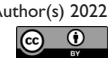
Article

new media & society
2024, Vol. 26(4) 1871–1890
© The Author(s) 2022

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14614448221079626
journals.sagepub.com/home/nms


Pre-service teachers' insights on data agency

Henriikka Vartiainen, Lotta Pellas, Juho Kahila, Teemu Valtonen and Matti Tedre
University of Eastern Finland, Finland

Abstract
With growing concerns over children's data agencies, researchers have begun to draw attention to children's and young people's privacy practices in social media environments. However, little is known about the experiences of pre-service teachers who play a key role in educating future generations. This study aimed to address this gap by exploring Finnish pre-service teachers' conceptions and experiences of data agency in social media environments. Drawing from in-depth interviews of pre-service teachers ($N = 14$), the analysis revealed that pre-service teachers construct their data agency in terms of social frames and shared social norms, and they also recognize the lack of understanding regarding wider socio-technical systems within which data agencies are situated. This research argues that without a sophisticated understanding of algorithmic governance and commercial use of data, it is unlikely that these future teachers would be prepared to facilitate children's and youth's agentive actions in a data-driven society.

Keywords
Data agency, media practices, pre-service teachers, relational privacy, surveillance capitalism, teacher education

Introduction

The idea that education should support people to develop their capacities for agentic and autonomous action has formed an important tradition in Western society as well as in prominent educational research (Eteläpelto et al., 2013). Over the years, the concept of

Corresponding author:
Henriikka Vartiainen, School of Applied Educational Science and Teacher Education, University of Eastern Finland (UEF), P.O. Box 111, FI-80101 Joensuu, Finland.
Email: henriikka.vartiainen@uef.fi

agency has been defined in a multitude of competing ways and applied in educational research from a broad variety of different disciplines. For example, social cognitive theory approaches agency from the perspective of intentionality which includes action plans and regulation strategies for realizing those plans (Bandura, 2001). From a socio-cultural perspective, agency is seen as goal-directed action that is always constrained and resourced by certain socio-cultural, technological, and material circumstances (Wertsch, 2007). Moreover, sociological approaches typically emphasize that agency cannot be divorced from social and economic structures that shape or constrain human agency (Eteläpelto et al., 2013). Recent discussion on agency has also extended its temporal dimension from the present to the past and to the future, indicating that agency is a complex, relational, and dynamic process that develops over time (Emirbayer and Mische, 1998).

On the other hand, most theories of agency were formulated long before the breakthrough of machine learning (ML) in the early 2000s and the new social realities these technologies have created. Currently, the complexity of human agency is increasingly under critical discussion as many aspects of everyday life, such as our everyday interactions, action, and information of all kinds, are increasingly being mediated, augmented, produced, and regulated by algorithmic governance (Kitchin, 2012). Personal and behavioral data which once were a mere byproduct of online participation have now become a valuable economic resource for platform owners collecting massive amounts of contextual information about people's daily actions and interactions, such as browsing habits, locations, purchases, status updates, ratings, comments, and emotional reactions, among endlessly many other things (McCosker, 2017; van Dijck, 2013; Zuboff, 2015). Such multidimensional user data that most people voluntarily share enable the creation of rich data profiles on a massive scale, indicating that our personal choices are increasingly influenced by the choices of our social networks and like-minded others (van Dijck, 2013). What is more, massive-scale data collection also enables targeted marketing, behavior engineering, attention harvesting, and targeting of mis- and disinformation, among an endless number of other uses (Hendricks and Vestergaard, 2018; Kramer et al., 2014; Valtonen et al., 2019).

Hallinan and Striphas (2016) argued that human agency is increasingly situated in a new kind of "algorithmic culture" which they describe as the "use of computational processes to sort, classify, and hierarchize people, places, objects, and ideas, and also the habits of thought, conduct, and expression that arise in relationship to those processes." Their view illustrates the socio-technological trajectories through which computational processes are increasingly recommending which objects, ideas, and practices enter our cultural realm, and in what form (Klinger and Svensson, 2018). According to Klinger and Svensson (2018), algorithms base their statistical and ML models on people's past behavior to project their future, most often in the service of advertisers and social media companies. Hallinan and Striphas (2016) argued that this kind of cultural production also produces interactions, emotional reactions, behavior, and new habits of thought that likely would not exist in the absence of those algorithmic processes. Accordingly, data-driven systems pose severe challenges for our agency as they can construct cultural realities, bubbles, and echo chambers that may be internalized by individuals in unconscious ways (Pangrazio and Selwyn, 2019; Valtonen et al., 2019).

Against this context, there is a pronounced need to understand how individuals exercise and develop new forms of agency within the socio-cultural and technological constraints and resources of datafied society.

The increasing datafication of society has also prompted calls for new kinds of digital competences which should be acquired by all citizens, to ensure personal development, social inclusion, and active citizenship. For example, the framework offered by the European Union (2016) proposes that today's citizens need the skills to communicate, collaborate, and solve problems with the help of online tools as well as have the skills to search, evaluate, edit, and produce digital content. This perspective recognizes that digital citizens need to understand common terms of service, know how to protect personal data and privacy in digital environments, and need to be able to protect oneself and others from online threats that stem from the pervasive use of digital technologies (Vuorikari et al., 2016). Along with safe, responsible, and ethical use of technology, digital citizenship is associated with civic engagement, acting as a force for change, and for critical resistance (Choi et al., 2018; Jones and Mitchell, 2016). To some extent, these calls for digital citizenship are also aligned with the data literacies that typically emphasize the ability to identify how and where personal data are generated and processed (Pangrazio and Selwyn, 2019), ability to manage and control personal data trails and practices (McCosker, 2017; Pangrazio and Selwyn, 2019), as well as ability to constructively engage in society through and about data (Data-Pop Alliance and Internews, 2015). According to Pangrazio and Selwyn (2019), data literacies should also include critical understandings of social media logic, including awareness of reconstitutions and recirculation of personal data.

In accordance with the data literacy approach, we suggest that data agency calls for a critical attitude toward digital being, ownership, and control of one's own data as well as informed ethical and moral decision-making grounded in understanding of how data are generated, processed, and used for different purposes (Tedre et al., 2020). While data agency can be seen as having capacity and volition to make informed decisions and personal data strategies, we also see a need to understand data agency as a relational and dynamic process that develops over time. In line with the socio-cultural theories of learning and participation, agency is not seen merely as individual attitudes or competences because online actions and interactions are tightly connected to social practices and tool-mediated actions taking place in various informal and formal settings. In this respect, data agency needs to be considered beyond the individual level and take into account the socio-cultural practices, computational processes, and wider structures in which the agency is situated. Stressing that agency should not be considered to be a stable property or competence, this perspective also highlights that agency varies over time and conditions (Eteläpelto et al., 2013) and may develop via learning. In this process, more mature members of the community, such as teachers and parents, play a key role in providing means, guidance, and metacognitive support adjusted to children's evolving skills and understanding (Rogoff et al., 1993).

However, many of the issues that impact children and young people's data agency are absent from their education (Pangrazio and Selwyn, 2019; Stoilova et al., 2020). According to Livingstone and Yoo (2018), enormous amounts of personal data are being collected throughout childhood, with or without consent, with or without

knowledge of who is providing those data, and furthermore with or without sufficient security provision. While it is assumed that parents and educators play a role in safeguarding and supporting children's privacy rights, many of these issues occur without adult awareness of them (Livingstone and Yoo, 2018). These actions are also embedded in a context of legal and policy regulations, such as General Data Protection Regulation, which makes assumptions that teenagers and parents are both able and willing to make informed decisions when engaging with complex online data consent mechanisms of diverse services (Keen, 2020). Moreover, existing barriers to accessing age-restricted content, goods, and services are mostly ineffective as they can be easily bypassed, for example, by making a wrong self-declaration of age or by using gift cards (Smirnova et al., 2021). In all, this complexity of data-driven society suggests the need to understand how agency is practiced and how the interpersonal, institutional, commercial, and technological factors shape and constrain agentive actions (Keen, 2020; Livingstone et al., 2019).

So far, researchers have investigated agency and related online privacy practices from the perspective of youth (e.g. boyd, 2014; boyd and Marwick, 2011; Davis and James, 2013; Keen, 2020; Lapenta and Jørgensen, 2015; Marwick and boyd, 2014; Pangrazio and Selwyn, 2019; Tufekci, 2008), younger children (e.g. Livingstone et al., 2019; Tedre et al., 2020), and their parents (Stoilova et al., 2020). A recent study by Stoilova et al. (2020) also provided some insights on the perspectives of teachers. However, little is known about the experiences of pre-service teachers while they play a crucial role in transferring data literacies and agencies into educational practice. Moreover, understanding of how pre-service teachers perceive their data agency is important for teacher training that needs to ensure that future teachers are able to support students in learning the competencies needed today and in future. The current study addresses this gap by exploring Finnish pre-service teachers' ($N = 14$) conceptions and experiences of data agency in social media environments. We start by introducing previous research on youth data practices with the aim of mapping how they make decisions and exercise control over their personal data in social media environments and how this is constrained and bounded by interpersonal and commercial factors (Livingstone et al., 2019). Then, we built upon this previous research and applied it in an empirical analysis of interview data with the aim of identifying elements of data agency in practice, as experienced by pre-service teachers. The article concludes with a discussion on pre-service teachers' experiences of agentive actions and gaps in knowledge, and how this should be taken into account in teacher education.

Literature review: perspectives on youth data agency in social media environments

To capture the complexity of children and youth's agency online, Livingstone et al. (2019: 4) have theorized three "types of privacy," namely, the interpersonal, institutional, and commercial. Interpersonal relates to "how a 'data self' is created, accessed and multiplied via social connections," whereas institutional refers to "how public agencies such as government, educational and health institutions gather and handle data about a person" and commercial refers to "how personal data is harvested and used for business and

marketing purposes" (Livingstone et al., 2019: 4). As this study is centered on pre-service teachers' conceptions and experiences of data agency in social media environments, we focus on the interpersonal and commercial levels which are at the core of understanding how agency is experienced and exercised in that particular context. Next, the extant literature is reviewed to identify how data agency is practiced by the youth.

Data agency on the interpersonal level

While models of data sharing are typically approached through the lens of individual rights and controls, previous research on youth practices has highlighted that their data agency is socially constructed (boyd and Marwick, 2011; Livingstone et al., 2019; Pangrazio and Selwyn, 2019). First, the presence of peers and friends is an important factor when youth make decisions on what services or applications they use and how much they disclose online (boyd and Marwick, 2011; Davis and James, 2013). According to boyd and Marwick (2011), social media sites have become the modern-day equivalent of the physical places where youth hang out with friends and run into other friends and peers. However, while youth want to participate, they also want to have control over their online visibility to others (boyd and Marwick, 2011).

Previous research has identified a number of strategies that youth use for protecting their privacy from both known and unknown others online. For example, youth are using different services for interacting with different social cohorts (boyd and Marwick, 2011; Lapenta and Jørgensen, 2015), using friends-only controls when posting online content (boyd and Marwick, 2011; Davis and James, 2013; Hargittai and Marwick, 2016; Marwick and boyd, 2014; Tufekci, 2008), switching between multiple accounts (Davis and James, 2013; Hargittai and Marwick 2016), using pseudonyms (Hargittai and Marwick, 2016), providing false information (Davis and James, 2013), deleting or blocking people (Stoilova et al., 2020), cleaning and deleting wall posts, comments, and tags (Davis and James, 2013; Raynes-Goldie, 2010), and using language and images that hold a particular meaning only to a particular group of people (Marwick and boyd, 2014).

Previous research has also identified strategies that involve withholding content from online spaces. For example, youth do not want to share their detailed contact information (e.g. full names, addresses, and phone numbers), credit card numbers, intimate personal matters (e.g. health records, political beliefs, "emotional things," such as parents' divorce and break-up with girlfriends and boyfriends), or content that they consider "inappropriate," such as sexual content or "duck-face selfies" (Davis and James, 2013; Hargittai and Marwick, 2016; Lapenta and Jørgensen, 2015). However, they also use more private messaging channels, such as Snapchat, for communication of more confidential issues with limited groups of friends (boyd and Marwick, 2011). These various personal data strategies and decisions what to share and what to withhold may vary among users (Hargittai and Marwick, 2016); however, these deliberate acts implies that youth value their privacy and are proactively taking agentive action regarding their current privacy concerns (Marwick and boyd, 2014; Stoilova et al., 2020; Tufekci, 2008). What is more, previous research has also demonstrated that youth are concerned about social risks related to sharing, such as cyberbullying, sexting, being tracked by strangers, and unwanted communications from strangers (Keen, 2020; Stoilova et al., 2019).

While youth exhibit control over what data they share and where, they have far less control over what friends, family members, and parents post about them (boyd and Marwick, 2011; Marwick and boyd, 2014). Parents have a long history of using the web to discuss and share parenting practices on diverse sites, and today, they often share multimedia information about their children (Leaver, 2017; Lupton and Williamson, 2017). In addition, sharing pictures of friends, such as group photographs, videos, and other media are part of contemporary youths' social bonding. Previous research has found that youth often rely on trust and negotiate commonly shared social norms to manage what is shared about them and by whom (Lapenta and Jørgensen, 2015; Marwick and boyd, 2014). However, Marwick and boyd (2014) note that how such posts are (re)read or viewed depends greatly on the context that varies across one's lifetime and thus may lead to "context collapse." For example, sharing of online "drunken narratives" are part of young people's social bonding, but there is a great risk that such posts may end up being accessed far beyond those they were intended for (Goodwin et al., 2014). Accordingly, what is appropriate for close friends to see may not be so appropriate for their future employer (Raynes-Goldie, 2010). Tufekci (2008) suggested that youth generally make choices about publicity based on their current concerns and thus may be shortsighted about future problems.

Data agency at commercial level

Previous research has also demonstrated that young people are more concerned about social privacy rather than how companies and their partners might use their personal information (boyd, 2014; Keen, 2020; Lapenta and Jørgensen, 2015; Livingstone et al., 2019; Raynes-Goldie, 2010). In other words, privacy is often seen in terms of controlling and limiting what they share with friends, parents, and others (boyd, 2014) and they perceive circulation and the sharing of data with advertisers as an acceptable trade-off for participation in networked activities (Lapenta and Jørgensen, 2015; Selwyn and Pangrazio, 2018). However, there is evidence that some youth attempt to control commercial tracking, for example, by turning on incognito options in their browser, opting out of certain apps or sites, deleting cookies, and even using Do-Not-Track browser plugins as well as password-management apps (Hargittai and Marwick, 2016). On the other hand, existing research also suggests that youth may provide personal data unconsciously and with little understanding of where, how, or why their personal data are being collected (Keen, 2020; Lapenta and Jørgensen, 2015; Selwyn and Pangrazio, 2018; Tedre et al., 2020). User agreements are hardly ever read and data privacy is typically seen as a responsibility of the platform provider (Keen, 2020; Lapenta and Jørgensen, 2015; Selwyn and Pangrazio, 2018; Tedre et al., 2020). Keen (2020) found that youth viewed the platform owners as impersonal and innocuous, and their relationship with them was seen as functional, non-judgmental, and convenient. Moreover, they felt that their data were circulating within commercial contexts, unlikely to be released in a manner that may result in reputational, material, or physical harm. Consequently, they saw no need to control what companies did with personal data generated about them (Keen, 2020).

Hargittai and Marwick (2016) argued that normalization of data collection and surveillance by states and companies may also lead to the feeling of a lack of control, generating

"apathy," "cynicism," and the impression that "privacy violations are inevitable." Similar experiences were also demonstrated by Selwyn and Pangrazio (2018) who found that while youth were generally interested about online privacy, they did not feel empowered to make changes, nor did they feel in control of their privacy. Moreover, Stoilova et al. (2020) found that there are substantial differences in the support children and youth receive from parents concerning their privacy issues, which creates inequality and makes some children more vulnerable. They also found that teachers typically focus on children's e-safety with little understanding of the data economy or digital infrastructure underpinning everyday interactions (Stoilova et al., 2019). However, this aspect has received little attention in teacher education (Valtonen et al., 2019) and there are, to our knowledge, no research on pre-service teachers' insights on data agency in social media environments. Nevertheless, understanding existing data practices and experiences of pre-service teachers is vital for providing future educators with readiness to prepare their students for participation in data-driven society. In this study, we aim to enrich these perspectives of data agency by addressing the following research question: *How do pre-service teachers experience their data agency at the interpersonal and commercial levels*?

Methodology

Context of the study

This article draws from 14 in-depth interviews with pre-service teachers conducted at one university in Finland. In Finland, teacher education, especially for elementary teachers, is among the most desired academic programs at universities that attract highly talented applicants (Niemi and Lavonen, 2020). All teachers go through a 5-year academic program that contains the Bachelor of Arts (Education) degree (180 European Credit Transfer and Accumulation System [ECTS]) and the Master of Arts (Education) degree (120 ECTS). The Bachelor of Arts degree typically covers the first 2–3 years of teacher education and the Master of Arts degree again from 2 to 3 years. A special feature that differentiates the Finnish teacher education from many other countries is that teacher training is designed to cultivate new teachers as reflective professionals who are able to make use of research-based practices within their work (Niemi and Lavonen, 2020). Based on recent longitudinal studies, the Finnish pre-service teachers are typically rather confident with their pedagogical expertise and unconfident with areas related to the technology (Valtonen et al., 2021).

The teacher training aligns with the Finnish National Core Curriculum for basic education (FNBoE, 2014) that is informed by a strong equity ethos. This means providing equal opportunities and high-quality teaching to every child regardless of their social, ethnic, and economic background (Niemi and Lavonen, 2020). Along with special subjects, the curriculum contains seven transversal competencies that need to be considered as part of all teaching provided. These seven competencies cover areas like *taking care of oneself and managing daily life, multiliteracy, information and communication technology* (*ICT*) *competencies*, and *participation, involvement, and building sustainable future* (FNBoE, 2014). Accordingly, Finland, like many countries, has included media literacy, ICT skills, and computational thinking in curricula. For example, the students are expected to gain

understanding of the digitalization and its impacts on various areas of life, to be able to observe their significance in their daily life, in interpersonal interaction, and as a channel of influence. Again, students need to gain skills for evaluating information in different modes and developing critical thinking skills. Altogether, the aim is that basic education provides students with the knowledge and skills needed for active agency and citizenship (FNBoE, 2014). While not explicitly stated nor taught in school praxis, data agency and understanding of underlying algorithmic processes are in key positions within these national goals (Valtonen et al., 2019).

Data collection

The research was conducted aligning with the instructions of the Finnish National Board on Research Integrity (TENK, 2019). All data collection procedures received institutional approval by the administration of the department of teacher education. In spring 2021, the researchers distributed an introductory email to 213 pre-service teachers, including the participant information sheet required by the university. This participant information sheet included the description of the purpose of the study, statement that the participation is entirely voluntary and the participant can withdraw from the study at any time, information about types of personal data that will be collected, how it will be processed, and how research results are reported. Moreover, the participant information sheet provided information about participants' rights as well as contact information of the data controller of the study.

The target group consisted of pre-service teachers who were studying, as part of their major or minor subject, for the qualification of elementary-class-teacher education. While this call for participation was sent to all students participating in an obligatory course that is situated at the end of a Bachelor's degree, this study was not part of the course tasks and all participation was voluntary. Out of 213 pre-service teachers, a voluntary response sample was drawn, from which 14 teacher students with a mean age of 26 years (range, 20–40 years; 50% female) volunteered to participate in the interviews and all gave their permission to use the data collected. These interviews were organized late in spring 2021 and they were carried out by the second and first authors. To ensure the anonymity of the participants, no demographic details other than age and gender were collected in any systematic manner. These one-to-one interviews with participants were conducted in spring 2021 through video-conferencing due to the COVID-19 pandemic. These interviews resulted in a total of 12.48 hours of audio data (average length of an interview was 54:55 minutes). All interviews were transcribed verbatim, and real names and other data related to personal identification are not reported.

In these in-depth interviews, a number of questions were intended to capture students' conceptions of (data) agency. First, the students were asked general questions about their experiences of agency (e.g. decision-making in life, how they formulate opinions, how do they influence societal issues). Second, they were asked to elaborate what social media services they use, for what, and how they make decisions on what applications they use. Third, the interview included questions related to interpersonal factors of data sharing (e.g. how they define their personal privacy, what kind of data they share, for whom and why, what kind of data is shared about them by others, and how they would

feel if a future employer, school pupils or their parents searched their data traces). Commercial factors were also raised (e.g. how and why their personal data are being collected, how it is processed and used, what kind of experiences they have on targeted content, what kind of risks they saw at commercial level, and how collection of personal data affects their future). Finally, the students were asked to reflect on the development of media education in schools as well as in teacher education. In summation, the assumption was that the topic was novel, and thus, these interviews were expected to provide students with an opportunity for sense-making and reflection of data agency in a manner that connected their past experiences, engagement with the present, as well as orientations toward the future.

Data analysis

Epistemologically, this study is based on the critical paradigm in educational research that aims not only to understand subjective meaning making but includes critical consciousness needed for questioning the status quo and dominant practices (Mack, 2010). In this respect, analyzing pre-service teachers' lived experiences was suspected to provide important insights for harnessing good practices and identified challenges for taking informed action aimed at improving teaching and learning praxis.

In this study, the analysis of data relied on content analysis (Elo and Kyngäs, 2008; Hsieh and Shannon, 2005). In particular, the analysis followed a deductive approach in which a coding scheme was developed based on previous empirical research (Elo and Kyngäs, 2008). In the first phase, prior research and discussion related to data agency and how it is practiced in interpersonal and commercial levels were traced from recent scholarly literature. In other words, existing literature helped to determine the initial coding scheme as well as to identify explicit definitions for each code. After a scheme was developed, the data were reviewed and coded with the help of data analysis software Atlas.ti. The first author was responsible for conducting the analysis that progressed through iterative steps of reading and re-reading the transcribed interviews, and identifying different theory driven data strategies in students' reflections of their lived experiences. Subsequently, new codes were created inductively from the stories and perspectives shared by our interview participants. To increase the reliability of the data analysis, the researcher discussed the analysis and interpretations with the other researchers. In line with the deductive approach, prior research also guided the discussion of findings (Hsieh and Shannon, 2005). Supplemental Appendix 1 describes the employed coding scheme. In the following section, data examples are provided for readers to consider and evaluate these interpretations (Hammer and Berland, 2014).

Results

Experiences of data agency at interpersonal level

The interviews revealed that these pre-service teachers all were active users in many social media services and that they had been using social media for a long time. At the moment, all of them were using Instagram as well as more private messaging channels

such as WhatsApp and Snapchat. Most of them had also been using Facebook for several years, but apparently their interest in sharing personal data in Facebook had faded away and Instagram was described as the key platform for participation on social networks. Some of the students also mentioned Twitter, Tumblr, Jodel, TikTok, Messenger, Signal, LinkedIn, and YouTube as used social media services. Different social media sites and applications were part of their everyday activities and used for several purposes, as Olivia describes:

> I use Instagram to follow content from friends and the kind of celebrities and people I follow and then to produce my own content, and then a little bit for this kind of enjoyment, just watching something from there whenever. And to some extent also, like . . . for example, I read a lot of news and other information from Instagram. And Facebook . . . Well, from there I read some news, but it's more for my own enjoyment, I read about different things and get to know the topics of groups and then, some buying and selling things there as well. And then . . . Well, in Snapchat, I . . . I mainly use it to communicate, I don't really put anything there. For communicating with friends. And then the same on Whatsapp, I use to communicate with friends, close friends. Then I use Tumblr for my own . . . for my own, entertaining purposes and then a bit of LinkedIn for employment purposes.

The interviews of pre-service teachers revealed that they were protecting their privacy using a number of strategies identified by previous research. Students were using different tools to communicate with different groups of people and also adjusted privacy settings accordingly. Pre-service teachers were also very aware of the way in which they represented themselves for these different audiences. Lily explains,

> On Facebook, I have friends that are not so close. The closest are on Instagram, where I have a private profile, I am able to limit who can see, and I have been able to select a little from the beginning. Well, in Facebook, I could have also chosen not to take everyone as a friend, but I have had it longer, so there are more. There are just friends, people I know, and those I don't want to share any information with.

In addition to adjusting profile privacy settings, students were typically extremely careful about what kind of self-image they wanted to share for others. Stella elaborates, "It's more of surface scratching, there was a nice moment and a nice thing. But really, maybe that's just, I could say scenery building, I don't share, for example, that it was a really shitty day." Managing personal privacy also included withholding content, such as intimate personal information about oneself or intimate information about close others. Moreover, many of the students were reflecting their previous and current activities in terms of their future profession, like Robert illustrates: "I share such content that if and when it goes somewhere else, other than my own wall, then it will not cause me anything in my civilian life nor in working life." Many students also reflected changes in their practices and told us that they are not sharing personal content as much as before when they were young. Students had also googled their data as well as deleted content that they had shared when they were younger.

In terms of publishing content that involves other people, all the students highlighted social norms that they had negotiated with their friends. Caroline elaborates

how they had created social rules for making sure that everyone would be asked before publishing:

> My friends are those who ask for permission to publish. It is subject to permission and somehow, for me, it even creates a sense of belonging to a group or something like that, kind of that my existence is recognized [laughs]. Maybe a little too deeply said, but something like that. I think, maybe nowadays, it is a kind of way of showing that somebody means something to you and is somehow important.

Moreover, there were also deliberate attempts to collaboratively prevent potential context collapses by undertaking regulatory activities together, as Stella describes:

> When life situations have changed, it has decreased, for example, when one friend went to a certain kind of job, less of a certain kind of content was produced after that. Of course, it's a kind of image thing, you have to think about it from the point of view of work life. Although these contents have not been insane, still a little effort has been made to limit it.

While students had great trust in their friends, some of them had also experienced privacy violations, such as when their parents shared embarrassing photos of them. Isabella describes: "a relative, like a family member or something, has shared some pictures or told really openly about something that relates to me, so I was like don't, I don't like it." Samuel was also reflecting context collapse that he had experienced as teacher:

> When I was younger, I shared more and maybe it makes me regret it nowadays. For example, I just ended up in a situation where pupils found my old blog, you know, in which young 18-year-old Samuel had messed up something.

Samuel solved this tension using his own experience as an educational example when discussing media practices with his pupils: "That situation was handled in a way that we're trying to bring understanding of how people are changing and those things are kind of OK. That you need to be ready to bear consequences if you mess around."

Experiences of data agency at commercial level

The interviews of pre-service teachers also confirmed findings of many previous studies related to youths' approaches and understanding of commercial-level issues. When asked about reading the Terms of Service, all students with one exception admitted that they did not read them or they just quickly browsed them though. Decision-making was typically reflected in terms of experienced difficulties making sense of complicated legal language. Moreover, many of the students reflected how non-participation in important social activities was not a real option, as Lillian illustrated:

> Interviewer: Well, do you read user agreements of applications before installing them?
>
> Lillian: No. [laughs] It's very bad, but many times they use very difficult language. They are not very user friendly. They are long. I'm guessing that it is purposeful so that people wouldn't read them. I don't know if

> this is a little conspiracy theory, but like . . . there are so many applications and many times, when you join in, it's not the best time you could sit down and read through these agreements. And what we have discussed with the group of friends is that if you decide not to use that app based on what it does, what it collects from you, what it says in the user agreement, then, in the end, you're the only one who loses in it. First of all, this data is collected from everyone and we are just one of those individuals. If I don't use that app, then I will lose all the benefits it could provide. For example Whatsapp, if I decide that I don't want to use it because it's a bad app that is managed poorly, then I can't send a message to people other than via text message or email, which is just impractical. Then, a little nihilism strikes as to whether it even matters that these user agreements are terrible when they have to be accepted anyway.

In addition, some of the students experienced difficulties in abandoning services that had become an important part of their everyday social interaction. Brian reflected how habits may go ahead when new app privacy updates are announced:

> In Whatsapp, there is coming . . . Today, there was again a reminder that 15.5. it will change some privacy thing. I don't know if I feel like changing it when everything is on Whatsapp. It is something that you are used to. That's the biggest reason to use them . . . I have the feeling that you could just give up because you are under observation anyway. It just feels the same, when every website asks for cookies to be accepted, then in Whatsapp it starts to feel the same, that one may give that. In principle, I wouldn't like to give all my data, but are there any options left?

In general, the students were quite aware of how, where, and why personal data are generated online, as Adam described: "Larger companies in particular collect that data in a purely commercial sense so that it can be targeted and sold to those who need it." However, only two of the students took deliberate actions for the prevention of commercial-level tracking, such as preventing apps from using cameras, deleting cookies, or using a virtual private network (VPN). Moreover, many of the students recognized the lack of knowledge on how personal data are processed. Lillian describes,

> I don't think that any human does that. Probably it is processed in such a way that they have these algorithms and so-called robots that will do the job, so they gather that data and transform it into a form that artificial intelligence can handle it, and then based on that . . . well, I don't think that there are concerns that some people could have seen it, unless then something really suspicious has happened.

On the other hand, all students recognized being targeted with tailored content and advertising based on what one has previously done online. This caused mixed feelings, as Caroline reflects:

> In a way, I am not bothered by the fact that data is collected about me. I somehow think that it is good that they know about me so that they can provide things I may need. But maybe it will

> also become seemingly so that they offer something that is not really needed, but it will become such when you see it everywhere and then you start thinking that I will need that and that. And then, maybe my thought is that . . . In a negative sense, there could be such things such as political decision-making, like elections or such things. So how data collection can be used to influence, when it can be seen that certain kinds of people are living in certain kinds of areas, then you could probably target election advertising or propaganda and whatever, so maybe those are the biggest concerns, so that . . . One thinks that you are free to make some decisions, but in reality, someone has brainwashed you to make those decisions.

As the excerpt above illustrates, few of the students were concerned about the ways in which companies could profile, predict, and engineer an individual's decision-making, opinions or actions. Some of the students had also recognized being targeted with fake news. For example, Aaron told that his friend had shared political news that he suspected to be fake: "a story of the prime minister came some time ago, at least it was somehow tricked around." However, when reflecting commercial-level risks, students were mostly concerned about data breaches, leaks, and identity thefts, often demanding responsibility, transparency, and security from the companies, like Olivia explains:

> There should be really strong security systems and so on, and that personal data should be properly and actively protected, and that all users should be aware of how it is taken care of and what they are used for.

Andrew had also recently experienced commercial-level privacy violations, but he was not concerned about the consequences:

> A while ago, when there was a data leak on Facebook, I checked that my data was taken. But then again, I feel that if someone really wants to know my gender, my address- which I have changed quite a few times-, and then, for example, my email address, that someone will get them if one wants to. And if someone leaks that to China or India, then it will leak. If someone there makes a fake profile of me, then I will have a good laugh.

While many of the students cared about their commercial-level privacy, some of the students also felt that their data would not be used in a manner that would cause them harm in the future, as James reflects:

> I don't think that data about ordinary people has any effect. On the other hand, if you were a celebrity, politician or something else, and if something sensitive would be revealed, that information could be used to play with that data. But I can't really see any scenarios on how it could affect me.

Moreover, some of the students reflected that their lack of knowledge affected their perceived ability to change current situation, as Caroline elaborated: "I certainly don't know about these issues and thus, I am probably really carefree about these things, because I think I can't influence them, so it's like whatever."

After asking students to reflect on their existing practices and data strategies, the students were also challenged to ponder the development of media education. Many of the

pre-service teachers thought that children should be supported to develop their critical understanding and they should become more aware of risks of sharing personal data with other people, particularly unknown others. Olivia stated,

> . . . there can be basically anyone behind the name and that no personal information should be shared there, no exact addresses or phone numbers or anything else. And then, you have to really know online people before meeting them and especially the children have to be in touch with their parents before actually doing anything like that. And this kind of general caution and also in terms of fake news and other things, so I would emphasize criticality in terms of source and media.

Pre-service teachers also emphasized understanding of what is good and bad behavior in relation to sharing images and other information with and about other people. While most of the students reflected the development of media education at interpersonal level, at the end of the interview, some students pondered about the need for educating commercial-level issues. William reflected on this:

> Like in school or youth centers, it could somehow be approached in more detail, like what it means that data is collected about you, how it is used and what kind of things it contains. Somehow we are just telling children that, do not accept that they collect data, but it doesn't tell much. But that tells us that we just don't have knowledge about what they collect. I just noticed, like when you asked about what data is collected, that I didn't even know what to answer.

Similarly, many of the students told us that this interview was the first time that they were challenged to ponder their data actions and perspectives on data agency, as Adam elaborates:

> I had to rethink a little bit about how I perceive these things. And it got me really to think about some of these things that I have considered as self-evident. I probably don't have any questions for you, but it was a good discussion and a good package of questions, because it aroused new thoughts and sometimes, it also put my finger in the mouth, to ponder how I really think about these issues.

Discussion

This study explored Finnish pre-service teachers' conceptions and experiences of data agency in social media environments. The findings corroborate a number of findings in previous research on the data strategies and attitudes of the youth. This study also found clear agentic dispositions, skills, and volition related to pre-service teachers shaping their online personae to what they believe best satisfies the expectations of their future professional environment. The students, who are soon about to apply for teacher jobs, were actively reflecting and shaping their digital selves from the perspective of future employees and school communities, and also had created collaborative strategies to prevent social risks and potential context collapse.

While pre-service teachers shared similar strategies to protect their interpersonal privacy, their privacy awareness, protective strategies, and desire for agency differed from

each other at the commercial level. In line with previous research studies, many of the pre-service teachers saw the sharing of data with advertisers as an acceptable trade-off or as a necessary price to be paid to participate (Keen, 2020; Selwyn and Pangrazio, 2018; Tedre et al., 2020). Although participants recognized how personal data are created and sold within commercial contexts, there was hardly any recognition of the mechanisms of data-driven techniques used in analysis, simulation, and automation of everyday information flows. Participants were able to describe how they were being tracked, but showed little insight into how those traces were used for profiling, modeling, predicting, and influencing people. Recognition of negative consequences and ethical concerns of algorithmic governance—such as algorithmic bias and invisible processes of prioritization and marginalization—remained largely black-boxed. While few students were concerned about behavior engineering and recognized potential threats to democracy driven by surveillance, they also felt that lack of knowledge prevents them from interacting in a resistant manner.

At the interpersonal level, pre-service teachers' data agency relied on classical approaches to digitalization, but at the commercial level, data agency would require understanding of data-driven and ML-based automation. At the former level, the participants' concerns are closely related to basic questions about data, such as visibility of their online selves and traces they leave. Those concerns are close to people's everyday lives, and mechanisms like tracking, content filtering, and curation are relatively easy to understand. At the latter, commercial level, the mechanisms are much more remote from people's everyday experiences, and understanding them requires more than common sense. Education on basic artificial intelligence (AI)/ML concepts is needed for understanding how profiling, clustering, and modeling similar profiles works and how those can be used for dynamic content creation, prediction based on similar interests, and behavior engineering (Valtonen et al., 2019). It is needed for understanding how ML systems can analyze one's likes to predict very accurately things like political alignments (Conover et al., 2011), sexual orientations, or ethnic groups (Kosinski et al., 2013). It is needed for understanding how services use attention engineering to keep users engaged, how loyalty loops and involvement spirals are designed and implemented (Siebert et al., 2020), and how people's voting behaviors can be influenced by social media (Bond et al., 2012).

In line with results from previous research studies, the pre-service teachers saw themselves as responsible actors in educating future generations to navigate in complex new media environments. Students typically emphasized taking a proactive stance to achieve interpersonal privacy goals and called for critical thinking, awareness of social risks, as well as comprehension of long-term consequences of sharing personal data about oneself or others. While these concerns are relevant, the results of the study confirm that most of the pre-service teachers find it hard to think of ways to support children's data agency beyond the interpersonal level (Davis and James, 2013; Stoilova et al., 2020). Without a sophisticated understanding of commercial-level issues and related mechanisms of ML and data-driven approaches to automation, it is unlikely that these future teachers would be prepared to facilitate children and youth's agentive actions in data-driven society. Meanwhile, we are reaching the point in which there are quite pronounced differences in the support children and youth receive for building their data agency (Stoilova et al., 2020). As Finnish educational system is built on purposeful policy aimed at equity and a

high-quality education for all (Niemi and Lavonen, 2020), the development of in-service and pre-service teacher education is needed for preserving those leading educational principles. However, the question remains whether transformative efforts should focus only on education. Especially at a commercial level, we cannot overlook the misalignment between individual teachers responding to the risks that action of dominant technology companies make accessible to children. Arguably, educational efforts toward data agency should be supported by greater regulation of the technology sector, which itself should take greater actions to respect children's rights and well-being (Lupton and Williamson, 2017; Stoilova et al., 2020; UNICEF, 2018).

Limitations and future research

One limitation of this qualitative study is that the sample is small and it was not drawn at random. Moreover, interviews focused on students' interpretation of data strategies rather than direct observations. Accordingly, because the study design is not suited for generalization, we do not suggest that the findings are representative of all pre-service teachers in Finland. Therefore, future research is necessary to examine the experiences of a broader range of pre-service teachers. Nevertheless, this study provides initial insight into the decision-making processes that inform pre-service teachers' data agency and privacy strategies as part of their lived experience.

An important contribution of this study is that it also improves the understanding of pre-service teachers' concerns and gaps in knowledge regarding the wider socio-technical systems within which data agencies are situated. The results of the study revealed several consistencies between our findings and previous research on youth data practices and it also raises several important questions for teacher education. Accordingly, there is an evident need for teacher education that responds to the real needs and practices of the youth as well as helps pre-service teachers to broaden their understanding of (1) how data agency is afforded and constrained by different but interacting levels of privacy; (2) how data agency is mediated, augmented, produced, and regulated by statistical and ML techniques; and (3) what kind of social, political, and economic implications these socio-technical systems may drive. As Selwyn and Pangrazio (2018) argue, it would seem sensible to presume that supporting students' data agency might be more effective if carried out by groups and collectives rather than individuals. Accordingly, an important future line of research is to develop conceptual tools, technologies, and pedagogical models that support students to collaboratively frame and analyze problems in existing socio-technical systems and help them to understand the mechanisms of ML to envision and take informed actions toward alternative possibilities.

Authors' note

All authors have agreed to the submission and that the article is not currently being considered for publication by any other print or electronic journal.

Acknowledgements

The authors thank the January Collective for their altruistic support as well as for the original idea for this study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Suomen Akatemia/Academy of Finland (grant number 296799).

ORCID iDs

Henriikka Vartiainen https://orcid.org/0000-0001-6005-907X

Juho Kahila https://orcid.org/0000-0002-9913-0627

Supplemental material

Supplemental material for this article is available online.

References

Bandura A (2001) Social cognitive theory: an agentic perspective. *Annual Review of Psychology* 52: 1–26.

Bond RM, Fariss CJ, Jones JJ, et al. (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415): 295–298.

boyd d (2014) *It's Complicated: The Social Lives of Networked Teens*. New Haven, CT: Yale University Press.

boyd d and Marwick A (2011) Social privacy in networked publics: teens' attitudes, practices, and strategies—a decade in Internet time: symposium on the dynamics of the Internet and society. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128

Choi M, Cristol D and Gimbert B (2018) Teachers as digital citizens: the influence of individual backgrounds, internet use and psychological characteristics on teachers' levels of digital citizenship. *Computers and Education* 121: 143–161.

Conover MD, Gonçalves B, Ratkiewicz J, et al. (2011) Predicting the political alignment of twitter users. In: *Proceedings of the 2011 IEEE international conference on privacy, security, risk and trust*, Boston, MA, 9–11 October, pp. 192–199. New York: IEEE.

Data-Pop Alliance and Internews (2015) Beyond data literacy: reinventing community engagement and empowerment in the age of data. New York. Available at: https://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf

Davis K and James C (2013) Tweens' conceptions of privacy online: implications for educators. *Learning, Media and Technology* 38(1): 4–25.

Elo S and Kyngäs H (2008) The qualitative content analysis process. *Journal of Advanced Nursing* 62(1): 107–115.

Emirbayer M and Mische A (1998) What is agency ? *The American Journal of Sociology* 103(4): 962–1023.

Eteläpelto A, Vähäsantanen K, Hökkä P, et al. (2013) What is agency? Conceptualizing professional agency at work. *Educational Research Review* 10: 45–65.

Goodwin I, Lyons A, Griffin C, et al. (2014) Ending up online: interrogating mediated youth drinking cultures. In: Bennett A and Robards B (eds) *Mediated Youth Cultures*. New York: Palgrave Macmillan, pp. 59–74.

Hallinan B and Striphas T (2016) Recommended for you: the Netflix Prize and the production of algorithmic culture. *New Media and Society* 18(1): 117–137.

Hammer D and Berland LK (2014) Confusing claims for data: a critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences* 23(1): 37–46.

Hargittai E and Marwick A (2016) "What can i really do?" Explaining the privacy paradox with online apathy. *International Journal of Communication* 10: 3737–3757.

Hendricks VF and Vestergaard M (2018) *Reality Lost: Markets of Attention, Misinformation and Manipulation*. Berlin: Springer.

Hsieh H-F and Shannon SE (2005) Three approaches to qualitative content analysis. *Qualitative Health Research* 15(9): 1277–1288.

Jones LM and Mitchell KJ (2016) Defining and measuring youth digital citizenship. *New Media and Society* 18(9): 2063–2079.

Keen C (2020) Apathy, convenience or irrelevance? Identifying conceptual barriers to safeguarding children's data privacy. *New Media and Society* 24: 50–69.

Kitchin R (2012) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Klinger U and Svensson J (2018) The end of media logics? On algorithms and agency. *New Media and Society* 20(12): 4653–4670.

Kosinski M, Stillwell D and Graepel T (2013) Digital records of behavior expose personal traits. *Proceedings of the National Academy of Sciences* 110(15): 5802–5805.

Kramer ADI, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America* 111(24): 8788–8790.

Lapenta GH and Jørgensen RF (2015) Youth, privacy and online media: framing the right to privacy in public policy-making. *First Monday* 20(3): 5568.

Leaver T (2017) Intimate surveillance: normalizing parental monitoring and mediation of infants online. *Social Media and Society* 3(2): 707192.

Livingstone S and Yoo D (2018) More clarity brings more confusion: debating what the European General Data Protection Regulation means for children in the UK. Available at: https://blogs.lse.ac.uk/medialse/2018/01/18/more-clarity-brings-more-confusion-debating-what-the-european-general-data-protection-regulation-means-for-children-in-the-uk/

Livingstone S, Stoilova M and Nandagiri R (2019) Talking to children about data and privacy online: research methodology. Available at: https://www.lse.ac.uk/media-and-communications/assets/documents/research/projects/childrens-privacy-online/Talking-to-children-about-data-and-privacy-online-methodology-final.pdf

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

McCosker A (2017) Data literacies for the postdemographic social media self. *First Monday* 22(10): 7307.

Mack L (2010) The philosophical underpinnings of educational research. *Polyglossia* 19: 5–11.

Marwick AE and boyd d (2014) Networked privacy: how teenagers negotiate context in social media. *New Media and Society* 16(7): 1051–1067.

Niemi H and Lavonen J (2020) Teacher education in Finland: persistent efforts for high-quality teachers. In: Lefty L and Fraser JW (eds) *Teaching the World's Teachers*. Baltimore, MD: Johns Hopkins University Press, pp. 153–178.

Pangrazio L and Selwyn N (2019) 'Personal data literacies': a critical literacies approach to enhancing understandings of personal digital data. *New Media and Society* 21(2): 419–437.

Raynes-Goldie K (2010) Aliases, creeping, and wall cleaning: understanding privacy in the age of Facebook. *First Monday* 15(1): 2775.

Rogoff B, Mistry J, Göncü A, et al. (1993) Guided participation in cultural activity by Toddlers and caregivers. *Monographs of the Society for Research in Child Development* 58(8): i+iii+v–vi+1–179.

Selwyn N and Pangrazio L (2018) Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data and Society* 5(1): 765021.

Siebert A, Gopaldas A, Lindridge A, et al. (2020) Customer experience journeys: loyalty loops versus involvement spirals. *Journal of Marketing* 84(4): 45–66.

Smirnova S, Livingstone S and Stoilova M (2021) Understanding of user needs and problems: a rapid evidence review of age assurance and parental controls. London. Available at: https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review-of-age-assurance-and-parental-controls/

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

Tedre M, Vartiainen H, Kahila J, et al. (2020) Machine learning introduces new perspectives to data agency in K-12 computing education. In: *Proceedings of the Frontiers in education conference (FIE)*, Uppsala, 21–24 October. New York: IEEE.

TENK (2019) The ethical principles of research with human participants and ethical review in the human sciences in Finland Finnish National Board on Research Integrity TENK guidelines 2019. Available at: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf

Tufekci Z (2008) Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society* 28(1): 20–36.

The Finnish National Board of Education (2014) *Perusopetuksen opetussuunnitelman perusteet 2014* [Finnish national core curriculum for basic education]. Available at: https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf

UNICEF (2018) *Children's Online Privacy and Freedom of Expression*. New York: UNICEF.

Valtonen T, Hoang N, Sointu E, et al. (2021) How pre-service teachers perceive their 21st-century skills and dispositions: a longitudinal perspective. *Computers in Human Behavior* 116: 106643.

Valtonen T, Tedre M, Mäkitalo K, et al. (2019) Media literacy education in the age of machine learning. *Journal of Media Literacy Education* 11(2): 20–36.

van Dijck J (2013) 'You have one identity': performing the self on Facebook and LinkedIn. *Media, Culture and Society* 35(2): 199–215.

Vuorikari R, Punie Y, Carretero S, et al. (2016) *Digcomp 2.0: The Digital Competence Framework for Citizens. Update Phase 1—The Conceptual Reference Model*. Luxembourg: Luxembourg Publication Office of the European Union.

Wertsch JV (2007) Mediation. In: Daniels H, Cole M and Wertsch JV (eds) *The Cambridge Companion to Vygotsky*. New York: Cambridge University Press, pp. 178–192.

Zuboff S (2015) Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1): 75–89.

Author biographies

Henriikka Vartiainen (PhD, Ed) is a senior researcher in the School of Applied Educational Science and Teacher Education, University of Eastern Finland. Her research interests include participatory learning, data agency, and technology in education.

Lotta Pellas (Master of Arts) majored in culture research with a specialization in media culture and communication. In her MA thesis, she examined social media influencers' discursive expressions of power in web articles.

Juho Kahila received MEd degree from the University of Jyväskylä, Finland. He is currently pursuing PhD degree in the School of Applied Educational Science and Teacher Education, University of Eastern Finland. His research interest includes children's learning related to digital games and metagame.

Teemu Valtonen is a Professor of Education in the University of Eastern Finland. His research interests include the use of information and communication technology in education, targeting, especially pre-service teachers' skills, and readiness to use ICT in education. In addition, his current research interest includes integration of the flipped classroom model into higher education teaching practices.

Matti Tedre is Professor of Computer Science, especially computing education and the philosophy of computer science, in the University of Eastern Finland.