(How) is using data for learning ethical?

26 June 2023
About the Data4Learning webinar series

This webinar series focuses on the meaningful and ethical use of digitally processed data for student learning. The webinar series is planned to take place between October 2022 and June 2023. The webinar series addresses a range of topics around trust and safety aspects, ownership of data, interpretation and usability of data, exchange of data and ethical use of data, as well as aspects around inclusion and equity, and implications for ecosystem governance, technical infrastructure, and continuous professional development.

The focus of the webinars is to facilitate an open exchange between relevant actors in education on what works and what does not – and most importantly on how to move forward together. The webinars are run for European Schoolnet’s steering committee members and a closed community of stakeholders. Although participation to the webinars remains upon invitation only, their summaries are shared publicly through European Schoolnet’s channels.

(How) is using data for learning ethical?

Speakers May 2023

**Duuk Baten**, responsible AI advisor at SURF, the Dutch National Research and Education Network (NREN).

**Wayne Holmes** (PhD, University of Oxford), Associate Professor in the UCL Knowledge Lab at University College London.

**Sonia Livingstone** FBA, OBE, Professor of Social Psychology at the Department of Media and Communications at the London School of Economics and Political Science.
Webinar Summary

What do we mean by ‘ethical’ data in education?

Addressing the question, an interesting linguistic observation was made by Duuk Baten who challenged the idea that there is such thing as “ethical” data or “ethical” Artificial Intelligence (AI). Instead, Duuk suggested that the better way to phrase it is to talk about ethics of data or data ethics. In this way, we avoid giving agency to data or an AI system. Building on this argument, Wayne Holmes emphasised the importance of understanding that “words have life of their own” when chosen to describe things, but also that “once they’re being used, they can start to limit the way that we’re thinking about the thing in question”. He argued that if researchers 60 years ago had given, for example, the name “advanced cybernetics” instead of AI to the systems they were developing, we might have not been debating whether these systems have sentience or intelligence. Therefore, when talking about ethics, it is important to understand that it is humans who make choices and have agency rather than the machines or the information collected and used.

Going back to providing a definition of ethical data or ethics of data, Wayne Holmes suggested that we refer to “something that by design and in practice protects learners’, teachers’, and other stakeholders’ in education human rights”. That includes, in Wayne’s view, areas around agency, autonomy, safety, dignity, and protection from commercial exploitation. However, Wayne Holmes believes that many issues appear with the data themselves throughout their life cycle. The issues begin with the way data is collected, including questions of who collects it, for which purposes, and the choice people have in consenting for their data being collected. A second issue where ethics come into play relates to storage and ownership of data. In particular the safety of the storage, the accessibility to data sets, and the anonymity that pertains it. Important also for Wayne Holmes is the question of ownership of the data and the ability for the owners to use the information for future purposes, the usefulness of which cannot be yet determined. Next issue concerns the analysis and interpretation of the data, the impact of the outcomes to the data subjects, and the way the results will be used – e.g., to support learning or commercial objectives. Finally, important issues to consider are the fair and equitable treatment of data throughout their life cycle.

Feeding to the discussion, Sonia Livingstone, connected the point about human rights to more particularly children’s rights and proposed the United Nations Convention on the Rights of the Child as the guiding framework when thinking about crucial issues affecting young individuals. She also added to the questions that need to be considered when we talk about ethics of data, underlying the need to be clear about “to whom should data use be transparent, to whom should terms and conditions be comprehensible, and how does one claim remedy against wrongs?”. These questions are even more challenging when children are the data subjects, because it is very hard to imagine what the long-term implications could be for them in the future. Sonia Livingstone also underlined the duty we have to care and to use data to benefit children’s education. For this reason, data should be used in a way that respects their privacy and allows their voices to be heard. Moreover, she pointed out that although we live in a world of public-private
arrangements, we should be wary of the delicate boundaries between use of data by businesses for mandated purposes and commercial exploitation of students’ data.

“The overarching questions are how do we make a risk-benefit calculation when we use data and [how] do we know enough about the risks of when data use can go wrong and become unethical? How do we weigh that against the potential benefits? And do we have evidence of the benefits of using education data?” – Sonia Livingstone

Duuk Baten brought up the element of “justified choice” when it comes to the use of data that is part of the European Commission Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators. In his view, it is important to consider the ‘why’ and ‘how’ someone is using educational data and how are the systems used set up. Duuk Baten also brought up the importance of the ecosystem of stakeholders that are involved in making sure of the ethical use of data. This ecosystem consists of the governments that develop the policies and regulations, the designers of the systems, the actors that sell the systems, those that procure them, the project leaders and IT administrators that manage them, and the teachers that use and implement them. This, Duuk Baten, argued is a “super complex and technical system” and every step has part of the responsibility. In such a complex environment, it is natural that there exists a “diversity of thoughts”, but what is important, according to Duuk Baten, is “weighing on values, weighing on the things that we find important to make those right choices”.

Who is responsible that data in learning is used ethically?

In the question of responsibility when it comes to ethical use of data, Duuk Baten, underlined his previous mention to the complex system and the interplay between different actors in the system. In his mind, it is important for public sector education institutions to accept and get more responsibility controlling and owning education data, reversing the current situation where “a lot of responsibility is actually sort of given away or deferred a bit to avoid taking the responsibility by either outsourcing or hiring consultants or taking certain EdTech providers on board”. In this way, Duuk Baten believes education institutions will have more control over their data and will be able to better understand and analyse them for the improvement of education. Of course, he argues that more control means also more risk, but this is the way to “enable us to be in the driving seat [determining] the future of education instead of becoming more and more passive and allowing more and more of these tools to be developed outside of our school walls”. To achieve this, it is important to provide the right support and incentives for schools to be able to take this responsibility.

In this sort of complex social technical system, there is a whole lot of different people in, responsible and involved, and especially in education you have this public-private interaction, there is an element of also taking responsibility and giving responsibility away – Duuk Baten
Sonia Livingstone from her side, emphasised that those that have the responsibility to act are the governments and educational authorities of each system. However, what is often the case is that “they don’t necessarily have the power to act if they don’t own, or indeed don’t understand the data that is being collected and the uses to which it would be put”. On the other hand, Sonia Livingstone understands that “in the world of complex cybernetics, it would be beyond the capacity of the education system perhaps to generate the insights that are potential from the data” and therefore, “the commercial sector, the EdTech sector is needed both for the provision [of technology] within the schools, but also for the behind-the-scenes expert analysis [of data]”. That creates a delicate balance between the public and the private sector which is important to focus on “harnessing the insights […] and bringing those back to the classroom”. What is important for Sonia Livingstone is that we can reflect upon the questions: “Do children understand the data being collected and how it’s used? What is the relationship of public-private partnership in the world of analytics and data analysis? How does it [data use] come back fairly and does it come back sufficiently fairly that the collective good is served?”. For her, there seems to be a delicate balance between the responsibility of governments and education authorities to legislate and regulate, and the responsibility of each school to make decisions about the technology they are using. However, schools seldom have the “enormous specialist expertise” needed which is, according to Sonia Livingstone, “unrealistic” as it would require significant amount of technical, legal, and technological upskilling of schools which are, at the same time, facing teacher shortages and many other pressing issues.

Wayne Holmes brought a slightly different dimension to the discussion, questioning not the actor responsible for ethical use of data, but the data use for learning altogether. He argued that “the so-called data for learning” is primarily being collected by commercial actors via software provided to schools for profit. For this data collection to be considered ethical, Wayne Holmes believes it is fundamental that it supports students’ learning. If not, then he argues that “however else it might be ethical, the collection of that data just cannot be ethical, because at the very least, we’re wasting the time of our young people to have to engage with these tools that actually don’t do what they say”.

"The key question about the ethics of data collected in learning is whether that data actually does what they claim it does, which is to support the learning of young people in otherwise ethical ways." – Wayne Holmes

Wayne Holmes further argued that “there is so little independent evidence at scale that these tools do anything in the classroom of any benefit to anyone”. Although there are some small-scale academic studies, when comparing with studies done in the medical field to test the effectiveness of medication, there is not enough evidence to prove that particular technologies do what they promise to achieve. Furthering his criticism, Wayne Holmes disputed the promise of learning analytics as a way to understand students’ learning. He reasoned that “the problem with the learning analytics research space is [that] so much effort is put into checking that the prediction that they [researchers] are drawing from their analysis of the data that they have collected is an accurate prediction”, making it a “technical discussion”. In that sense, Wayne Holmes continued, “there’s a huge disconnect between the prediction and the actual
impact of that prediction on the individual children and their learning trajectory, on their way of existing in the classroom, and the relationship between the student and the teacher and the student and other students in the classroom”. In essence, he also added academics in the equation of responsibility for using data ethically.

**In times where ‘open data’ for research is advocated across Europe, can education data be ‘open’ for public use?**

Peter Karlberg, Director of Education at the Swedish National Agency for Education posed the question of ‘open data’ in education and the concern he has around educational authorities publishing educational data for public use. Addressing this question and concern, Sonia Livingstone highlighted that the topic is very important as although open data refers often to anonymised data, in education this can be reidentified as high numbers of sensitive data is collected within this context that can be put together and create very detailed profiles. Sonia Livingstone also raised the issue that in education, companies collect so many details that one can end up with micro-data like keypresses and time one takes to complete an activity. In an area where consent is already questionable, she believes it would be unethical to make students’ data public as this would allow private entities to be the ones to benefit most by using it for profit.

Duuk Baten drew a distinction between data collected in public space, which makes sense to be made public and further used for other public purposes, and sensitive learning data that could be used for unknown purposes. What Duuk Baten suggested is the idea of “federated access” where access to data is given to different groups of individuals based on the sensitivity they carry. In that way, researchers can still have some access to a dataset to conduct research, making sure specific checks and balances are in place.

“We realised that openness is not ideal because openness also means openness to everybody and access to the strong players which are usually these commercial players to keep using it in ways that we might not have wanted to.” – Duuk Baten

Wayne Holmes suggested caution to the context certain words are used for the meaning they carry. The word “open”, he argued, is a positive word when it comes to for example open educational resources (OER), but it has a less positive connotation when it comes to open educational data. Wayne Holmes, similarly to the others, argued that he is in favour of open data about infrastructure for example, but sharing individuals’ data -even anonymised- can be “hugely problematic”. This “Silicon Valley push” carries the “notion that actually, if we share data, we make data open for everybody and everything and then the world will be a better place”. Wayne Holmes thinks that “the principle of open is good” but one needs to be aware of “the way that it’s implemented and to whom it is imposed”.

“We often hear people saying data is the new oil, but the difference is you burn oil and it’s gone. It pollutes and does all the other nasty stuff, and it moves our cars, but it’s gone. You can’t reuse it once it’s been burnt, but
data is not like that. Data exists and will continue to exist in perpetuity and can be repurposed, and it only takes, you know, one slip of a of a company which is meant to be protecting the data to forget to renew a certificate or something, and suddenly that data is accessible to the whole world in one way or another.” – Wayne Holmes

How do we define what is ethical for technology?

In his reply to the question, Wayne Holmes argued that when we talk about technology or data in education, we often forget about the ethics of education itself. This is something that has “been questioned and debated over 2000 years and probably before that”. That is why, when we talk about technology in education, it is important to look at the pedagogical soundness of a particular tool, whether it “actually does what it claims to do”, and the evidence that support this. Although Wayne Holmes understands that “one person’s evidence is somebody else’s hearsay”, he argued that the bottom line is to know whether “these technologies enhance teaching, they empower teachers or they disempower teachers, do they give agency to students?”. Another important consideration relates to the development of technology for education. Wayne Holmes shared that it is known that “most developments of technology for use in education mostly aim at the average student because what they’re [companies] trying to develop is tough”. This way, he argued, EdTech developers “aim at the middle of the bell curve” of the distribution of students, but then transferring the application to those at the edges of the bell curve (the students with bigger learning challenges) is very difficult. What Wayne Holmes suggested is that “we should start by working with those who have the biggest inclusion needs, and we should be making education appropriate for them, not trying to force them to be appropriate for the education that we have”. In this way, he continued, “it becomes much easier to transfer what you’ve learned from those at the edges of the bell curve to those in the middle”, considering “equity and inclusion from the very beginning, not as a bolt-on at the end”. Adding to that, Wayne Holmes also highlighted the importance of safety and wellbeing of students. He argued that it is important to make sure “technologies that are being developed are, before they’re put into use, actually safe by design”. Companies should also “prioritise wellbeing of the students by design and avoid making young people accidentally coming to contact with inappropriate materials”. All these, Wayne Holmes stressed should “be done by design, not as an afterthought”. Finally, he also underlined the importance of developing and enhancing digital citizenship of everyone, including young people to help them “self-actualise, do it safely, maintain their wellbeing, and help them contribute to society”. Sonia Livingstone argued that, when answering this question, one should consider whether to question the technology itself or the company that develops it. Moreover, one should distinguish between “winner takes all” companies that are giants in the education market and “hoover” students’ data, and smaller players that are more targeted to specific areas of the curriculum or education provision. She also brought up the different kinds of technology use in education. On the one hand, there is the technology use for developing skills, needed facts, some kind of personalisation, rote learning, and assessment. On the other hand, Sonia
Livingstone argued, there are affordances that technology can provide but are “enormously underused”. Those relate to opportunities for networking among students, collaborative and creative learning, tinkering and making, and to support “child-led learning”. However, Sonia Livingstone admitted that pedagogy and use of data is a “serious political question”, and different actors have different views on what beneficial use of technology in schools is.

“An even more personalised, ever more individualised, ever more quantified learning - is that really the route to produce the creative, agile, canny workers of the future? I don’t think so.” – Sonia Livingstone

Duuk Baten brought to the discussion the important principle that “technology is not a neutral thing”, but rather it is filled with values which then have an impact on the ways education is considered, designed, and delivered. Technology use, like learning analytics, Duuk Baten argued, brings up certain aspects of a student that will influence the teacher and their perception they have about the particular student. On top of this value struggle, one has to add the power dynamics related to the ecosystem actors, discussed earlier, which raises questions about ownership and development of technology tools and the primary beneficiaries of its application. Schools, Duuk Baten continued, have certain control over what is happening within their walls, but what takes place outside can also influence the way they work. Students use technology not only in school, but much more outside of it, and this can affect the way they see the purpose of their education.

“Schools cannot be the only ones bearing the responsibility for this [technology use] because students are also autonomous beings. They’ll use interesting stuff, shiny new tools, and that influences their perception of the way they see education, the way they learn from education.”

– Duuk Baten

In defining what is ethical for technology, Duuk Baten suggested three elements to be taken into account. First element relates to legal and regulatory frameworks that should put safeguards and should be developed with human and children’s rights in mind. Second element to be considered is the values that drive the different approaches to education and act as a “common language”. Those, according to Duuk Baten, can be formed on a national, sectorial, or even school level and should encompass what is considered important. These values would then be part of design and procurement systems, funding mechanisms, and data infrastructure. The final element relates to “ethics or responsibility as a practice” where every actor and professional must act responsibly, “not hype things that we don’t know about enough, try to be honest about what we’re doing, try to enable our own virtuous behaviour as individuals”.

Duuk Baten argued that by acknowledging that the “reality is dirty, it’s gritty, and it’s difficult” and that “there are no easy right answers” we can “go for that technology worth wanting”.

Sonia Livingstone echoing the suggestions Duuk made, mentioned the work that the Digital Futures Commission is doing to develop a set of 11 principles for designing digital products for children. She,
however, also underlined the need that educators have for more “pragmatic requirements” that can help them when choosing technology to use in their lessons which fulfils some minimum standards. The Digital Futures Commission is therefore also developing a Blueprint for certification, which can allow schools to assess technology based on some minimum criteria. Setting minimum standards and certification, Sonia Livingstone argued, should be the work of education authorities which would help guide schools and at the same time, provide certainty for the industry on the regulations they need to abide by.

Adding to the conversation, Wayne Holmes also shared the work that the Council of Europe is doing on working towards a Convention that focuses on Teaching with AI and data, looking at “how we can use technologies effectively and innovatively in education settings”. It will act as a Blueprint or guidance for governments to set, as Sonia mentioned, some minimum standards that can be adapted according to particular contexts and circumstances. With this opportunity, Wayne Holmes also raised the often-quoted promise from EdTech providers about “automating and speeding up what they think we do in education”. He claimed that “the difficulty of doing this kind of work is the problem-based versus solution-based approaches to developing education technologies, and almost all education technologies are solution based”. Wayne Holmes argued that EdTech companies should start asking educators what the problems are they want technology to help them with, rather than the other way around. In his view, “the Edtech companies [tend to] go for the low hanging fruit, what's the easiest way to do something and actually don't really explore the possibilities” which can help us address the question of “what do we want from education” and “what is its purpose”. However, as mentioned above, Wayne Holmes concedes that “we don't have agreement about that” and therefore it is “very difficult to understand what [the role of] technology should be in education and what we perceive as being ethical uses of technology”.

“One of the interesting things about working in the field of technology in education is that soon as you think about how a particular technology might work in a classroom setting or what it might try to do, it makes us think about what's the education that we're trying to do actually is.”
— Wayne Holmes

How can schools, teachers, and parents say yes or no to something that they do not really understand?

Addressing this question, Sonia Livingstone suggested that “no one should ever say yes to anything they don't understand, and that would not be informed consent if they did so - it would be invalid”. Sonia Livingstone underlined that it is paramount that schools, teachers, parents, and students should be able to understand to what they are consenting to. In her view, “if we don't understand we cannot consent, so it means something is wrong with the architecture of choice”. In this sense, consenting to something should be a genuine choice. Coming back to minimum standards, Sonia Livingstone offered the idea of a sandwich when it comes to consent. The bottom part refers to certification and minimum standards set by education authorities, while the top part refers to principles, values, and children’s rights. In-between those, Sonia Livingstone suggested should be the ecosystem that “sets out possible visions that advises on
procurement that can offer independent evaluations of what's working”. This part should be “a kind of a moving space because people are experimenting and trying things out”, where also ideas about education are shaped and changed over time with the advance of technical and empirical knowledge.

Wayne Holmes built on Sonia’s arguments and added that although he considers informed consent as essential, the license approach consent that most technology products currently use does not work. In his view, teachers are already overwhelmed with their duties and thus responsibility should be with the “ministries to ensure that teachers have good quality professional development in this domain”. Teachers, Wayne Holmes argued, should be given real space for professional development – this is the only way to consider it an opportunity for them. He also mentioned that providers of technology used in education should be “open and transparent in genuine ways”.

From his side, Duuk Baten argued that “responsibilities are so stretched throughout this complex ecosystem of decisions being made and there is an interesting interaction happening where responsibility needs to come with an ability”. This ability refers to think and understand what one is reading, for example in the Terms and Conditions of a particular product. Finally, Duuk Baten also questioned the ability of students to really consent within the framework of the school, due to its hierarchical relationship between the school and the student.

Further reading suggested by the speakers:

Other publications the speakers shared: Data4Learning, (How) is using data for learning ethical, Resources


(How) is using data for learning ethical?

https://op.europa.eu/en/publication-detail/-/publication/50c53c01-abe8-83e1-01aa75ed71a1/language-en


**Resources suggested by participants during the webinar for future reflection:**

Weighing Values (kennisnet.nl)

Microsoft Word - Problems with data governance in UK schools- the cases of Google Classroom and ClassDojo.docx (digitalfuturescommission.org.uk)

Update afspraken en gesprekken SURF en Google | SURF.nl

Child Rights by Design - SRights Foundation & DFC (digitalfuturescommission.org.uk)

A Blueprint for Education Data (digitalfuturescommission.org.uk)

The Ethics of Artificial Intelligence in Education: Practices, Challenges (e-bronnen.be)