be commodified to improve algorithmic services, create competitive difference (in terms of how many faces are analysed) and serve business and strategic contexts for which the student data was not intended (such as testing responses to ads, linking well to van der Hof et al., 2020).

Second is emphasis of §103 on *personal* data, rather than simply child data. This is significant because in many instances automated empathy vendors will argue that their systems only deal in aggregate impressions (such as overall levels of pupil attention and happiness), and that data collected cannot be linked back to an individual. There is good technical legal debate in that personal data must exist in these systems for a fragment of a second as the 'insight' is collected and aggregated (George et al., 2019), but in practice, this has not stopped use of this approach to emotion recognition in out-of-home advertising in Europe (McStay, 2020).

This essay recommends critical attention to aggregated as well as identifying practices, especially given the scope for chilling effects, self-censorship and surveillant experience of being 'always-on'. After all, this is the antithesis of the social and emotional learning that should take place in education. Moreover, the moral basis for recommended critical attention is not that aggregated data about students may conceivably be personal data due to the fraction of a second processing of personal data. Although it should be noted that EU and UK data protection does not prescribe a minimum amount of time personal data should exist within a data processing system for it to be governed by legal rights over personal data, the moral basis argued here is that privacy and related rights may be held by a group as well as individuals (Floridi, 2014; Wachter, 2020).

Conclusion

Between rights to freedom of thought, privacy and access to education, there is the glaring question of whether automated empathy can do what is claimed. However, even with improvements in methodology, automated empathy in education does not align with the need for mental and emotional reserve to ensure human flourishing. This essay concludes that automated empathy technologies are incommensurable with current and near future social values. The core methodological and normative problems are as follows:

- Serious questions about effectiveness, validity and social representativeness of training data
- Lack of alignment between financial incentives in automated empathy and the wellbeing of schoolchildren
- Moral problems in using aggregated inferences about children's emotions to train neural networks that will be deployed for other commercial purposes
- Mission creep, where in-class data may be used for other socially determining purposes (such as social scoring)
- Already demonstrated risk of self-surveillance and chilling effects in the classroom
- Data minimisation questions that ask whether automated empathy is necessary for successful education.

Andrejevic, M., & Selwyn, N. (2020). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology, 45*(2), 115-128

ARTICLE 19 (2021). *Emotional entanglement: China's emotion recognition market and its implications for human rights*

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1-68

Class. (2022). *Add teaching & learning tools to Zoom*. www.class.com

Clynes, M. (1977). *Sentics: The touch of the emotions*. Anchor Press/Doubleday

Council of Europe. (2021). *Consultative Committee of the Convention for the Protection of Individuals with regard to automatic processing of personal data. Convention 108: Guidelines on facial recognition*

Crampton, N. (2022). Microsoft's framework for building AI systems responsibly

Daniela, L. (2020). *New perspectives on virtual and augmented reality*. Taylor & Francis

D'Mello, S. K. (2017). Emotional learning analytics. In C. Lang, G. Siemens, W. Alyssa, & D. Gašević (Eds.), *Handbook of learning analytics and educational data mining* (pp. 115-127). Society for Learning Analytics Research.

D'Mello, S. K., & Graesser, A. (2012). AutoTutor and Affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems, 2*(4), 23, 1-39

Dror, O. E. (2001). Counting the affects: Discoursing in numbers. Social Indicators Research, 68(2), 357-378

EDPB (European Data Protection Board). (2021). *EDPB–EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. 18 June

European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. 21 April

Floridi, L. (2014). Open data, data protection, and group privacy. *Philosophy & Technology*, 27, 1-3

Forbes-Riley, K., & Litman, D. J. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication, 53*(9-10), 1115-1136

George, D., Reutimann, K., & Tamò-Larrieux, A. (2019). GDPR bypass by design? Transient processing of data under the GDPR. *International Data Privacy Law, 9*(4), 285-298

Google Cloud. (2022). *Vision AI*

GOV.UK. (2021). *AI Barometer Part 5 – Education*

Intel Education. (2022). *Applying artificial intelligence to transform how we learn*

McDuff, D., & Czerwinski, M. (2018). Designing emotionally sentient agents. *Communications of the ACM, 61*(12), 74-83

McStay, A. (2019). Emotional AI and EdTech: Serving the public good. *Learning Media & Technology, 45*(3), 270-283

McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy, *Big Data & Society*

McStay, A. (2022: forthcoming). *Automating empathy: When technologies claim to feel-into everyday life*. Oxford University Press

McStay, A., & Rosner, G. (2021). Emotional Artificial Intelligence in children's toys and devices: Ethics, governance and practical remedies. *Big Data & Society, 8*(1)

McStay, A., & Urquhart, L. (2019). 'This time with feeling?' Assessing EU data governance implications of out of home appraisal based emotional AI. *First Monday, 24*(10)

McStay, A., & Urquhart, L. (2022). In cars (are we really safest of all?): Interior sensing and emotional opacity. *International Review of Law, Computers & Technology*

McStay, A., Miyashita, H., Rosner, G., & Urquhart, L. (2021). *Comment on children's rights in relation to emotional AI and the digital environment*

Microsoft Azure. (2018). Face API

Microsoft Education. (2022). Face detection and attributes

OECD (Organisation for Economic Co-operation and Development). (2015). *Skills for social progress: The power of social and emotional skills*. OECD Skills Studies, OECD Publishing

Picard, R., Cassell, J., Kort, B., Reilly, R., Bickmore, T., Kapoor, A., Mota, S., & Vaucelle, C. (2001). *Affective learning companion*

Rhue, L. (2018). Racial influence on automated perceptions of emotions. *SSRN*, 9 November

Roach, J. (2021). Mesh for Microsoft Teams aims to make collaboration in the 'metaverse' personal and fun

Stark, L., & Hutson, J. (2021). Physiognomic artificial intelligence. *Fordham Intellectual Property, Media & Entertainment Law Journal, SSRN*. 20 September

UNCRC (United Nations Convention on the Rights of the Child). (2021). *General Comment No. 25 (2021) on children's rights in relation to the digital environment*

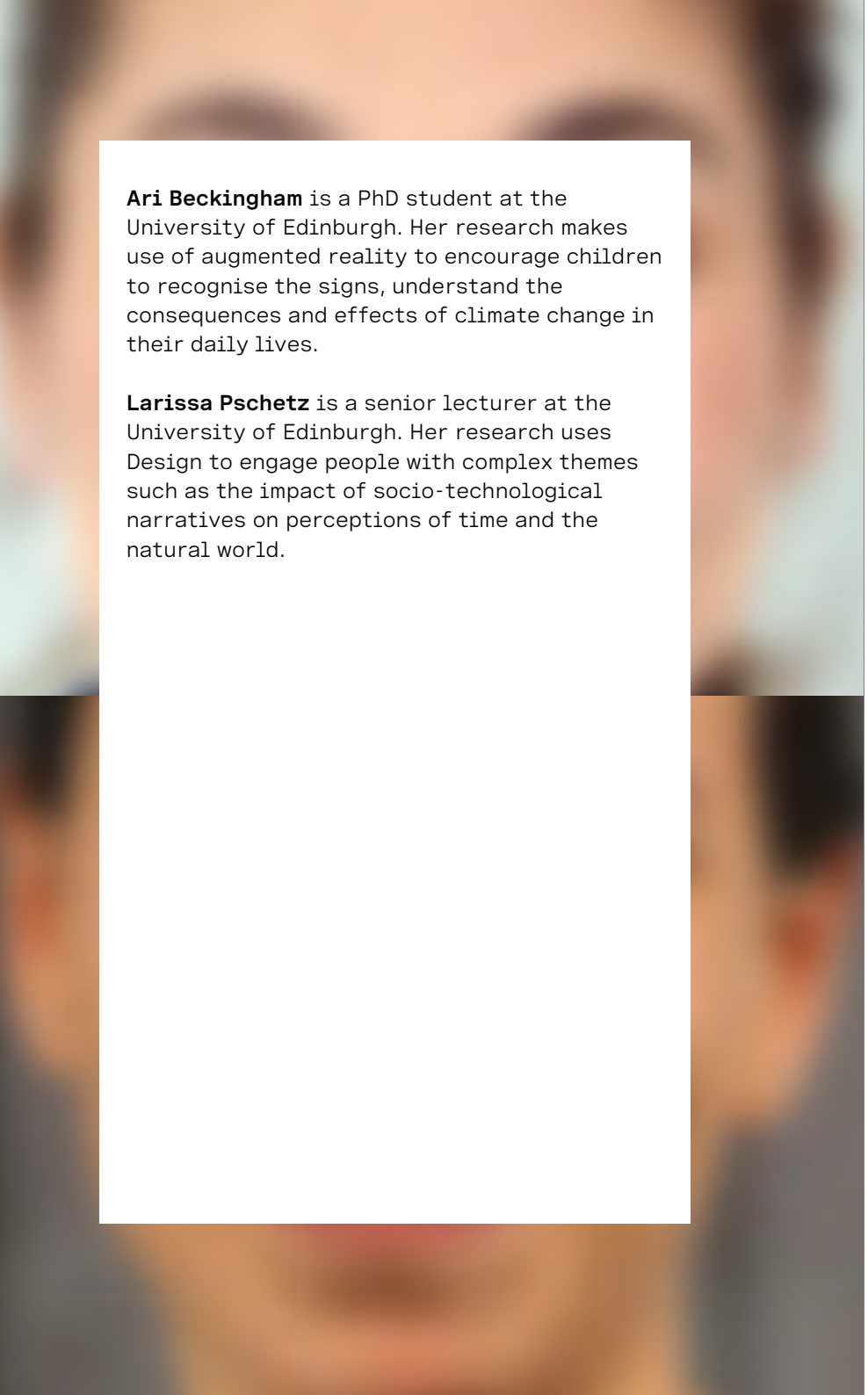
UNESCO. (2021). *Reimagining our futures together: A new social contract for education*

UN (United Nations) General Assembly. (2021). *Resolution adopted by the Human Rights Council on 7 October 2021. 48/4. Right to privacy in the digital age*

van der Hof, S., Lievens, E., Milkaite, I., Verdoodt, V., Hannema, T., & Liefaard, T. (2020). The child's right to protection against economic exploitation in the digital world. *The International Journal of Children's Rights, 28*(4), 833-859

Wachter, S. (2020). Affinity profiling and discrimination by association in online behavioural advertising. *Berkeley Technology Law Journal, 35*(2)

Williamson, B. (2019) Psychodata, 7 October



Ari Beckingham is a PhD student at the University of Edinburgh. Her research makes use of augmented reality to encourage children to recognise the signs, understand the consequences and effects of climate change in their daily lives.

Larissa Pschetz is a senior lecturer at the University of Edinburgh. Her research uses Design to engage people with complex themes such as the impact of socio-technological narratives on perceptions of time and the natural world.

Rethinking pace, context and modes of learning in EdTech design

Ari Beckingham and Larissa Pschetz,
University of Edinburgh

The coronavirus (COVID-19) pandemic created an urgent demand for digital technologies to support children's remote learning and socialisation needs. With lockdowns and school closures, many children started to adopt digital platforms and applications to carry out school, social and extracurricular activities. Formal learning gained new informal and unstructured characteristics, as it started to take place across multiple settings and relied more on (sometimes also revealing the lack of) support from parents (including caregivers) and other groups beyond the school.

However, most technologies adopted for learning during the pandemic were not designed for such a purpose. For example, video conferencing platforms (e.g., Zoom and Microsoft Teams) were designed to support adult engagement and productivity, often with office-based environments, routines and etiquette in mind. They tended to be context-agnostic, treating physical environments as either irrelevant or a distraction to the learning activity, rather than an important aspect of learning. Importantly, these technologies assumed an awareness of practices and implications of data recording and sharing, which children (and many adults) indeed often lack.

Here, we argue that designers could start rethinking

educational tools for remote learning by considering different paces and modes of engagement, connection to different contexts and physical and social environments, as well as increased awareness of data processes and practices. We further discuss how design considerations, combined with emerging technologies such as augmented reality (AR), could help create a renewed agenda for EdTech (educational technologies) design. The intended benefit of rethinking learning tools in this way is to expand the breadth of learning tools and possibilities, without compromising on transparency and ethics of data processing. However, the possibility of improved learning outcomes to result from our suggested design considerations, especially in relation to the use of AR, remains difficult to measure.

Considering more nuanced approaches to pace and engagement

A key aspect of technology design is a recurrent focus on productivity (Hallnäs & Redström, 2001; Odom et al., 2012), which usually translates into keeping people engaged, focusing attention towards accomplishment of specific tasks, and optimisation of time spent on these tasks. While such assumptions can be problematic in general (Pschetz & Bastian, 2018), when applied to children's technologies, they can be hazardous. Addiction to technology usage among children and teenagers is a common concern in the literature (Hawi et al., 2019), with studies suggesting links between greater time spent online and reported feelings of loneliness as well as lack of physical activity and disrupted sleep (Nalwa & Anand, 2003). Coban (2020) discusses that even well-meaning EdTech, such as video-animated storybooks, which help keep students engaged in stories and where they can practise reading with or without adult supervision, can lead to a slippery slope of dependency on technology, especially with young children who are at a vulnerable stage of development.

As reported by the 5Rights Foundation (2021a), there is a culture among companies of generating revenue by maximising the retention, reach and activity of children, which may sacrifice their safety and wellbeing. As such, designers have been designing interactions, interfaces and content for 'more time, more people, more activity' (5Rights Foundation, 2021b), leaving children feeling that they are spending too much time on their devices, that they have too much exposure and/or feel too much pressure to get engagement on their profiles.

Some designers have been advocating for a change of practice, for example by exploring whether less engagement could translate into longer term relationships. Challenging assumptions of success in the industry is a way of treating the symptom of a wider issue of following the narratives of success in design as a whole. This would open up space for exploring different qualities of engagement and new modes of interaction (e.g., by exploring boredom, including attention breaks, and supporting connection to the environment).

A different approach to engagement is exemplified by the PlayStation games Flow (2008) and Journey (2012), which invite players to slow down while exploring a scenic virtual landscape. Exploration is the main aspect of the game, and players interact with each other through a quiet mode of communication - when they meet in the game's virtual environment, they can only communicate by emitting bird-like tweets, rather than through words or text. These games invite slowing down and offer a meditative experience rather than focusing on task completion, competition and active communication.

The consideration of more nuanced notions of pace and engagement allows for incorporation of multiple aspects of children's experience, which can be a starting point for creating radically novel interfaces and learning experiences, while providing alternatives for exploitative models of technological development.

Rethinking context

During the COVID-19 pandemic, technologies widely adopted in formal education focused on keeping communication channels open. In this context, children's physical environment was either disregarded or treated as background noise to the mediated interaction. While supporting communication is important, children could indeed benefit from technologies that take their social and physical environment into account.

Research shows that learning can be improved when

applied across different locations, representations and activities, and immersive technologies can have a key role in promoting retention, communication and engagement (Luckin et al., 2012). According to research by Zimmerman and Bell (2012), learners' performance and competency faltered when in a formal setting such as school, over an informal setting such as out-of-school programmes. Mediated cross-setting learning could therefore improve education, beyond the need to accommodate for self-isolation. It could support capturing, storing, comparison and integration of data from several places and contexts, while also keeping groups in and out of different processes through multiple ways of sharing and communicating. Furthermore, due to the accessibility of devices such as mobile phones, and students' immediate access to capturing, storing and managing data (e.g., images, audio and video), children have the opportunity to view their learning material from a variety of different perspectives (Furió et al., 2015).

The introduction of context, however, cannot be taken lightly. The environments children are in can make or break learning, and research has shown that remote learning can contribute to reinforcing differences between the well-supported and ill-supported (Engzell et al., 2020). Understanding these factors is important to designing technologies that reduce rather than increase differences. While modes of parental engagement in mediating access to learning technologies has been extensively discussed in the design and human-computer interaction (HCI) literature (Yu et al., 2021), the pandemic demonstrated that shifting the focus from schools to caregivers may lead to greater inequality in learning. Particularly due to different levels of engagement and availability from parents or caregivers (Anzani et al., 2020), basic things such as setting up an account for a service can exclude children who cannot engage caregivers for permission (Keaton & Gilbert, 2020).

Thus designers need to consider the importance of multiple stakeholders in learning, including communication between teachers, students, parents and/or other stakeholders. One such effort in this direction has been made by ClassDojo, an application for primary education that aims to build links between the school and the home - mainly through updates provided through a variety of formats that are available to students, parents and teachers. The ambition could be extended to considering other stakeholders such as Scouts, charities and community groups that could not only extend children's learning to other settings, but also extend their support network.

Instead of treating children's environment as a distraction, designers could look for ways to account and reduce differences in support, allowing for connection and potential recreation of environments - for example through tasks that involve exploration of particular natural settings or through use of immersive technologies such as AR.

Changing responses to pervasive data processing

Systems adopted for children's remote education further assume an awareness of implications of data recording and sharing, which students - and indeed, many adults - often lack (von Struensee, 2021). Online communities and group chats, for instance, are often not monitored for the very young, and can risk exposing them to inappropriate content - for example profiles with 'child age' restrictions can be faced with extreme diet cultures or even self-harm content (eSafety Commissioner, 2019).

The need to minimise these risks is often interpreted as a need to restrict technology usage - for example there are guidelines for reframing technology for young learners, in particular by the American Academy of Paediatrics and the National Association for the Education of Young Children, that look at preventing children from being overexposed to digital devices and decrease potential technology addiction. Indeed, designers need to consider the implications of the technologies they design at every step of the process, such as when considering format, technology, interaction, context, remit, etc.

However, simply restricting usage is not the answer to minimise risks, as it doesn't increase understanding of what data is and how it can be used (for good or bad). In the HCI literature, there have been several discussions as to how parents and caregivers can take a role in mediating children's

interaction with technology (Yu et al., 2021), but these assume a level of parental involvement that children may lack. Instead, an attitude to supporting data literacy would look at supporting data literacy education, which, according to Ridsdale et al. (2015) includes (a) supporting diverse and creative learning approaches that make effective use of technology; (b) iterative learning of data-related issues with complementary skills integrated (such as in project-based learning); (c) an emphasis on the mechanics of data integration as well as concepts; and (d) increasing engagement with content by using real-world data.

Much of digital technologies' data processing takes place in the background of users' awareness, and while it is often taken for granted that adults are aware of these practices (when accepting terms and conditions) - even though many are not - the same assumptions become problematic when considering child users. With a general lack of transparency of data processes, assumptions become charged with concerns and often fears that platforms would constantly track users' data on behalf of businesses, potentially leaking, or selling to third parties (Pschetz et al., 2017), which is problematic for adults but could present increased risks to children.

Instead of parents and companies simply restricting children's access to technology, more transparent data processes, adequate ways to present terms of services or request consent (e.g., including different stakeholders and exploring other media beyond the legal contract), and extending the curricula to approach critical issues around data literacy could allow children to nurture a more positive attitude to data that could be transformational in their lives and for society as a whole.

Changing pace, exploring settings and increasing data awareness through AR

In our ongoing research, we are looking at ways to support cross-setting learning and informed data practices through apps that can help children understand environmental issues. We focus on issues of climate change as a way to connect children to a pressing issue, and to cultivate what Anna Tsing (2015) calls 'the art of noticing'. Although effects of changes in climate are manifested in many ways around us, noticing them requires stepping out of narratives of productivity and time saving to connect to the environment around us. Thus, we ask how children can engage with the changes that selected species (trees, insects, birds and mammals) experience as a result of climate heating. By engaging in, producing and contributing to recording data on biological phenomena like blooming dates in relation to climatic conditions (and therefore 'real-world' data), we invite children to engage in the multiple temporal patterns of nature and understand the delicate balance between species' temporalities, which allow for brief encounters through which species remake themselves and adapt to a changing environment.

Through the production of phenology records, children are invited to understand what a single data point represents in a larger context. Using data-capturing apps, children begin to understand how to produce records and how these can be interpreted in a larger context. While it has been shown that children have great ability to interpret climate change-related information (Eide & Kunelius, 2021); we aim to support them in translating a broad subject into tangible accounts, and to further allow them to see themselves as active participants of a database for the public good.

These approaches to pace and data practices are brought together through the use of AR - a real-time experience where one's physical world is enhanced with a layer of computer-generated information (Carmigniani et al, 2011). Since its inception in the late 1990s, AR has been extensively explored in education, with research stressing potentials and limitations alike. AR can have a key role in promoting retention, communication and engagement, by allowing children to apply learning across contexts and to deepen it through the addition of new layers, for example by making invisible aspects of an ecosystem such as pollution levels visible, or allowing inspection of every detail of a small insect.† AR has also been shown to increase confidence as it enables multisensory learning and allows students to learn by doing (Lu & Liu, 2015),

† A great example of this is the Smithsonian's AR app that allows the user to view replicas of popular exhibits in their own surroundings (Smithsonian, 2020).

and due to its ability to simulate real-life situations, it also has the ability to meaningfully engage the user (Fan et al., 2020). AR is also seen as a big industry that is likely to generate innovation, particularly in the education sector. UKRI (UK Research and Innovation) predicts that, by 2024, the immersive technologies industry will be worth £101 billion (Chitty, 2022), and in a 2020 survey by Perkins Coie and the XR Association, respondents named 'education' as the second most likely sector to adopt immersive technologies (Dick et al., 2021).

Nevertheless, the usage and benefits of the technologies are still dependent on teachers and students' technical skills, knowledge or familiarity with the software, and due to its mode of knowledge construction and reception, many of the benefits discussed in the context of learning - specifically, formal education - are difficult to measure because they cannot be compared with current curriculum targets. Furthermore, it presents challenges both in terms of education and data practices. Educational challenges include the potential of AR to: (a) distract students from learning aspects, if not properly used (Chiang et al, 2014); (b) be difficult to use by students and teachers alike, particularly without a well-designed interface or guidance (Munoz-Cristobal et al., 2014), (c) be content inflexible, which does not allow teachers to incorporate or connect it to their lessons (Fan et al., 2020), and (d) present challenges for inclusion of large groups (Furió et al., 2013). In fact, many educational institutes have not been keen on pursuing immersive technologies in their classroom due to available budget, existing ICT (information and communications technology) infrastructure (hardware, software and internet) and limited time to train teachers (Weerakanto, 2019).

In terms of data practices, AR apps are often part of technology ecosystems that derive value from gathering data from users, in ways that can present increased risks to children. For example, when investigating the implication of AR video games such as Pokémon, Das et al. (2017) found that the real-time location tracking functions increased threats of physical harm as well as posed risks to mental health through potential risk of addiction. Additionally, decisions to use AR in the classroom are mainly driven by the potential of these technologies to support teaching (McKnight et al., 2016), and include little reflection on how data will be processed as children interact with such systems. Design considerations, such as whether the AR experience will be consumed individually or collaboratively, in an informal environment, for example home, or in a formal environment such as school, and whether a parent or guardian will be present to supervise or provide context for the experience, could help address the problem of data infrastructure when designing education AR experiences.

Through our work we advocate that designers consider three factors in their initial design stage. First, consider 'how', by thinking of solutions to offer help and provide easy access to update content by teachers or parents and caregivers; previous researchers (Fan et al., 2020) noted a lack of flexibility with AR content as a disadvantage. Second, consider 'where', and specifically designing strategies to support collaborative learning, for example will the student be using the AR application at school with the teacher, superimposing it on to objects or sharing the device with a teacher or guardian to find information (Sytwu & Wang, 2016)? Third, consider 'who' - allow teachers and learners to explore or choose between various contexts such as learning style, groups, for example age, and learning contexts, such as individual or collaborative.

Conclusion

This essay presents the potential of cross-setting data-driven platforms to enhance children's learning, and discusses design considerations for these new ways of delivering education through AR. We draw attention to the need for:

- Integration of nuanced notions of pace and engagement to enrich learning experience and provide alternatives to exploitative models of technological development

- Experimentation with approaches to account for and reduce differences in learning support, particularly at home, connecting multiple

stakeholders and allowing exploration (and potential recreation) of real-world environments

- Experimentation with approaches to explain data processing, terms of services and acquisition of consent in a more meaningful, accessible, and age-appropriate manner while supporting collective decision-making concerning technology uses between parents and children

- Context and child-centric consideration and design of immersive technologies, such as AR, as a means for enriching children's learning experience.

We do not claim that cross-setting data-driven platforms could account for children's learning needs. Instead, we provide design considerations to inspire alternative approaches to EdTech design, combining understanding of formal subjects with informal explorations of the natural world that are underlined by critical data capture, and reinforced through AR explorations.

5Rights Foundation. (2021a). Pathways: How digital design puts children at risk

5Rights Foundation. (2021b). *Playful by Design® Free play in a digital world*

Anzani, D. R. A., Zaeni, I. A., Nuqul, F. L., & Mualifah, M. (2020). Relationship between parents' education level and parental engagement in the pandemic period of Covid-19. International Webinar Series - Educational Revolution in Post Covid Era, pp. 30-38

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications, 51*, 341-377. doi:10.1007/s11042-010-0660-6

Chiang, T. H., Yang, S. J., & Hwang, G. J. (2014). Students' online interactive patterns in augmented reality-based inquiry activities. *Computers & Education, 78*, 97-108.

Chitty, A. (2022). *Tackling challenges, building prosperity: The Industrial Strategy Challenge Fund*. UKRI

Coban, A. (2020). On pandemics, technology, and early childhood education: An opinion piece. *Childhood Education, 96*(6), 66-69. doi:10.1080/00094056.2020.1846393

Das, P., Zhu, M. O., McLaughlin, L., Bilgrami, Z., & Milanaik, R. L. (2017). Augmented reality video games: New possibilities and implications for children and adolescents. *Multimodal Technologies and Interaction, 1*(2), 8.

Dick, E. (2021). *The promise of immersive learning: Augmented and virtual reality's potential in education*. Information Technology and Innovation Foundation

eSafety Commissioner. (2019). *Safety by Design overview*

Eide, E., & Kunelius, R. (2021). Voices of a generation: The communicative power of youth activism. *Climatic Change, 169*(6)

Engzell, P., Frey, A., & Verhagen, M. D. (2020). Learning inequality during the COVID-19 pandemic. *SocArXiv*

Fan, M., Antle, A. N., & Warren, J. L. (2020). Augmented reality for early language learning: A systematic review of augmented reality application design, instructional strategies, and evaluation outcomes. *Journal of Educational Computing Research, 58*(6), 1059-1100

Furió, D., Juan, M. C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: A comparative study. *Journal of Computer Assisted Learning, 31*(3), 189-201

Furió, D., González-Gancedo, S., Juan, M. C., Seguí, I., & Costa, M. (2013). The effects of the size and weight of a mobile device on an educational game. *Computers & Education, 64*, 24-41

Hallnäs, L., & Redström, J. (2001). Slow technology - Designing for reflection. *Personal and Ubiquitous Computing, 5*(3), 201-212

Hawi, N. S., Samaha, M., & Griffiths, M. D. (2019). The digital addiction scale for children: Development and validation. *Cyberpsychology, Behaviour, and Social Networking, 22*(12), 771-778

Keaton, W., & Gilbert, A. (2020). Successful online learning: What does learner interaction with peers, instructors and parents look like? *Journal of Online Learning Research, 6*(2), 129-154

Lu, S. J., & Liu, Y. C. (2015). Integrating augmented reality technology to enhance children's learning in marine education. *Environmental Education Research*, 21(4), 525-541

Luckin, R., Bligh, B., Manches, A., Ainsworth, S., Crook, C., & Noss, R. (2012). Decoding learning: *The proof, promise and potential of digital education*

McKnight, K., O'Malley, K., Ruzic, R., Horsley, M. K., Franey, J. J., & Bassett, K. (2016). Teaching in a digital age: How educators use technology to improve student learning. *Journal of Research on Technology in Education, 48*(3), 194-211

Munoz-Cristobal, J. A., Jorrin-Abellan, I. M., Asensio-Perez, J. I., Martinez-Mones, A., Prieto, L. P., & Dimitriadis, Y. (2014). Supporting teacher orchestration in ubiquitous learning environments: A study in primary education. *IEEE Transactions on Learning Technologies, 8*(1), 83-97

Nalwa, K., & Anand, A. P. (2003). Internet addiction in students: A cause of concern. *Cyberpsychology & Behaviour, 6*(6), 653-656

Odom, W., Banks, R., Durrant, A., Kirk, D., & Pierce, J. (2012). Slow technology: Critical reflection and future directions. In *Proceedings of the Designing Interactive Systems Conference* (pp. 816-817)

Pschetz, L., & Bastian, M. (2018). Temporal design: Rethinking time in design. *Design Studies, 56*, 169-184

Pschetz, L., Tallyn, E., Gianni, R., & Speed, C. (2017). Bitbarista: Exploring perceptions of data transactions in the Internet of Things. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2964-2975

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*. Dalhousie University

Smithsonian. (2020). *Voyager + AR*

Sytwu, T. A., & Wang, C. H. (2016). An investigation of the effects of individual differences on mobile-based augmented reality English vocabulary learning. In D. Churchill, J. Lu, T. Chiu, & B. Fox (Eds.), *Mobile learning design: Lecture notes in educational technology* (pp. 387-410)

Tsing, A. L. (2015). *The mushroom at the end of the world*. Princeton University Press

von Struensee, S. (2021). Eye on developments in artificial intelligence and children's rights: Artificial Intelligence in Education (AIEd), EdTech, surveillance, and harmful content. *SSRN Electronic Journal*. doi:10.2139/ssrn.3882296

Weerakanto, P. (2019). Digital literacies of English language teachers and students and their perceptions of technology-enhanced language learning and teaching in Thailand. Doctoral dissertation, The University of Arizona

Yu, J., Granados, J., Hayden, R., & Roque, R. (2021). Parental facilitation of young children's technology-based learning experiences from nondominant groups during the COVID-19 pandemic. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 1-27

Zimmerman, H. T., & Bell, P. (2012). Everyday expertise: Learning within and across formal and informal settings. *Theoretical Foundations of Learning Environments, 2*, 224-241

RETHINKING DATA FUTURES

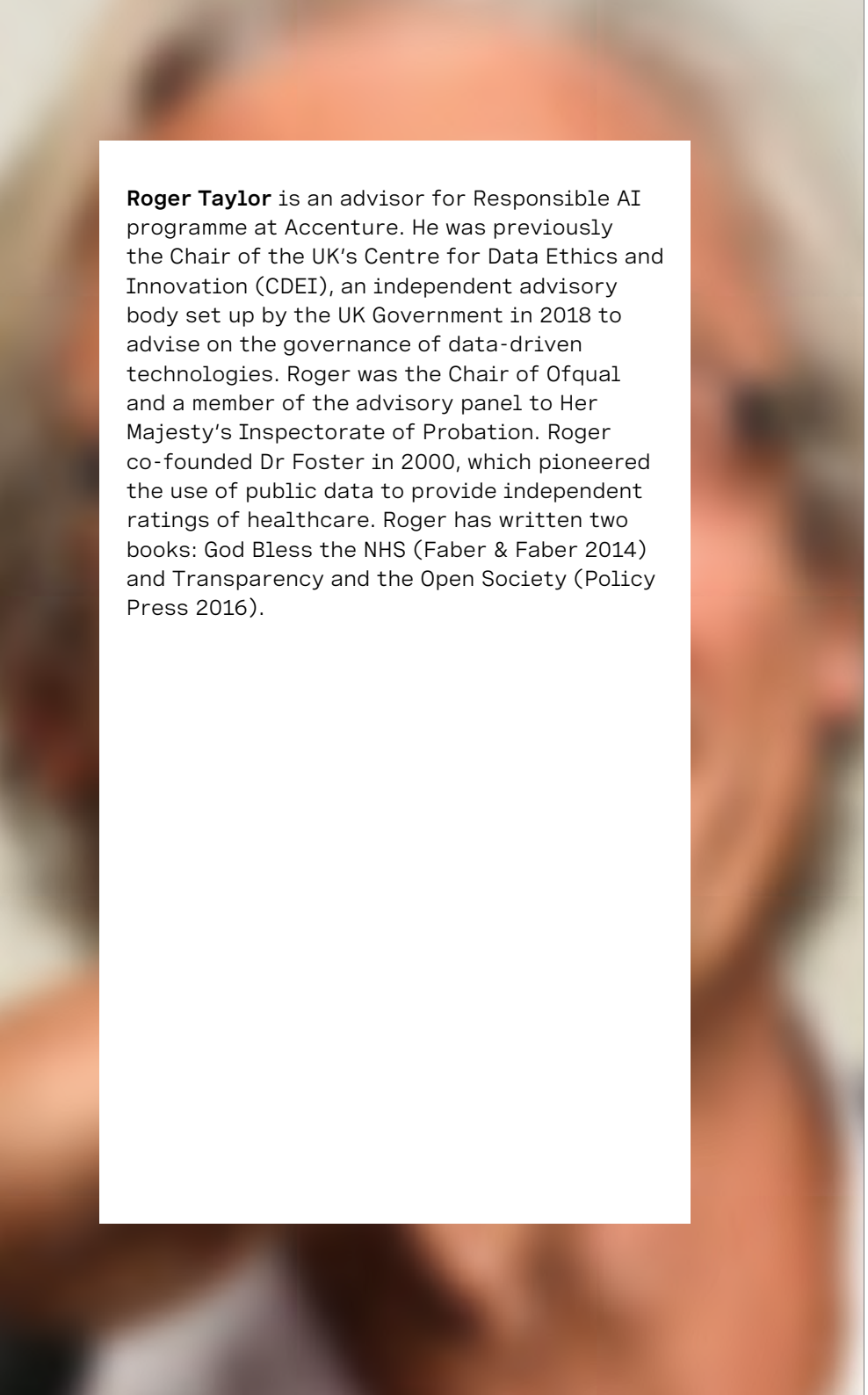
Seesaw helps because I know what I've done. It doesn't help because there is really no point (Girl, 10)

It's cool to do homework on ClassDojo and I like the photos they share of what we have been doing in school (Girl, 9)

Show My Homework lets me and my parents know what homework I have outstanding and the ones I have done (Boy, 15)

I don't like to use MyMaths because I like being taught by a teacher face to face and do not like homework (Boy, 16)

Digital Futures Commission (2022) EdTech survey



Roger Taylor is an advisor for Responsible AI programme at Accenture. He was previously the Chair of the UK's Centre for Data Ethics and Innovation (CDEI), an independent advisory body set up by the UK Government in 2018 to advise on the governance of data-driven technologies. Roger was the Chair of Ofqual and a member of the advisory panel to Her Majesty's Inspectorate of Probation. Roger co-founded Dr Foster in 2000, which pioneered the use of public data to provide independent ratings of healthcare. Roger has written two books: God Bless the NHS (Faber & Faber 2014) and Transparency and the Open Society (Policy Press 2016).

New approaches to data stewardship in education

Roger Taylor, Open Data Partners

At the heart of problems created by the digital economy lies control of data. Harms arise when an organisation has exclusive control over data about people, but has interests that diverge from theirs. Proprietary control over data, of the sort exercised by large platforms, also hinders innovation. It allows companies to achieve market dominance and to crush competition. Changing the way that data is controlled is a way to enable innovation while also providing a level of protection that current rights-based data regulation is unable to secure.

The current data protection regime is unable to provide adequate protection. Individual consent does not provide a way to understand how data about me is used. To truly understand this, I would need to know the impact of that data used, compared to other people. The regulatory focus on purpose - an important and valuable protection - is of limited use in today's data-driven world where people want personalised digital services but are at risk of being manipulated or discriminated against. My problem is not stopping people using my data for purposes I disapprove of; my problem is stopping people using my data for purposes I approve of, but doing it in a way that is ultimately damaging.

Look at education, for example. The use of data to drive

decisions and tools in education is potentially of enormous benefit, but it could also cause serious harm. The problem is not deciding whether to say 'yes' or 'no' to a particular purpose; it is knowing how we can protect ourselves from the poor use of data for purposes that are theoretically beneficial but prove, in practice, to be harmful.

This is why many technologists advocate a much more radical solution - breaking up the inherently monopolistic characteristics of digital markets by separating control over data from the provision of data-driven services. This would allow for unconflicted data stewardship organisations to monitor how data is being used, represent the interests of data subjects, prevent misuse of data and ensure appropriate levels of competition.

This is the principle behind the idea of data trusts, which has attracted many champions from technology industries because it offers a uniquely compelling vision of how to address problems in digital markets.

One barrier to the adoption of these approaches is the belief that data protection regulation can solve a problem it is not designed to solve (see Figure 1). We should be sceptical of regulatory solutions that rely on data protection regulation and do not take full account of how this approach could be applied in practice.

An example: Google Classroom

Google provides equipment and digital services free to schools, such as email or apps for setting and receiving assignments. Google says the data collected is used only to monitor and improve these services. Children can, with parental consent for under-13s, also use the Chrome browser or Google maps. If they use Chrome they will see adverts, but Google says no data from children is used to personalise these adverts.

Google's suite of education products has been criticised from many angles (e.g., Krutka et al., 2021). In a feature in Fast Company last year, it was accused of disguising its business model, 'making it almost impossible to ascertain what data it collected about students and what Google uses it for' (Williamson, 2021).

Last year New Mexico started legal proceedings against Google, claiming it was illegally tracking the online behaviour of children under 13. The case was initially dismissed, but New Mexico appealed and Google settled. They admitted no breach of the law but agreed to do more to police age-screening on the app store and to fund an education initiative in the state.

The claim from New Mexico was that Google had collected information without getting clear consent and concealing its activity. The complaint accused Google of 'infiltrating' schools; of claiming its product was free when in fact it 'comes at a very real cost which Google purposefully disguises'. It said the company was 'mining children's data' for commercial benefit.

After the settlement New Mexico Attorney General Hector Balderas said: 'There are incredible risks lurking online and we should do everything we can to protect the privacy of children'.[1] This is true. However, it is not clear that his actions have had any significant impact on these risks. Google may not be breaking the law, but it has not become any easier 'to ascertain what data it collected about students and what Google uses it for'. To the extent, if at all, that Google has been infiltrating schools and imposing a 'very real cost' on children, nothing of significance has changed. One commentator called it 'fundamentally a victory' for Google, and pointed out that a new Google-branded education institute in New Mexico was a win for the business (Gold, 2021).

While Google is no doubt gathering a large amount of data, which it will use for commercial benefit, and which may harm people, the problem is that data protection is an ineffective tool to combat this risk. Attacking Google for unauthorised data use or inadequate consent misses the target. Even with all the necessary consent and legal authorities in place, the risk that the data is used in a way that harms young people remains. The question is not whether the company has the legal authority to use the data; it is whether it is doing so in a way that is harmful or beneficial.

For example, companies such as Google will typically establish a legal basis that allows them to use data to improve their service. It makes little sense to object to this in principle, but 'improving services' could mean something relatively innocuous, such as designing better ways to present email. It might also mean using the data to build artificial intelligence

(AI) that reads children's essays, monitors the speed with which they write, sees what time in the evening they do their homework and starts to build an understanding that could inform recommendations to teachers based on highly personal profiling. The second of these might be an enormously beneficial thing to do, but equally, it might be extremely damaging. As things stand, we have little way of knowing whether Google's work to improve its products is innocuous, brilliant or destructive.

The need for a new standard of practice in data driven systems

This type of market failure is not new. There are many products where the market would not work without quite specific regulations regarding information. Medicines are one. You cannot tell whether a pill works by looking at it. And it is not safe to find out by trying it. Instead, we have an elaborate regulatory mechanism that sets standards for how information is generated and shared to assess a product's efficacy. Cars and airlines are similar, in that you are safe to choose a car on the basis of its shape or an airline on the basis of its food because regulation does the work of ensuring the wrong choice is unlikely to cause serious harm.

AI and other complex data-driven systems are similar in that the quality of the product or service can only be assessed with knowledge and quite specific datasets.

Data-driven systems present two additional challenges. First, it is difficult to tell in advance where the dangers might be. No one imagined that using machine learning to build recommendation systems in social media would help unleash a pandemic of misinformation. Second, we are not talking about one class of products - it is a fundamental technology that is altering a wide range of products and services, introducing new risks to all of them.

Data protection law was developed to control the purposes for which data is used, and is grounded in the concerns that arose during the initial development of databases. We face very different risks today that cannot be effectively addressed with data protection law.

What would a new approach to data governance look like?

A number of different elements have a role in reshaping digital markets, but the key ideas are:

- Separation of the data layer from the application layer in the architecture of digital services
- Independent governance and control of the data layer by organisations that are legally excluded from providing apps and services and that have duties towards data subjects (e.g., data trusts)
- Protection of individual data rights through personal data stores (or rights to data portability and reporting).

These ideas are independent and there are examples of each. For example, Open Banking in the UK is a mechanism to enforce portability of financial data;[2] the medical research field has a number of 'data governance' bodies that oversee access to data (e.g., HDR UK); and in the commercial area, shared data pools such as 'Skywise' allow companies involved in building Airbus aeroplanes to manage access to a shared pool of data.[3]

However, the biggest opportunities lie in bringing these ideas together and applying them to the provision of personal digital services. Together they create a virtuous circle that can support a market for demonstrably beneficial innovation. They allow for decentralised management of digital IDs and create the space for a market of 'digital agents' who represent the interests of individuals and communities, enabling individuals to maintain control over how data about them is used while at the same time empowering organisations capable of turning these rights into effective market or regulatory power.

Separating the data from the application creates a market incentive to drive the adoption of data standards to the extent that services and apps make use of common underlying data. This allows for greater competition in the provision of these services. It also allows for external experts or regulators to develop the skills to interrogate and interpret the data that are equal to those of the organisations providing services.

Establishing separate governance for the data layer means

that data users have to make the case for their use of data to an organisation that is its equal, both in terms of its ability to control data and to understand how data is being used.

The detail of how these mechanisms are best applied in any particular area depends on the context. However, if these arrangements were applied in education, we can imagine a scenario in which a personal data store would hold a defined set of information about the pupil, including the data generated by the school. The school would then operate a data trust on standard terms with pupils or parents and caregivers. Such trusts might be federated across similar schools and operated by an independent trustee body. The terms would set down not just the purposes to which data can be put, but also the way in which the impact of data use is assessed and the mechanisms by which data subjects are kept informed and able to exercise choice.

The trust would also set out the terms on which providers of digital education services could access data. For example, the trust might set requirements for data must be returned to the data layer (e.g., activities, test scores) in either standard or proprietary formats. Such arrangements might require, for example, that any assessment of bias or benefit would be based on data held in the independent data layer, not on the provider's own data systems.

In effect, this mechanism replaces regulation with market incentives. This can then ensure the appropriate level of resource going into these activities - activities that are value-creating for society and the economy, but that would likely be prohibitive if framed as a regulatory requirement.

This approach would end the pretence that individual consent is enabling people to exercise meaningful control over data use. Instead, agents acting on behalf of parents and caregivers, children and schools would have the powers and capabilities necessary to protect their interests.

Such an arrangement would also afford greater freedom to providers of digital services to innovate and improve services, without increasing risks of data misuse.

Why is there limited progress towards reforming digital markets?

Progress towards creating this new world is not due to lack of enthusiasm or hard work. For many years leaders in the technology industry have been calling for root-and-branch reform of the data economy and working to achieve it.

In 2021 Tim Berners-Lee launched Inrupt, a company that builds on the work of the SOLID data standard for personal data stores, recognising that we need to rebuild the data economy from the ground up. In the UK, Professor Irene Ng has a similar initiative, HatDex, which enables individuals to require their data to be held in a separate database that they own.

The Open Data Institute has championed the use of data trusts to create a new layer of governance over data use.[4] Neil Lawrence, former Director of Machine Learning at Amazon and now DeepMind Professor of Machine Learning at the University of Cambridge, has established the Data Trust Initiative to support the implementation of such arrangements (Gardner, 2020; see also Delacroix & Lawrence, 2019). The Mozilla foundation has also been active in encouraging new approaches to data management.

Despite this, these ideas receive insufficient attention in discussions about regulating digital services, whether in education or in any other area.

Defend Digital Me, which campaigns on the use of data in education, has made detailed recommendations to prevent abuse of data, but does not address the need for wholesale reform of the relationship between control over data and provision of data-driven services (Defend Digital Me, 2020).

The UK Government, which advocates a strongly pro-innovation stance towards data (DCMS, 2021), has been very clear in setting out how it intends to reform data protection to remove regulatory barriers to innovation. In comparison, its comments on data stewardship lack detail and substance. The consultation proposals for reform of the General Data Protection Regulation (GDPR) had extensive analysis of problems with current regulatory arrangements, but little to say about new approaches to data governance.

There are several things that can account for this. The first is that the problem is hard. It is hard for policymakers to get

their heads around the many difficult questions any implementation of these new arrangements raises. For example, to what extent is it necessary to impose a minimum level of custodianship on a market, or should this be something that individuals or market participants can opt in to? The second option is more appealing to governments because it requires less action, but is also less likely to succeed.

If a minimum standard is imposed, what are the legal mechanisms that are best suited to doing this? What institutions are required to oversee this? How far would its powers extend in setting standards and/or requiring data sharing with end service providers or intermediate data agency /data trust services?

A second problem is that the answers to these questions are very context-dependent. They would vary, for example, according to whether the service under discussion is safety-critical, highly regulated, state-provided, foundational (e.g., identity) or an entirely optional consumer service.

It is easier for governments to set out frameworks and overarching mechanisms. The EU is implementing exactly this sort of approach through its Data Governance Act that establishes the basis on which data-sharing mechanisms might operate.† The UK has similarly been exploring 'enabling' frameworks to allow for such mechanisms to exist.

However, the market failure that makes new forms of data stewardship necessary is the same market failure, which means that simply 'enabling' solutions to exist will be insufficient. The role of government here is not to enable, but to deliver.

The last problem for governments is uncertainty. The complexity of the situation, and the range of possible solutions, means that there is no way to reliably and comprehensively design such a complex set of arrangements in advance. This is one of a class of problems in which the solution can only be identified by first making a commitment to put new arrangements in place, and then working through the issues with stakeholders. Such situations are not unusual in life, but they are never comfortable for governments for whom the appearance of control is so vital. The necessary engagement from interested parties will not be available without a commitment to implementation and funds to support the work. The only way to make progress is to recognise that reform is essential, decide where to implement reforms and have the political courage to commit to seeing it through. 'Commitment' here would mean establishing a competent legal authority to oversee reforms, giving it an appropriate budget, and setting a principles-based framework within which to operate along with target dates and reporting requirements.

This can be challenging and off-putting to government that may lack the skills and knowledge to feel comfortable about the risks. Elected representatives currently face no compelling reason to wade into such difficult waters. These policy ideas do not offer quick solutions. They require long-term strategic planning and significant investment to build digital services for the next generation that are trustworthy and beneficial. Politicians are happier leaving it as a nice idea and offering warm words of encouragement.

Applying new data stewardship models in education

If a government were to demonstrate the necessary vision and courage, education is an interesting and promising area where intervention to reshape the digital economy could bring significant benefits. The market is not one that has been staked and claimed - in the sense that there are no dominant education-specific services that generate their value from the proprietary exploitation of the knowledge contained within mass data collection. There is significant potential benefit from the use of AI and data-driven technologies if done right, and there are significant risks in leaving it to current market arrangements. There is widespread acceptance that government has a key role in assuring the quality of education, including digital education services.

Crucially in education there is a credible route to success. Moving to a new model is much easier if there is an 'on ramp' of deliverable benefits that start at a low level and build as a system develops. In education, relatively simple steps - such as giving people digital certificates for their qualifications -

† This is not a criticism of the EU as it would be difficult for an overarching body such as the EU to do more than this. The criticism is of national governments that could and should do more.

provide both a useful service (you do not have to find paper certificates for job applications) and creates the basis for the establishment of new data stewardship arrangements based on personal data stores.

The tools we need to move beyond the current debate about data protection and instead initiate a discussion about reform of the data economy are available to us. However, it needs the catalyst of political pressure and political will to change the way in which this market operates. Without it, we will not be able to deliver a safe, innovative, digitally supported education system.

Figure 1: GDPR, data protection and harmful data-driven services

A key assertion in this essay is that data protection laws as currently constructed cannot offer adequate consumer protection. This claim could be supported by examining the extent of its impact on the behaviour of the large data platforms. However, given limited space, it is simpler to ground in some more fundamental observations about the mechanisms of data protection (for a much longer discussion of these issues, see Taylor & Kelsey, 2016).

The core principle of data protection law (and a good principle, too) is that data should not be processed without lawful grounds. This gives consumers and regulators the power to 'pull the plug' and halt data processing. Consumers can do this by withholding consent, regulators by rejecting the legal basis of processing.

Although data privacy advocates recommend 'pulling the plug', this recommendation cannot protect consumers for the following reasons:

1. *People want personalised services*. If people were willing to do without, then the power to pull the plug would fix the problem. However, the majority of people in the UK, USA and Germany are in favour of the use of personalisation in a wide range of applications including recommendation systems and advertising (Kozyreva et al., 2021). People in the UK are also in favour of using personalisation to make recommendations in education by identifying educational needs (CDEI, 2020).

2. *Regulators do not have the powers and capabilities necessary to identify and address harmful personalisation*. People need protection, not from personalisation, but from personalisation that is biased, manipulative or harmful in other ways. You cannot achieve this by saying 'no' - partly for the obvious reason that rejecting things does not force people to give you what you want - but more importantly, because you first need to be able to

identify whether or not a particular use of data is harmful. This is rarely immediately apparent. For example, to identify bias you need to look at how a system treated large groups of people and compare it to similar systems.

3. *The scale of the task makes it implausible that a single regulator could not have the powers and capabilities necessary to identify and address harmful personalisation*. Europe has proposed a new AI law to address the problem of harmful decision-making. It does this by creating an obligation on users of AI systems to demonstrate they have a system in place to manage risks. This recognises the purely logistical problem of having a single regulator attempting to oversee the fairness of data-driven decisions in health, education, finance and life in general. That is an unfeasibly large task, and the issues concerned are, in many cases, covered by other existing legal obligations.

4. *The best way to protect the consumer is to ensure that someone other than the provider of a data-driven service has the capability, the power and responsibility to assess whether it is beneficial or harmful*. To assess the fairness, accuracy and/or harmfulness of data-driven systems requires an ability to compare between systems and to look beyond the immediate data on which the system runs (which, in the main, will confirm the vendor's view of the system). The harm that can come from misuse of data in education is that it hinders education or negatively effects children in ways that are not obvious to those using such systems. An assessment of whether something is harmful will depend crucially on the ability to compare it to alternative approaches to education and understand the impact in the wider context. The creation of data agencies, data trusts or regulators that manage shared data pools within key industries provides just such a mechanism.

CDEI (Centre for Data Ethics and Innovation). (2020). CDEI review of online targeting

Day, E. (2021). Governance of data for children's learning in UK state schools. Digital Futures Commission, 5Rights Foundation

DCMS (Department for Digital, Culture, Media & Sport). (2021). Data: A new direction. 10 September

defenddigitalme, (2020). The state of data 2020: Mapping a child's digital footprint in the state education landscape in England

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance. International Data Privacy Law, 9(4), 236-252

Gardner, R. (2020). New Data Trusts Initiative will spearhead community-focused data governance. Department of Computer Science and Technology, University of Cambridge, 21 October

Gold, A. (2021). Google settles children's privacy suits brought by New Mexico. Axios, 13 December

Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. Humanities and Social Sciences Communications, 8, 117

Krutka, D. G., Smits, R. M., & Willhelm, T. A. (2021). Don't be evil: Should we use Google in schools? TechTrends, 65, 421-431

Livingstone, S., Atabey, A., & Pothong, K. (2021). Addressing the problems and realising the benefits of processing children's education data: Report on an expert roundtable. Digital Futures Commission, 5Rights Foundation.

Taylor, R., & Kelsey, T. (2016). Transparency and the open society. Policy Press.

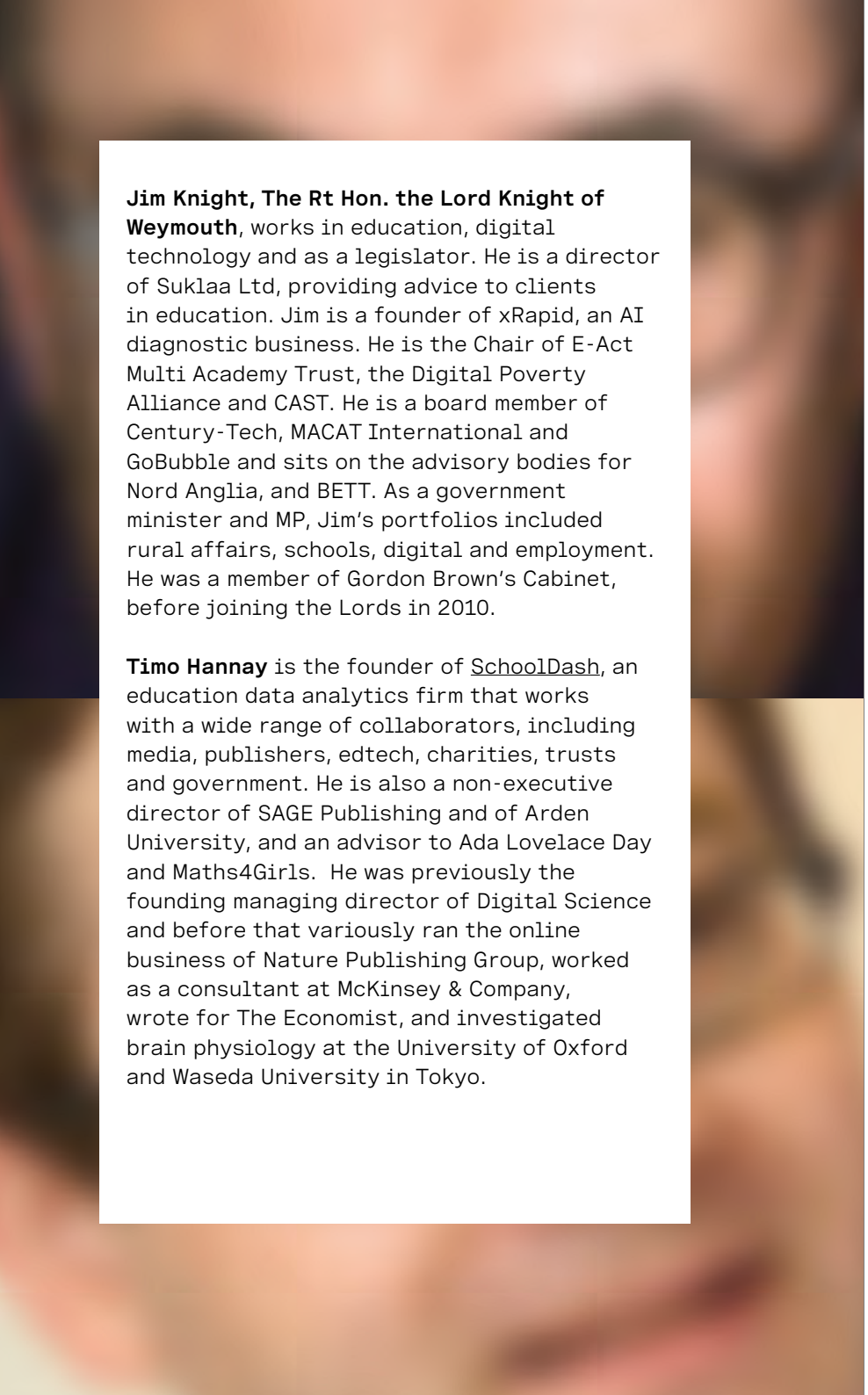
Williamson, B. (2021) Google's plans to bring AI to education make its dominance in classrooms more alarming. Fast Company, 28 June.

1 New Mexico press release: https://www.nmag.gov/attorney-general-hector-balderas-announces-landmark-settlements-with-google-over-childrens-online-privacy

2 https://openbanking.org.uk

3 https://aircraft.airbus.com/en/services/enhance/skywise

4 https://theodi.org/project/data-trusts



Jim Knight, The Rt Hon. the Lord Knight of Weymouth, works in education, digital technology and as a legislator. He is a director of Suklaa Ltd, providing advice to clients in education. Jim is a founder of xRapid, an AI diagnostic business. He is the Chair of E-Act Multi Academy Trust, the Digital Poverty Alliance and CAST. He is a board member of Century-Tech, MACAT International and GoBubble and sits on the advisory bodies for Nord Anglia, and BETT. As a government minister and MP, Jim's portfolios included rural affairs, schools, digital and employment. He was a member of Gordon Brown's Cabinet, before joining the Lords in 2010.

Timo Hannay is the founder of SchoolDash, an education data analytics firm that works with a wide range of collaborators, including media, publishers, edtech, charities, trusts and government. He is also a non-executive director of SAGE Publishing and of Arden University, and an advisor to Ada Lovelace Day and Maths4Girls. He was previously the founding managing director of Digital Science and before that variously ran the online business of Nature Publishing Group, worked as a consultant at McKinsey & Company, wrote for The Economist, and investigated brain physiology at the University of Oxford and Waseda University in Tokyo.

Trust in data, and data in trust

Jim Knight, Labour Peer and Timo Hannay, SchoolDash

Information is the currency of our age. Like more traditional forms of value, it can be used either to alleviate or exacerbate social ills. We believe that greater sharing and analysis of data is essential to more fully understand and address shortcomings in education. This must be done responsibly, benefiting the education system as a whole. One promising approach towards achieving this aim is the emerging concept of data trusts: legal entities that provide independent stewardship of data. This essay explores their potential in the context of education, particularly in the wake of the COVID-19 pandemic.

The risks of data processing and usage

In the online world, it is both the best and the worst of times. The internet has enabled access to information, opportunity and human connection in a way that was previously inconceivable.

Data is being generated, harvested and analysed at a scale that is transforming our economies and societies. To social scientists and policymakers, data provides a uniquely powerful observational tool - akin to the telescope for astronomers or the microscope for biologists. We no longer need to interview

1000 people and extrapolate; we can analyse the actual behaviour of whole populations in real time. In education, too, the adoption of online and artificial intelligence (AI)-driven approaches to learning has surged, especially in the wake of the COVID-19 pandemic. Indeed, we were perhaps lucky that it struck at a time when such alternatives to classroom-based teaching were even possible.

Yet any excitement at these remarkable and genuinely impressive developments must be tempered by deep concern about all sorts of adverse consequences.

Two years ago, the World Health Organization (WHO) declared that, alongside the COVID-19 pandemic, it was also fighting an 'infodemic' of misinformation (WHO, 2020). As we write, the ongoing war in Ukraine is being fought not only on land with guns and tanks, but also in cyberspace, with novel forms of propaganda and military intelligence. More than ever before, social media platforms have become geopolitical players and are struggling to tame the viral, runaway nature of their own networks and algorithms (Bushwick, 2022). It seems that every week brings news of yet another data leak or ransomware attack (Page, 2022). In the UK, the 5% of people who remain offline cite worries about privacy, identity theft and misuse of personal data as the most common reasons for foregoing the benefits of these new technologies (Lloyds Bank, 2021). The rise in online harms is prompting governments to seek to regulate large swathes of the internet through mechanisms such as the Online Safety Bill.

Data in education

Like many other domains, education generates vast quantities of data, most of it held by private organisations, including educational technology (EdTech) firms, test publishers, tuition providers and survey companies. So far, these proprietary datasets have been mostly invisible and largely unexamined, even to the organisations that hold them.

The pandemic has started to change these attitudes to education data.

Datasets have grown even bigger, driven by the move to online learning. Their usefulness has become much more apparent, due to an urgent need to understand the effects of lockdown and school closures on children's learning and wellbeing.

For example, the Education Policy Institute (EPI), an independent charity, was commissioned by the Department for Education (DfE) to analyse children's academic progress (or lack thereof) using data from Renaissance Learning, a commercial test provider (EPI, 2021). As schools closed during lockdown and then reopened, many of them used these tests to assess their pupils' individual progress and learning needs. By aggregating data from all such schools, the EPI was able to publish a series of studies during the course of the pandemic that characterised and quantified the national picture with respect to pupils' lost learning relative to previous cohorts. This stimulated the national debate and informed administrative decisions at every level, from central government to individual schools. It helped to answer such important questions as: exactly how much are pupils underperforming? How does this vary by age or geography? Which subjects and topics have been most affected?

On their own, such analyses do not solve problems, or even guarantee consensus about how to do so (indeed, there was considerable disagreement on this between the UK government and its own education recovery commissioner, leading to his eventual resignation; see Coughlan & Sellgren, 2021). But they were widely reported and discussed, and served the vital role of grounding the debate in objective reality rather than anecdote and preconception.

SchoolDash, a data analytics firm founded by one of us (TH), has been conducting similar work with another test publisher, Hodder Education (Hannay, 2021a). As well as providing alternative perspectives on the problem (since each dataset lent itself to slightly different analyses), a broad consistency between the EPI's results and those of SchoolDash helped to inspire confidence that the lost learning was real. In addition, SchoolDash has analysed data from EduKit (Hannay, 2020), a wellbeing survey company used by schools, and a number of online learning providers, including Oak National Academy (Hannay, 2021b), a government-funded initiative established during the pandemic to provide free online video lessons. These provided insights into how young people were

coping with home-based learning, both psychologically, in terms of social and emotional contentment, and practically, in terms of being able to access online lessons and engage effectively with the material provided.

These initiatives were largely products of COVID-19 and the associated lockdowns, when timely information about children's activities became especially important and official statistics were either too slow or, in some cases, absent altogether. But the potential of such analyses to help us understand and improve education goes well beyond the unique circumstances of the pandemic. If we can use such data to understand attainment gaps, technology divides and wellbeing deficits during lockdown, then why not during more normal times too? Educational inequalities and shortcomings are perennial challenges, and the mission to reduce them continues to deserve all the insight we can muster.

The risks in education data uses vs. public trust

These developments are exciting - but potentially scary too. We know from the activities of big tech companies and totalitarian regimes that unfettered use of personal information can have bad effects, whether intended or not. Perhaps understandably, trust in governments and technology companies to use information responsibly is in decline in the UK (Wisniewski, 2020), the USA (PAC, 2021) and elsewhere. How can we enjoy the collective benefits while minimising the risk that important data might be used to serve narrow commercial or political interests rather than the interests of learners and the common good? This is particularly concerning when the data refers to individual people, and all the more so when those people are children.

The EU's General Data Protection Regulation (GDPR) has created a regime that reduces potential harm while maintaining many of the benefits that come from gathering and analysing personal data. However, it has not resulted in an increase in public trust (Wisniewski, 2020). Furthermore, insofar as this is based on user consent, it is difficult for parents or children (or anyone) to give consent for use that may be highly technical or somewhat uncertain in terms of outcome, as is often the case in analyses of subjects such as educational disparities.

The Age Appropriate Design Code (AADC) further protects children's online activity in the UK, and this will soon be followed by an Online Safety Bill. These will help private and public sector technology providers to build on well-defined minimum standards. But additional, more flexible solutions are required to truly minimise risks and support the positive use of potentially sensitive information. It is abundantly clear that self-regulation by technology companies is not enough on its own. What other approaches might help?

Many kinds of organisation already use independent oversight to protect wider interests: schools have governors, charities have trustees and companies have non-executive directors. As data acquires increasing personal and societal relevance, perhaps it deserves similar safeguards too. This is the central idea behind 'data trusts', a relatively new concept that builds on existing data rights and trust law to provide independent fiduciary stewardship of data (ODI, 2018).

The potential of data trusts in building trust in data

In October 2017, the UK Government published an independent review, *Growing the artificial intelligence industry in the UK* (Hall & Pesenti, 2017). It called for the 'development of data trusts, to improve trust and ease around sharing data'. Since then, interest and activity around this idea has steadily increased, with organisations such as the Open Data Institute (2018), the Alan Turing Institute (2019), the Ada Lovelace Institute (2021), Nesta (Mulgan & Straub, 2019) and the Data Trusts Initiative[1] (a collaboration between the Universities of Cambridge and Birmingham), all exploring and promoting the concept. They are now being talked about in commercial contexts, for example to represent users of a particular service, and their application in sensitive domains like AI (Mehonic, 2018) and healthcare (Milne et al., 2020) is being actively pursued.

Data trusts can take a wide variety of forms (O'Hara, 2019). A common model is for the trustees to represent the interests of a well-defined group of people, such as a local community or a cohort that is the subject of a research project[2] (somewhat akin to an academic ethical review board). We believe that they

can also be useful in representing the wider interests of, say, the education system as a whole, and even society at large.

With this in mind, we are collaborating with other individuals and organisations to establish an education data trust[3] that will aggregate information across multiple proprietary data holders and public sources, using these to conduct analyses that will provide actionable insights for participants at all levels of the education system. This builds on the work conducted before and during the pandemic described above, and will help to answer questions such as: how well have children caught up on learning following the reopening of schools? What are the current demographic and socioeconomic disparities in attainment? How does access to technology vary by pupil age or location? How big a role is online education playing, and how is this evolving over time? These are just a few examples from an almost endless list of possibilities. Sounder, more timely answers to these questions would support better-informed public debate and more effective education policies, ultimately to the benefit of children.

Crucially, all of these activities will be overseen by a board of independent, knowledgeable and highly regarded trustees, who will be tasked with representing the interests of the data subjects and the wider education system. Unless they approve of a particular activity, it will not be allowed to happen. This will obviously constrain our ability to conduct whatever analyses we (and the data providers) choose to conduct. But, far from holding us back, we expect this to be an enabler. Done properly, a rigorous and transparent approach to data oversight will instil greater confidence in our activities, leading to more, not fewer, opportunities to access new sources of information and use them to derive valuable insights. It will initially focus on schools in England and other parts of the UK, but with the potential to extend into further or higher education, and into other territories. We also hope that it will serve as a template for those who would like to establish similar initiatives in other domains, just as we have been inspired by emerging projects elsewhere (see, for example, the Data Trust Initiative's 2022 pilot projects[4]).

Data trusts are not a substitute for legislation, public regulation or even industry self-regulation, but they are an important addition to the mix. Neither are they a fool-proof, catch-all solution, just as the existence of a governing body does not guarantee the proper running of a school. But oversight by a group of independent, knowledgeable and credible trustees who are answerable not to senior executives or shareholders but to data subjects and wider society surely represents a step in the right direction. Indeed, not to make such a move towards better stewardship and greater transparency is to invite further scepticism towards data analysis of any kind, decreasing its potential in solving real-world problems and letting go a huge opportunity to pursue the common good.

We expect that there will be further applications of data trusts that go beyond the kinds of analyses described here. The use of AI in education is a particularly pressing area. It offers huge opportunities by helping teachers to better match pedagogy and resources to individual learners, and to do so far more efficiently than has ever been possible before. But it requires vast quantities of training data, which, in turn, raises ethical questions.

An instructive example is the use by Google's DeepMind of the health records of 1.5 million patients at an NHS Trust (BBC News, 2021). The company used these to train an AI system that detects people at risk of kidney injury. Beyond the concerns of the Information Commissioner around data privacy, there was also controversy over the intellectual and commercial rights, since NHS patients had shared data with a public service provider only to see it being used to create a commercial product that was then sold back to the NHS at taxpayers' expense.

Similar conflicts of interest could arise in schools. For example, if a company uses one product to collect learner data and then applies it in developing a second product, how should the resulting value be shared between the company and the school, and who is looking out for the interests of individual learners? There are no easy or universal answers to such questions, which makes it all the more important to establish systems of oversight that take adequate account of the full range of interests involved.

Towards responsible data uses

Despite increased regulation, trust in technology is in decline and this backlash will continue to constrain the adoption of new, potentially beneficial innovations. While not a panacea, data trusts represent an important part of the solution. We anticipate that they will become a standard means to ensure responsible use of data across education and beyond, especially where the information concerns children or other potentially vulnerable groups. What Creative Commons has done for content sharing, and Wikipedia has done for knowledge dissemination, data trusts might yet do for online information. By harnessing the power of openness and collaboration, we hope that they will help to support the internet as a force for individual and collective good.

Ada Lovelace Institute & AI Council. (2021). Exploring legal mechanisms for data stewardship. 4 March

Alan Turing Institute, The (2019). Data trusts workshop

BBC News. (2021). DeepMind faces legal action over NHS data use. 1 October

Bushwick, S. (2022). Russia's information war is being waged on social media platforms. *Scientific American*, 8 March

Coughlan, S., & Sellgren, K. (2021). School catch-up tsar resigns over lack of funding. BBC News, 2 June

EPI (Education Policy Institute). (2021). *EPI research for the Department for Education on pupil learning loss*. 29 October

Hall, W., & Pesenti, J. (2017). *Growing the artificial intelligence industry in the UK*. Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy

Hannay, T. (2020). The lockdown experiences of pupils in England. SchoolDash, 27 August

Hannay, T. (2021a). The effects of educational disruption on primary school attainment in summer 2021. SchoolDash, 31 August

Hannay, T. (2021b). What Oak National Academy usage tells us about education during the pandemic. SchoolDash, 5 November

Lloyds Bank. (2021). *UK Consumer Digital Index 2021*

Mehonic, A. (2018). Can data trusts be the backbone of our future AI ecosystem? The Alan Turing Institute, 3 October

Milne, R., Sorbie, A., & Dixon-Woods, M. (2020). What can data trusts for health research learn from participatory governance in biobanks? *Journal of Medical Ethics, 48*(5)

Mulgan, G., & Straub, V. (2019). The new ecosystem of trust. Nesta, 21 February

ODI (Open Data Institute). (2018). What is a data trust? 10 July

O'Hara, K. (2019). *Data trusts: Ethics, architecture and governance for trustworthy data stewardship*. WSI White Paper # 1, February. Web Science Institute

PAC (Public Affairs Council). (2021). *2021 Public Affairs Pulse Survey report*

Page, C. (2022). Thousands of Nvidia employee passwords leak online as hackers' ransom deadline looms. TechCrunch, 4 March

WHO (World Health Organization). (2020). Munich Security Conference Speech. WHO Director-General, 15 February

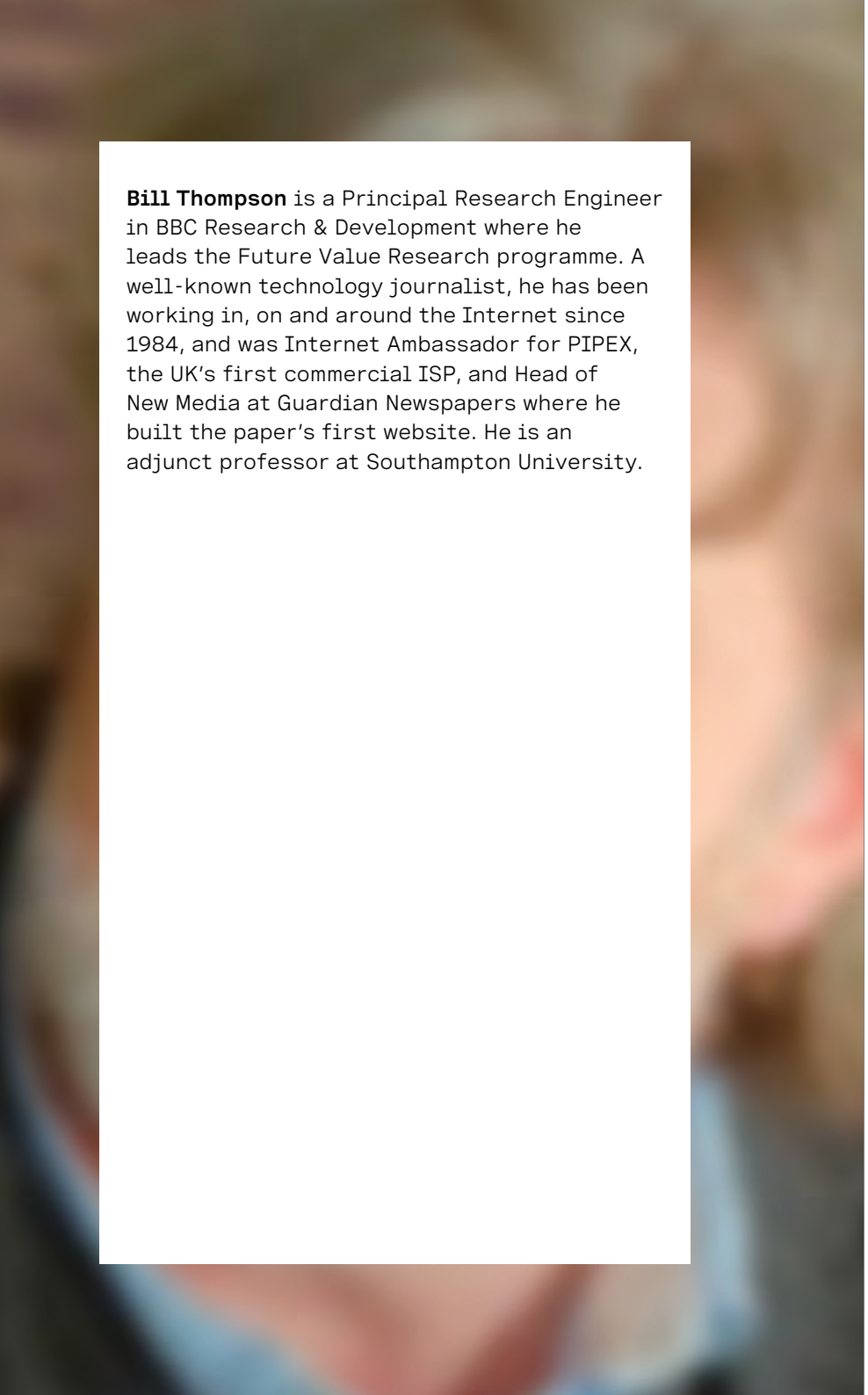
Wisniewski, G. (2020). Losing faith: The UK's faltering trust in tech. Edelman, 30 January

1 See: https://datatrusts.uk

2 For more information, visit: https://datatrusts.uk/pilot-projects-1

3 More information about a project concerning education data trust will soon be available at: https://educationdatatrust.com

4 For more information, visit: https://datatrusts.uk/pilot-projects-1



Bill Thompson is a Principal Research Engineer in BBC Research & Development where he leads the Future Value Research programme. A well-known technology journalist, he has been working in, on and around the Internet since 1984, and was Internet Ambassador for PIPEX, the UK's first commercial ISP, and Head of New Media at Guardian Newspapers where he built the paper's first website. He is an adjunct professor at Southampton University.

Data from cradle to grave: How personal data stores could transform the uses of data about children and young people

Bill Thompson, BBC Research & Development

An age of watchers

The growth of 'surveillance capitalism', a term popularised by Shoshana Zuboff in her book *The age of surveillance capitalism* (2019) and now widely adopted, should not surprise us. Computer systems, from the earliest mainframes to the modern pocket-sized networked supercomputers we still call 'phones', have always maintained records of how they are being used, in order to manage performance, monitor security and, where relevant, charge for resources.

Keeping records - or 'logging' - is what computers do, and from the very earliest days of online publishing on the website owners were promised that the ability to track visitors was one of the great advantages. But we have come a long way from using web server logs to let us know how many people read a web page and for roughly how long, and now vast amounts of data are being collected about every computer-related transaction.

The extent of logging and the ways the records are being used is the result of a set of choices and imperatives that have largely been driven by the increased commercialisation of the online environment since the mid-1990s (Naughton, 2012). This complex, expensive, flawed and intrusive system provides more

and more data points around any interaction a user may have with a networked system, processing that data in order to make inferences about them that can be used to drive advertisements, shape online experiences, or even alert the state to activity that it deems noteworthy or dangerous (Zuboff, 2019).

And we are not at the end of this process. As technology develops, so does tracking. With the emergence of virtual reality (VR) and augmented reality (AR) systems and the push to develop shared, persistent, online spaces that allow for property ownership, as multiplayer games fuse into what has been called the 'metaverse', this will only get more extreme. Unless we do something to avoid it, we can look forward to an age of 'omniscient capitalism' (Pesce, 2019), in which everyone who straps on a headset or puts on a pair of smart glasses becomes part of a virtual panopticon, with consequent risks to privacy.

We need to ask ourselves what can be done about this, particularly with regard to data about children and young people.

Data about children and young people

In the ongoing debate about the balance of interests between the technology companies that want to monitor users to support their business models, governments that want to track their citizens, and individuals who want ways to preserve their privacy and exert control over their data (European Parliament, 2022), it is generally recognised that there are special considerations concerning children and young people.

Some of these are legal, relating to their ability to give informed consent, while there are also issues around how much we want the online activities of children and young people to be monitored, profiled and used to present ads, shape what they see on social media, or even influence their life trajectory.

In some cases, they create the problem for themselves. Many young people have found a way to work around the age limits of services like TikTok, Snapchat and Facebook by lying on registration forms, or had accounts created by compliant parents or caregivers so that they are not 'missing out'. They may even be using mobile phones registered to adults, which means that personal data is collected from them as if they are over 18, clicking through consent screens without having the legal authority to accept the terms (Ofcom, 2022).

Even when someone's age is apparent there can be problems, and discussions about an appropriate regulatory environment have started to consider areas where data about children may be slipping through the gaps. In the UK the recent Digital Futures Commission report on data-driven education systems, *Governance of data for children's learning in UK state schools* (Day, 2021), prompted in part by the shift from in-person to online teaching during the COVID-19 pandemic, pointed out just how regulatory uncertainties and common business practices around excessive data protection and retention had created an environment where children's education data was largely uncontrolled, compromising the many potential benefits that data processing could offer.

This issue is not limited to education data. Recently in the USA the Federal Trade Commission (FTC) ordered WW (formerly 'Weight Watchers') to delete a dataset and the training model derived from it because it had been illegally acquired from young people (Brody, 2022). This should not surprise us, as many aspects of young people's lives are now mediated by electronic systems that play such an important part in our lives, and so they have become full members of Zuboff's 'surveillance society'.

Looking beyond regulation

Whatever regulations are in place, data management is fundamentally a technical issue, involving the collection, storage and processing of computerised records. At the moment the standard model for organisations that want to use data about people is for them to set up a central database or user activity store, often linked to user accounts with some form of validation or login capability. Keeping this data secure is a challenge, and there are significant reputational, regulatory and financial risks, as well as the costs of storage and service provision. It is also very easy for data to be used in ways that go beyond the original purposes, with potential privacy implications.

Away from this standard approach there is much

experimentation and innovation around data management. One of the most promising alternatives to the monolithic model is the personal data store, or PDS. A PDS is a storage system in which data can be securely stored, using a range of encryption and other technologies. Some are 'data vaults', like Solid[1] or Mydex[2]. Others like Databox[3] are complete computing environments where data is both stored and processed without leaving the secure area, with only the results made available to third parties. The owner of a PDS can store their personal data and control which systems have access to it, changing their mind at any time.

The PDS is an enabling technology that is capable of supporting a wide range of business models, from open source to fully commercial, with a range of tools to facilitate the integration of third-party apps (Bolychevsky & Worthington, 2018), but the technology is mature enough for the Flanders government to have started a project to give every citizen a Solid PDS to hold citizen data (Berners-Lee & Bruce, 2021).

The PDS is an element in a broader effort to rethink data processing around the trusted processing of data, and a number of organisations, including the BBC, have developed a model they call the 'public service data ecosystem' (PSDE), 'a set of components which work together to provide a secure and effective platform for public service applications, and which are able to integrate personal data with open data, aggregate data, and data from sources such as Internet of Things devices' (Sharp et al., 2021).

The idea of a data ecosystem came from an ongoing discussion over the future of data governance more generally, and detailed exploration of new legal models for data stewardship, where the control of the data about people used by an organisation is managed by a data trust or cooperative that balances the interests of the individuals and the organisation (Ada Lovelace Institute, 2021).

At the heart of the public service data ecosystem is a personal data store that provides a user-centred approach to the storage and access of both legally defined 'personal data' and other data the user might want to control independent of any particular application. Within the public service data ecosystem each user can have one or more data stores, which

can be located in the cloud or on their own hardware, and third parties cannot copy or perform any processing of the data without the user's explicit permission.

The ecosystem also incorporates other data sources, whether open data or licensed, that can be used in combination with the data stored in the PDS to support a range of services. These could include finance, health and entertainment applications, all of which benefit from the additional control over data use that the PDS provides.

PDS-based systems are already widely used. The CitizenMe app is used to support a model called 'zero-party data', holding user data within their app, while allowing it to be used in controlled and transparent ways (Deakins, 2022). Mydex is working with the Scottish Government to provide identity verification services by storing authorised credentials (Mydex, 2020).

In 2021 BBC Research & Development (R&D) published its work with the Solid system to develop a cross-media recommender that used combined transaction data from multiple media services including the BBC iPlayer to create a user profile that could improve the quality of recommendations from those services, without the need to share the data across services (Sharp, 2021; Sharp et al., 2021).

Delivering the potential of the PDS

While we have been talking about PDS in some form for well over a decade, the market has so far failed to deliver on their potential. A study on PDS conducted at the Cambridge Judge Business School (University of Cambridge) in 2015 noted that:

> As an innovative concept, the personal data store faces significant obstacles to widespread diffusion. In particular, PDS providers must reach critical mass in the context of a double-sided market: the PDS system must attract a sufficient number of individuals and businesses if it is to flourish as a platform for data exchange, but neither individuals nor businesses are easily captured without the other first in place. (Brochot et al., 2015)

Things now seem to be changing, partly as a result of legal rulings including the FTC ruling against WW referred to earlier, and notably the Belgian Data Protection Authority's recent decision to fine the online advertiser IAB Europe over its transparency and consent framework (Bryant, 2022), as well as new regulations like the UK Age Appropriate Design Code (AADC). All of these have raised awareness of how data can be abused, and the risks organisations take when they store data.

At the same time, research into attitudes to personal data use has shown that people dislike the current approach in which commercial organisations control their personal data, preferring approaches that give them control over their data that include oversight from regulatory bodies or that enable them to opt out of data gathering (Hartman et al., 2020).

And we now have a population, including young people, which is more familiar with security practices for online services and smartphones, while advertising companies like Apple have raised awareness both of surveillance and how to counter it. Apple advertising makes much of the fact that their photos app processes your images on your phone while Google sends your data to its cloud.

Young people and personal data

The combination of a shifting regulatory environment, increased consumer awareness and technical maturity creates an opportunity to propose PDS-based approaches to the management of the data needed to deliver services to children and young people, offering a large enough market to make it worth investing in, and a compelling use case that can drive providers towards the technology.

A PDS-based approach would let service providers, especially in the educational technology (EdTech) market, offer advanced functionality to schools while protecting user data, and young people could then be encouraged to use their PDS for other purposes, perhaps encouraging social media platforms to offer children-oriented services using the same technology. It does not matter who provides the PDS as long as a service can make use of it through a standard interface with a suitable data model.

Within the broader context of the public service data ecosystem, other data sources, including open data, could be safely combined with data held in a PDS to support a range of services - for example advanced profiling and recommendations, as discussed by BBC R&D (Sharp et al., 2021). Furthermore, the data would remain under the control of the individual and would not have to be deleted when they left school or added to the data lakes maintained by the companies providing services. And, of course, a PDS-based model for EdTech would also be useful for adult learners, so it could be that many would choose to retain their PDS as they leave formal education, putting pressure on platforms and others to adapt to the new model.

Remaining issues

A public service data ecosystem based around PDS does not solve all of the issues around the use and abuse of data about people. There are still consent issues to be dealt with when it comes to children and young people, both over the initial provision of a PDS to someone under 18 and to their agreement to allow data to be accessed by services.

Nor can it deal with the 'selling a kidney' problem, where an individual who has control of all their personal data decides to give access - or a copy - to an organisation without really considering the implications, or on the basis of misleading offers.

There are also, inevitably, going to be security breaches, bugs in code and other ways in which there could be data breaches. As always, we will need a strong regulatory framework, and proper enforcement of penalties against those who abuse data about people, to reinforce the technical provisions. However, an approach to the public service data ecosystem based on personal data stores seems to be healthier and more likely to encourage people to understand and consider the consequences of their actions.

The benefits of control

Perhaps the biggest benefit will be increased transparency, because every application and service that wants to create and use data about a user via a PDS will have to be very clear about what it is storing, and when it wants to use that data.

There are parallels between the current debate concerning the ways we manage data about people, and the ongoing arguments about accessibility of websites and games. Early websites and games were generally designed for users who required no special accommodations (other than vision correction, because somehow, not being able to focus is not classed as a disability). Making the web more accessible and adding accessibility features to games was presented as limiting the creative expression of designers, and was resisted for many years.

There is a similar debate about data processing, where companies argue that regulation or technical limitations inhibit creativity. Yet, as with accessible websites and games, alternative approaches that put respect for people at the centre of their design can still deliver business objectives, perhaps more easily and certainly with less regulatory peril.

Computerised systems will remain important in all aspects of our lives, including in educational settings, and storing and processing data about users is a vital aspect of their operation that cannot be avoided. However, it is possible to develop approaches to data management that allow data about people to be controlled by them while still being available where they are needed to serve legitimate interests. The PDS model offers individuals a significant degree of control over their data without compromising the functionality of the systems, and merits serious consideration as an alternative to large-scale databases.

A world within which children and young people are encouraged to think about their personal data and where and how it is stored and used, with PDS as one option, may be one in which they grow up into adult data citizens instead of being seen as data 'subjects'. In order for this to happen we need further research to explore the capabilities of the technology, and a regulatory environment that can accommodate it as one option for the storage and processing of personal data.

Ada Lovelace Institute. (2021). Exploring legal mechanisms for data stewardship

Berners-Lee, T., & Bruce, J. (2021). Committing to a digital innovation economy in Flanders, built on Solid

Bolychevsky, I., & Worthington, S. (2018). Are personal data stores about to become the NEXT BIG THING? *Medium*, 4 October

Brochot, G., Brunini, J., Eisma, F., Larsen, R., & Lewis, D. (2015). Study on personal data stores. *Digital Single Market*, 7 August

Brody, B. (2022). Weight Watchers must delete algorithms built from kids' data. *Protocol*, 4 March

Bryant, J. (2022). Belgian DPA fines IAB Europe 250K euros over consent framework GDPR violations. Iapp.org, 2 February

Databox Project. (2017). EPSRC Databox Project

Day, E. (2021). *Governance of data for children's learning in UK state schools*. Digital Futures Commission & 5Rights Foundation

Deakins, S. (2022). The future of data is 'zero data'. *CitizenMe*, 15 June

European Parliament. (2022). Deal on Digital Markets Act: Ensuring fair competition and more choice for users. News

Hartman, T., Kennedy, H., Steedman, R., & Jones, R. (2020). Public perceptions of good data management: Findings from a UK-based survey. *Big Data & Society, 7*(1), 205395172093561

Mydex. (2020). *Smart entitlements: Recommendations and report for the Scottish government*

Naughton, J. (2012). *From Gutenberg to Zuckerberg: What you really need to know about the internet*. Quercus

Ofcom. (2022). Children and parents: Media use and attitudes report 2022

Pesce, M. (2019). *Augmented reality: Unboxing tech's next big thing*. Polity

Sharp, E. (2021). Personal data stores: Building and trialling trusted data services. BBC R&D

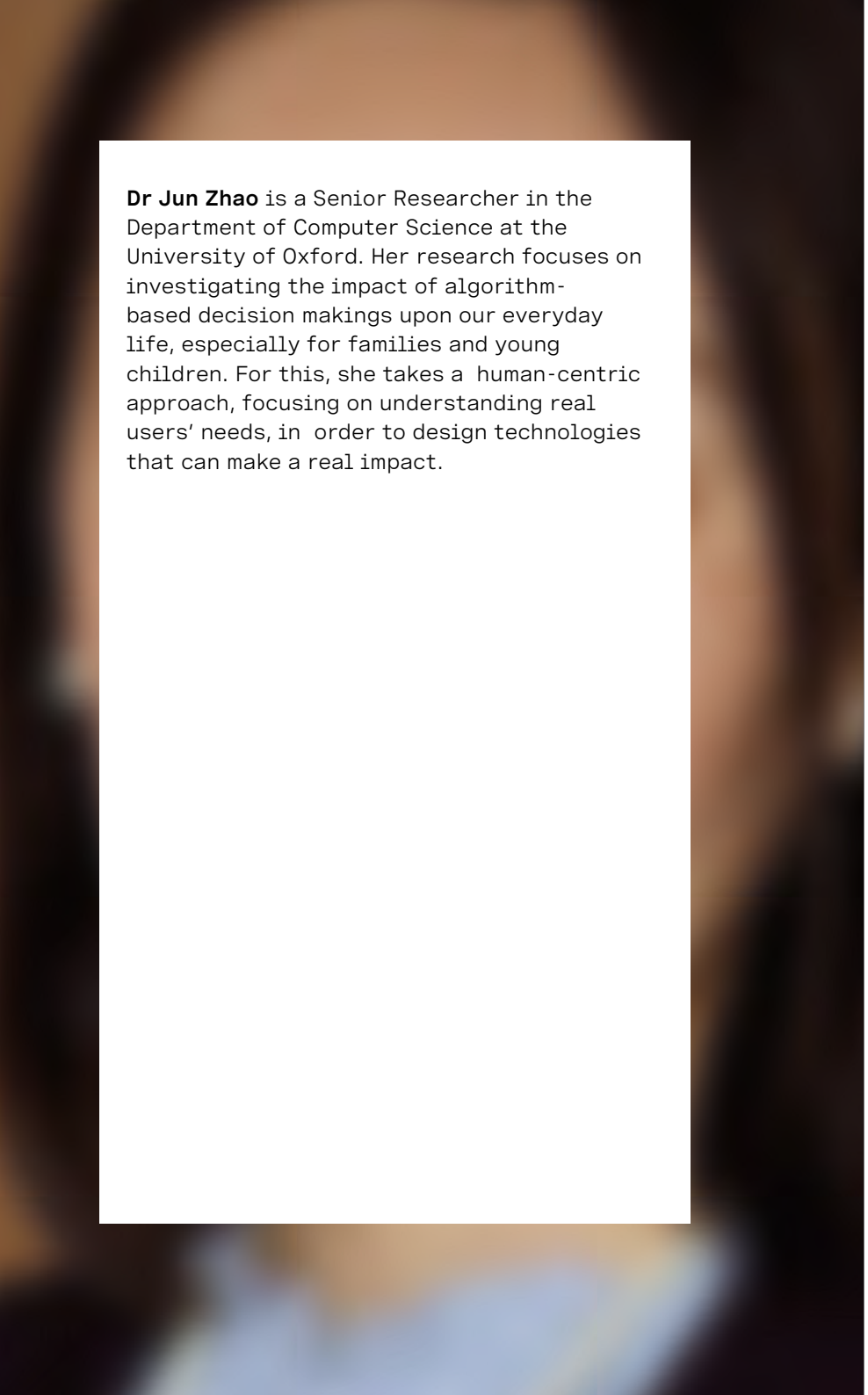
Sharp, E., Ricklefs, H., Leonard, M., Jones, R., Greenham, A., Carter, J., Broom, T., Bird, K., & Thompson, B. (2021). Enhancing media through the development of a Public Service Data ecosystem. *IBC 2021*

Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.

1 For more information about this product, see: https://solidproject.org

2 For more information about this product, see: https://mydex.org

3 For more information about the architecture of Databox, see https://imperial.ac.uk/systems-algorithms-design-lab/research/databox-project



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

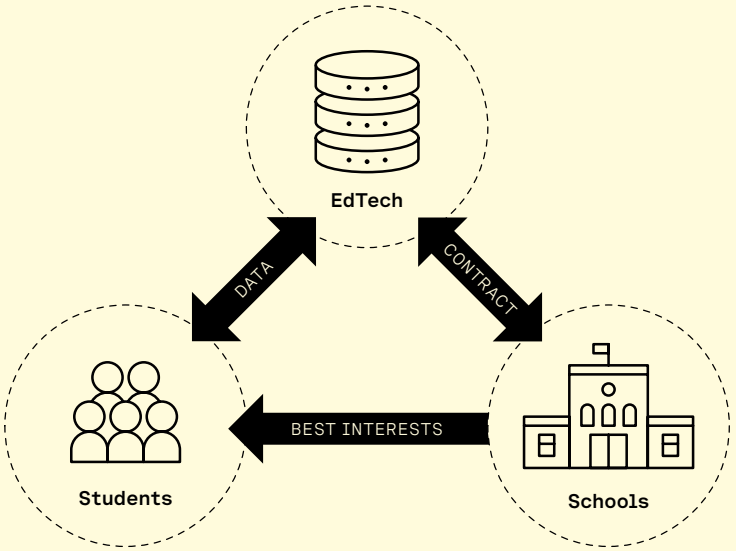


Figure 1(a): A hypothetical scenario of a school contracting an EdTech, with students' data held and processed by the EdTech company

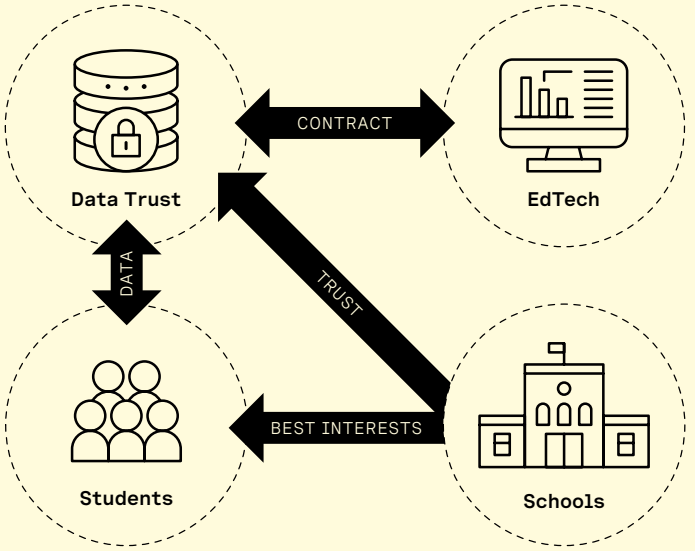


Figure 1(b): A hypothetical design of a school delegating the data protection and compliance responsibility to a trusted education data trust

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

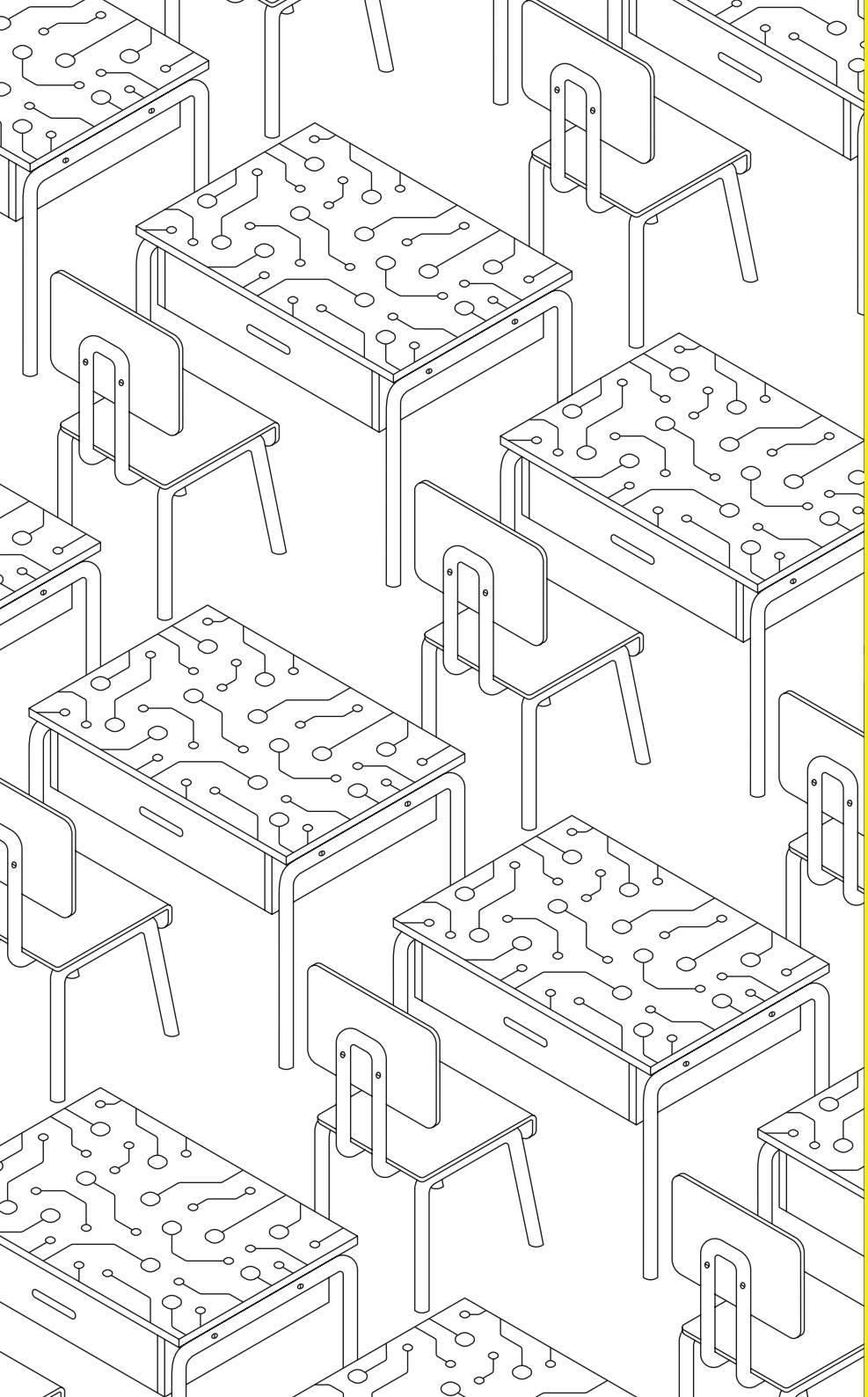
Walters, R. (2021). UK EdTech sector grows to £3.5bn as demand surges for digital classrooms and AR. *FE News*, 14 January

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



The data collected from children at or through their participation in school are exponentially increasing in variety, velocity and volume. But whose interests are served by this 'datafication' of education and childhood?

This essay collection offers critical, practical and creative reflections that identify exciting possibilities for beneficial uses of children's education data as well as tackling the exploitative uses or misuse of such data. Collectively, the essays set out principled yet practical proposals for our children's education data futures.