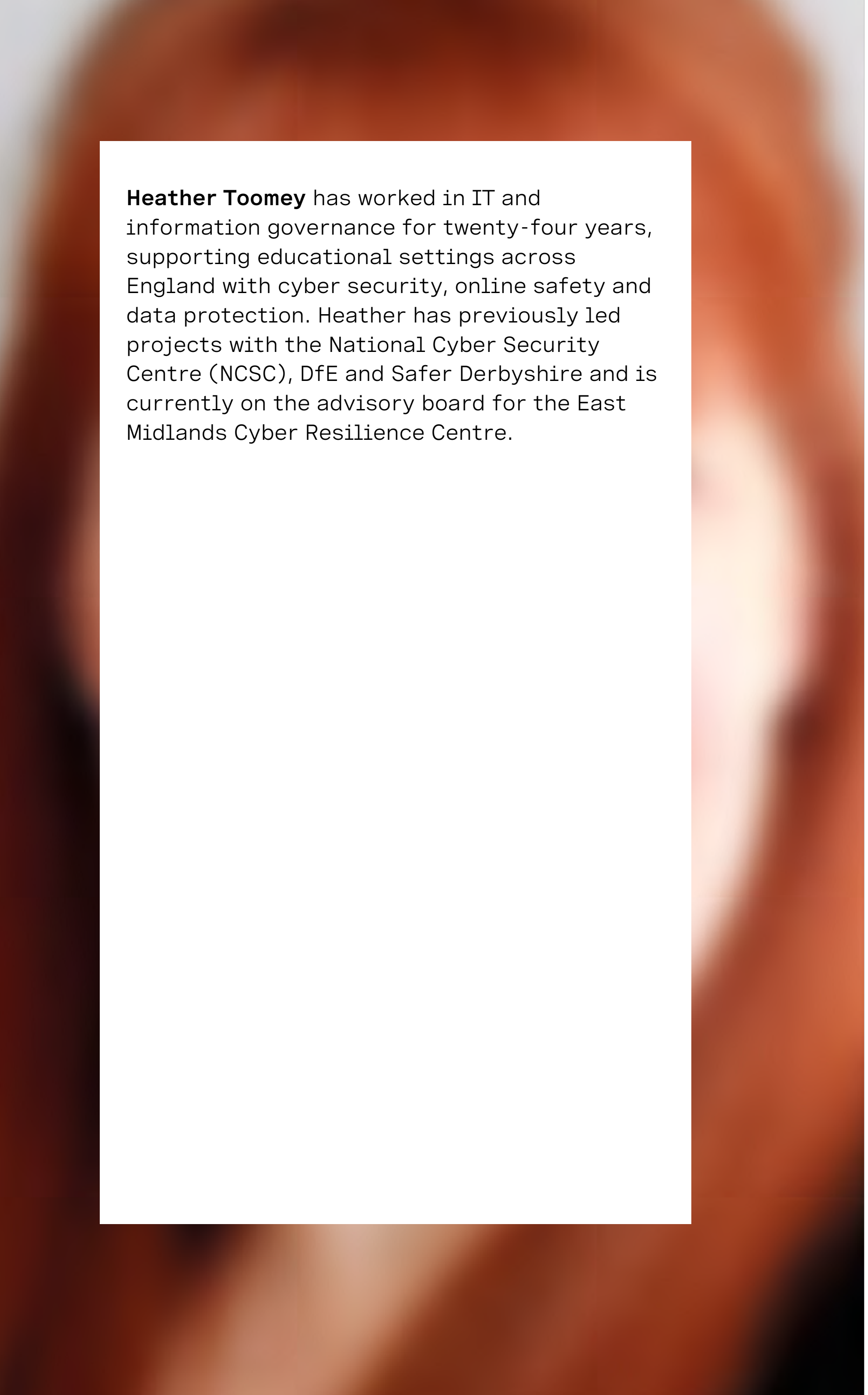

Heather Toomey has worked in IT and information governance for twenty-four years, supporting educational settings across England with cyber security, online safety and data protection. Heather has previously led projects with the National Cyber Security Centre (NCSC), DfE and Safer Derbyshire and is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

> (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

[†] 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

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