The Data Awareness Framework as part of Data Literacies in K-12 Education

Lukas Höper* **and Carsten Schulte**

Department of Computer Science, Paderborn University, Paderborn, Germany

(accepted for publication on 22 November 2023, article published online on 18 December 2023, issue published on 5 June 2024)

Author Accepted Manuscript of the paper:
Höper, L. and Schulte, C. (2024). The data awareness framework as part of data literacies in K-12 education, *Information and Learning Sciences*, Vol. 125, No. 7/8. https://doi.org/10.1108/ILS-06-2023-0075

Abstract

Purpose: In today's digital world, data-driven digital artefacts pose challenges for education, as many students lack an understanding of data and feel powerless when interacting with them. This article addresses these challenges and introduces the data awareness framework. It focuses on understanding data-driven technologies and reflecting on the role of data in everyday life. The paper also presents an empirical study on young school students' data awareness.

Design and methodology: The study involves a teaching unit on data awareness framed by a pretest-posttest design using a questionnaire on students' awareness and understanding of and reflection on data practices of data-driven digital artefacts.

Findings: The study's findings indicate that the data awareness framework supports students in understanding data practices of data-driven digital artefacts. The findings also suggest that the framework encourages students to reflect on these data practices and think about their daily behaviour.

Originality: Students learn a model about interactions with data-driven digital artefacts and use it to analyse data-driven applications. This approach appears to enable students to understand these artefacts from everyday life and reflect on these interactions. The work contributes to research on data and AI literacies and suggests a way to support students in developing self-determination and agency during interactions with data-driven digital artefacts.

Keywords: K-12, Computing Education, Data Awareness, Data Literacy, Big Data, Datafication, AI Literacy

*Corresponding author can be contacted at: lukas.hoeper@uni-paderborn.de

1. Introduction

Education about data and data-driven technologies has been advocated in various fields in recent years, particularly with the development and omnipresence of machine learning (ML) applications. This has led to the emergence of numerous approaches and perspectives, for example, the concept of data literacy (e.g., Ridsdale et al. 2015, Wolff et al. 2016) and related concepts of AI literacy (e.g., Long & Magerko 2020), data agency (Tedre et al. 2020), and algorithm literacy (e.g., Dogruel 2021). This emphasises addressing multiple literacies rather than a single, definitive literacy for data and data-driven digital artefacts (ddA).

Researchers argue for supporting students in perceiving their role within data-driven societies (e.g., as data sources) and in relating learning about data and ddA to their everyday lives (Gebre 2018, Bilstrup et al. 2022, Pangrazio & Selwyn 2019, Sander 2020). However, some studies indicate that students often perceive data as impersonal and unrelated to their personal lives (Gebre 2018, Bowler et al. 2017). Moreover, studies indicate that many students lack awareness and understanding of where, how, and why data about them are collected and processed (e.g., Bowler et al. 2017, Bucher 2017, Pangrazio & Selwyn 2020, 2019, Tedre et al. 2020, Goray & Schoenebeck 2022). Hence, researchers argue for the need of examining students' understanding of data practices of data-driven applications (e.g., Gebre 2018, Pangrazio & Selwyn 2018, 2020, Goray & Schoenebeck 2022, Acker & Bowler 2018, Sander 2020).

This paper presents the *data awareness framework* as a specific spot within a continuum of approaches for teaching about data and ddA in K-12. We believe that considering and understanding the role of data within interactions with ddA is a central aspect of data literacies. Hence, data awareness addresses a strand of data literacies that focuses on raising students' awareness of personal data and data practices (e.g., Pangrazio & Selwyn 2019, Gebre 2022, Acker & Bowler 2018) and combines it with concepts of AI literacy and data agency (e.g., Tedre et al. 2020, Long & Magerko 2020).

The framework focuses on interactions with ddA that adapt their outputs according to data models generated by data collection and processing (e.g., social media applications, streaming services, and search engines). It is based on the idea that people typically focus on their immediate goals during these interactions, like sharing personal experiences with friends, finding movies to watch, or searching for information about something. With their attention primarily focused on these tasks, the role of the data is often neglected. Data awareness means shifting the focus to data practices during these interactions, which is often perceived as difficult and without immediate benefit. The overall goal is to empower students to navigate in a data-driven world, make informed and reflected decisions, and actively manage these data flows.

This is an important educational goal as several studies reported that many people develop feelings of powerlessness, apathy, or resignation (e.g., Lutz et al. 2020, Keen 2020, Hargittai & Marwick 2016). Even if they are concerned about the collection and processing of personal data (Sander 2020, Dowthwaite et al. 2020), they feel unable to counteract these data practices (Dowthwaite et al. 2020, Bilstrup et al. 2022). The idea of the data awareness framework is to provide students with an easy to understand model of interactions with ddA and related data practices enabling them to shift their focus on essential elements of the role of data in everyday interactions with ddA, and to analyse ddA and reflect on these interactions accordingly. Hence, data awareness aligns with arguments for supporting students to relate learning about data and ddA to everyday life. Thus, the framework is meant to teach about data and ddA in a way that helps students overcome powerlessness when interacting with ddA, that is, to support students in developing self-determination and agency in a data-driven world.

Aside from introducing the data awareness framework, in this paper, we report on an empirical study among 11-13 years old students. First, we examine students' understanding of data practices during interactions with ddA before and after an intervention that implements the framework (i.e. whether students have learned and can apply the framework's model). Second, we examine whether the intervention encourages students to reflect on the role of data within these interactions, as opposed to resignation regarding ddAs' data practices. By doing so, we aim to evaluate the effectiveness and practical applicability of the framework in supporting individuals to become more aware of and reflect on the role of data within their interactions with ddA.

2. Related Work

In this section, we discuss research on students' understanding of ddAs' data practices, their privacy concerns, and feelings of powerlessness when interacting with ddA, which serves as foundation for this study. We then discuss approaches for teaching data and ddA to locate data awareness within their intersections.

2.1. Students' Understanding of Data Collection and Processing

Recent research examined how people understand the role of data within interactions with ddA. We discuss and analyse crucial results by focusing on awareness and understanding of (a) data collection and (b) data processing in ddA.

Goray & Schoenebeck (2022), for instance, found that some youth were aware of certain types of data collection, like about their interactions. However, Keen (2020) found that while some students had ideas of collected data, they had no concrete ideas of their secondary use or consequences of data practices. Similarly, Bowler et al. (2017) reported that while some students knew that data is collected in some contexts, they did not adequately understand which data and what happened after it was collected. Dowthwaite et al. (2020) reported that, in the context of online platforms, students could not adequately give examples of data collection. They also struggled to perceive the implications of the use of the data. During interviews with children, Sun et al. (2021) found that they were mostly unaware that data companies collect personal data and how they use them; instead, they perceived collecting personal data rather as an interpersonal issue. Further studies found that many students lacked an understanding of personal data and how it was generated or used in social media and could not assess the impacts of data practices (Pangrazio & Selwyn 2020, 2018). Tedre et al. (2020) conducted workshops about ML and reported that students lacked an understanding of (personal) data collection and their usage. In the context of fitness trackers, Gabriele & Chiasson (2020) reported that while some participants had ideas of which data was collected, they could not assess the purposes of using the data. When interviewing social media users, Bucher (2017) found that some participants were aware that data are collected and algorithms are used for data processing. However, Eslami et al. (2015) found that most users were unaware that social media use curation algorithms.

These results are interpreted in Figure 1. No study reported that participants had good awareness and understanding of data collection by ddA, and all studies report on lacking understanding of data processing and its purposes. This underscores the importance of the topic for education. The existing studies often describe findings regarding data collection separately from those regarding data processing. However, this empirical research relies on different concepts of how to uncover students' understanding of ddA and their data practices. It is rarely characterised what it means to understand the data collection or the processing and its purposes. The studies do not systematically distinguish between types of data collection or kinds of data processing purposes. In the framework presented below, we make such a distinction and examine students' understanding accordingly in the empirical study of this paper.

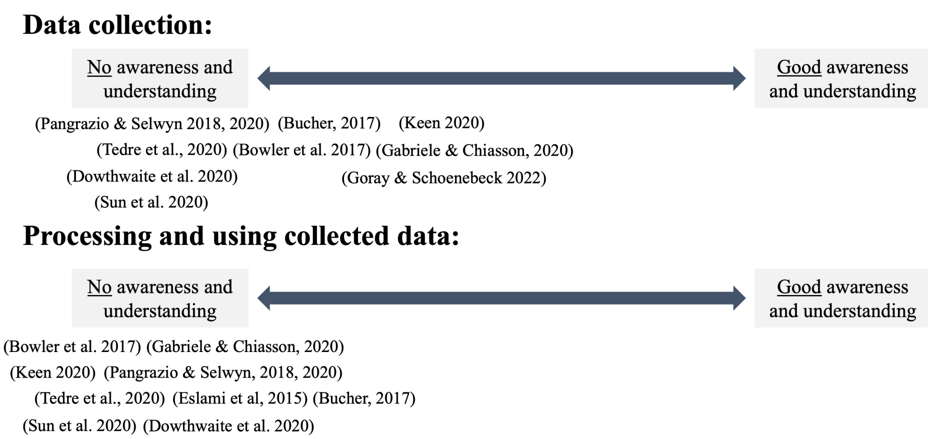

FIGURE 1: Overview of research on students' understanding of data collection and processing. (*Source:* Authors' own work)

2.2. *Students' Privacy and Feelings of Powerlessness*

If students have little or no understanding of ddAs' data practices, it raises the question of how they interact with ddA in everyday life. This section provides a brief overview of research findings on privacy and powerlessness.

The privacy paradox addresses the relation of one's concerns and habits: Although people generally value their privacy, they often disclose personal information (e.g., Acquisti & Grossklags 2005, Barnes 2006). The privacy calculus idea suggests that it derives from weighing individual benefits and privacy consequences (Dinev & Hart 2006), which may go beyond a rational consideration (Williams et al. 2017). Hence, people disclose personal information when the perceived benefits outweigh the assumed privacy consequences (Lutz et al. 2020).

In several studies, researchers examined students' concerns about their privacy and found that students are concerned about data practices of digital services (e.g., Sander 2020, Dowthwaite et al. 2020). Livingstone et al. (2019) conducted a literature review of students' understanding of privacy in different contexts and reported that school students struggle to assess data flows and their consequences, especially in commercial contexts.

Even students who understand basic concepts about data flows, tracking, and data-driven practices struggle to perceive their role as data sources when using data-driven applications (Gebre 2018), to interpret these concepts on a concrete personal level (Bowler et al. 2017), and to develop a critical stance towards these data-driven practices of their everyday lives (Vartiainen et al. 2021).

Moreover, studies indicate that students see no need for action regarding data practices of data-driven systems (Keen 2020). As reported in some studies, students tend to develop feelings of apathy, resignation, powerlessness, or having no control regarding the data practices of ddA providers (Sander 2020, Hargittai & Marwick 2016, Keen 2020, Lutz et al. 2020). Several possible reasons are discussed in the literature, such as lack of knowledge and control over personal data, users' incapabilities regarding respective actions, and their belief that privacy-protecting behaviour is useless (Lutz et al. 2020, Hargittai & Marwick 2016).

In summary, many students are concerned about ddAs' data collection and processing but feel powerless to counteract and control it (e.g., deciding whether to disclose certain personal data or making specific settings in a application). They struggle to understand data practices of ddA and encounter challenges that hinder their agency in making informed and self-determined decisions. Educational approaches are therefore needed that support students to understand ddA and their data practices in ways that help them to make informed decisions and develop self-determination and agency (i.e. as opposed to developing powerlessness) – as the work presented here aims to address.

2.3. Related Approaches

In recent years, several approaches for teaching about data, data practices, and data-driven technologies have been developed, discussed, and refined. Data literacies, for example, are mostly described as referring to abilities or skills for understanding and handling data, working with its representations, and communicating and reasoning about data (e.g., Pangrazio & Sefton-Green 2020, Ridsdale et al. 2015, Wolff et al. 2016, Gebre 2022, Calzada Prado & Marzal 2013). For instance, it encompasses collecting and analysing data to gain insights following inquiry workflows (e.g., Wolff et al. 2016). Aside from focusing on teaching technical skills and data-based inquiry processes (see characterisation in: Gebre 2022), other perspectives on data literacies are suggested: Critical data literacies include critical considerations of data systems, their data practices and implications (Bilstrup et al. 2022, Sander 2020, D'Ignazio & Bhargava 2015), personal data literacies focus on personal data and critically reflecting on respective data practices (Pangrazio & Selwyn 2019, Pangrazio & Sefton-Green 2020), and creative data literacy includes citizenship and empowerment for data-based engagement (D'Ignazio 2017).

In relation to AI technologies, the concept of AI literacy has been developed and discussed. It aims at enabling students to understand and use AI technologies and become creators of such technologies (Casal-Otero et al. 2023, Touretzky et al. 2019). Long & Magerko (2020) characterise AI literacy as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (p. 2). This includes understanding underlying AI concepts, evaluating impact of AI, and using AI as a tool (Casal-Otero et al. 2023, Long & Magerko 2020). Teaching approaches emphasise, for example, that students learn about AI and ML concepts and develop skills for designing AI applications (Marques et al. 2020, Casal-Otero et al. 2023).

Data agency extends data literacies and refers to “people’s volition, skills, attitudes, and capacity for informed actions that make a difference in their digital world” (Tedre et al. 2020, p.1). It aims at empowering students to actively participate in data-driven societies as makers and producers. For example, this also includes understanding ML concepts and developing ideas of ML applications (Vartiainen et al. 2020).

In addition, some approaches focus on algorithmic decision-making. For instance, algorithm awareness refers to recognising when algorithms are used in different contexts, like content curation algorithms (Eslami et al. 2015, Brodsky et al. 2020), or algorithm literacy which includes understanding, evaluating and using algorithms (Dogruel 2021).

These approaches have different focuses on competent handling of data, understanding data and algorithms, understanding AI and ML concepts, and designing and developing AI applications. Research on them reports on contributions to enabling students to inquiry data, make data-based arguments and develop ideas of ML applications. However, it remains unclear how students can relate these learning experiences to their everyday interactions with ddA and utilise their knowledge and skills to understand ddA and the data practices in order to make reflected, informed, and self-determined decisions. As several researchers emphasise supporting relations to students’ everyday lives (e.g., Bilstrup et al. 2022, Gebre 2018, Pangrazio & Selwyn 2019, McCosker 2017, Bowler et al. 2017, Sander 2020, Vartiainen et al. 2021), further research is needed to gain nuanced insights into how teaching about data and data-driven technologies can support students to develop an understanding that is usable in everyday interactions with ddA. The data awareness framework is meant to address this need. It aims at providing students with a conceptual understanding of ddA and data practices that is meaningful and useful for understanding ddA and reflecting on their daily interactions with ddA. While the most approaches focus on skills for carrying out data projects or designing ML applications, data awareness enables students to analyse and reflect on ddA and their data practices by using a model (see Figure 2). Thus, the framework starts to consider ddA from individual perspectives, but can be extended to critical reflections on societal issues, as suggested by critical data literacies. Overall, the framework incorporates objectives from different approaches and particularly contributes to the data literacies strand focusing on personal data (e.g., Pangrazio & Selwyn 2019, Acker & Bowler 2018, Gebre 2022).

3. Approach: Data Awareness Framework

In this section, we present the data awareness framework, including its model of the role of data in interactions with ddA which is intended to be used by students to understand ddA’s data practices and reflect on their daily interactions with ddA.

3.1. Role of Data as Facets of the Framework

First, we characterise the role of data during interactions with ddA as essential facets of the framework, which were indirectly used to analyse the empirical studies beforehand and build the foundation for the study presented in this paper.

Types of data collection Providers of ddA are particularly interested in personal data, e.g., using it to predict users' behaviour and make targeted recommendations to provoke that behaviour. They collect data through various means (e.g., Livingstone et al. 2019, Pangrazio & Selwyn 2019). As summarised by OECD (2014), data can be categorised into four types: *Provided data* is created actively by the user; *observed data* is gathered by observation and recording, of which the user is not necessarily aware; *derived data* is generated by processing existing data; *inferred data* is generated by probability-based processing. Hence, users intend to provide data that is collected, while other data is collected through observation, tracking or generated through data processing which goes alongside users' actions.

Hence, our framework distinguishes explicit and implicit data collection. *Explicit data collection* refers to data users intentionally provide when using ddA (provided data). Users are usually aware of this data collection because they actively disclose it. *Implicit data collection*, in contrast, refers to data collected through observation, tracking, and data processing that occurs alongside the user's action (observed, derived, and inferred data). Users are usually unaware of it as they do not intentionally provide this data.

The discussion of prior research in Section 2.1 suggests that students often struggle to recognise the implicit data collection, but some students seem to recognise explicitly collected data.

Kinds of data processing purposes Collecting and using personal data is typically driven by economic motives (Mayer-Schönberger & Cukier 2013, Pangrazio & Sefton-Green 2020, Zuboff 2019). Users' experiences are utilised as free raw material to predict their future behaviour and serve commercial interests. In this sense, Zuboff (2019) argues that providers of ddA do not collect personal data to offer services; instead, they offer services to collect such data. This leads to information and power asymmetries between commercial actors and users (West 2019), partly due to the challenges of understanding the purposes behind data practices (Brunton & Nissenbaum 2015).

The collected data are used and processed by ddA to provide features and generate outputs, which is more evident for users. We call them *primary purposes* of using and processing collected data, which are usually described from users' perspective. For example, a search engine provider uses the collected data to display results. However, from providers' perspective, data processing serves additional purposes beyond immediate output generation, such as predicting future behaviour, adapting features accordingly or using the data in other contexts (e.g., Tufekci 2014, Zuboff 2019). For instance, Google provides Captchas to distinguish humans from machines, but has also used the collected data to train ML models that have been temporarily used in military contexts (Denning & Denning 2020, Shane & Wakabayashi 2018). We call them *secondary purposes*, which are rather not apparent to users. Assessing secondary purposes in detail is often challenging, as they are sometimes obscured or overshadowed by primary purposes (Brunton & Nissenbaum 2015, Burrell 2016).

The discussed studies on students' understanding of ddAs' data practices suggest that students have limited understanding of these purposes, particularly regarding secondary purposes (see Section 2.1).

Data models about users Providers gather information about users and create *data models*, which serve as primary and secondary purposes. They are called as digital footprints, digital shadows, or digital doppelgänger (e.g., Bode & Kristensen 2016, Cheney-Lippold 2017, Tufekci 2014, Kitchin 2014). The collected or generated data constitute a model of the user that is continuously refined during these interactions, representing specific aspects of the user rather than being a 'copy' of the user. The data models can also contain predictions about users' behaviour (e.g., Zuboff 2019) and extract meaningful and sensitive information from seemingly unrelated data (e.g., Goray & Schoenebeck 2022, Mühlhoff 2021).

Research on peoples' understanding of data practices indicates that they struggle to comprehend the role of data within interactions with ddA (see Section 2.1). Consequently, they may lack an understanding of how data models about users are constructed.

3.2. *Perspective on Interactions with ddA*

The framework considers ddA embedded in interaction systems, as the role of data then becomes meaningful. This perspective is described in more detail in this section.

Interactions between users and ddA From an educational perspective, Schulte & Budde (2018) describe interactions between humans and digital artefacts as sequences of input, processing, and output. Users provide inputs, such as clicking on search engines results, which serves as a request to the ddA and a source of data collection (i.e. data about the clicked link). Using this data, the ddA can generate personalised outputs, such as the order of search results. By considering these interactions with ddA students are familiar with, we intend to increase students' engagement and support relations between understanding ddA and the role of data and the daily life. The framework uses this perspective to characterise the role of data in interactions with ddA. In this perspective, students can analyse and assess the role of data, changing outputs, and reciprocal influences between users and ddA.

Influences between users and ddA The interaction between a user and a ddA causes interrelations between them. Users influence ddA when explicitly generating data or performing actions that are implicitly collected, impacting the outputs and digital doppelgänger through data processing. Conversely, a ddA affects users (as discussed in: Tufekci 2014, Kramer et al. 2014, Rahwan et al. 2019, Susser et al. 2019). For example, they can effect behaviour and decision-making as provoked by nudging and other techniques of attention engineering. Beyond the individual perspective, massive data practices create societal issues. Phenomena like filter bubbles, targeted information, and misinformation can significantly influence many people and thus have broader implications on society (e.g., electoral processes) (Bond et al. 2012, Lazer et al. 2018). Zuboff (2019) and West (2019) characterise such effects within surveillance or data capitalism, highlighting their significance.

Considering ddA as embedded in interaction contexts can reveal interrelations on individual and societal levels. Reflecting on these interactions at an individual level is crucial for perceiving one's role within these interactions and making self-determined actions (e.g., to recognise effects such as by nudging). Extended to a societal level, these reflections could address ethical and societal issues arising from data practices. Thus, the framework starts at an individual level, supporting students to reflect on their daily interactions with ddA, and then explore and discuss societal issues.

3.3. *Summarising the Data Awareness Framework*

The data awareness framework is characterised by considering interactions between human and ddA and focusing on the role of data within these interactions. Personal data are collected explicitly and implicitly when interacting with ddA; these are used and processed for primary and secondary purposes, often creating and using a data model of the user (e.g., a digital doppelgänger). These aspects are the central facets of the framework's model as summarised in Figure 2.

Data awareness is located at the intersection of existing approaches, particularly those for data literacies that emphasise awareness of personal data and reflection on data practices (e.g., Pangrazio & Selwyn 2019, Tedre et al. 2020, Acker & Bowler 2018, Sander 2020, Dowthwaite et al. 2020, Gebre 2022), but also addresses understanding of data-driven technologies as mentioned in approaches for AI literacy and data agency (see Section 2.3). The framework's perspective on individual interactions with ddA could be extended by critical reflections on societal issues of data practices in a datafied world (as highlighted by critical data literacies (e.g., Sander 2020, Bilstrup et al. 2022)).

Data awareness goes beyond the vague idea that some data collection is happening and personal data is somehow processed for some purpose. One main goal of data awareness is to empower students to understand ddA and their data practices, and reflect on these interactions. In doing so, it provides a model of interactions with ddA which is meant to be usable for students in everyday life (see Figure 2). We intend to support them in identifying choices for actions and making informed and reflective decisions for their everyday interactions with ddA. Therefore, the data awareness framework aims at fostering students' self-determination and agency when interacting with ddA in a digital and datafied world. Data awareness can thus be defined as:

> *When interacting with ddA, data awareness means being able to become aware of the explicit and implicit data collection, the primary and secondary purposes, and the role of data models about oneself (e.g., digital doppelgänger), as well as one's own role in these interactions.*

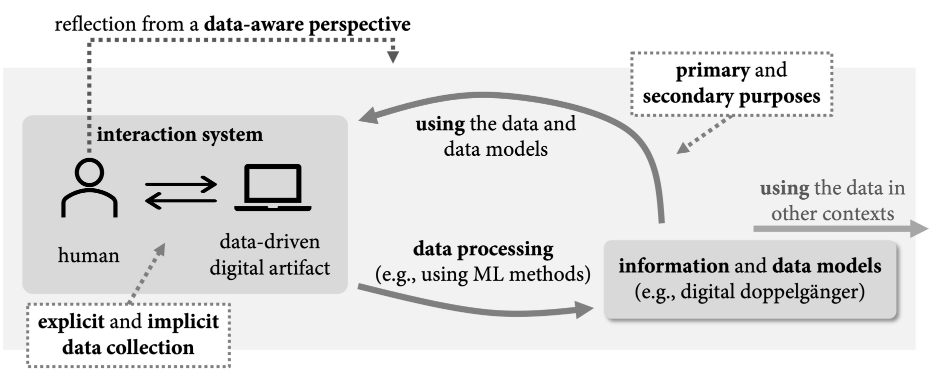

FIGURE 2: Model for describing the central facets of the data awareness framework. (*Source:* Authors' own work)

4. Method

In the following, we present an empirical study on school students' data awareness and how it can be fostered.

4.1. Design and Participants

The study is part of a larger design-based research project focused on developing and evaluating the data awareness framework. This study concentrates on students' understanding of and reflections on data practices of ddA from everyday life. We developed a teaching unit implementing the framework to examine students' data awareness. The study aims at examining (1) students' understanding of the framework's facets, particularly the types of data collection and the kinds of data processing purposes, and (2) the extent to which the framework encourages reflection on the role of data in interactions with ddA. Hence, we address the following research questions:

- *RQ1.* To what extent does the teaching unit enable students to comprehend and describe data collection and data processing purposes within an exemplary interaction with a ddA?
- *RQ2.* To what extent does the teaching unit encourage students to think about where, how, and why ddA collect and process data about them?
- *RQ3.* If students can describe both types of data collection and kinds of its processing purposes by an exemplary ddA, do they think more often about where, how, and why ddA collect and process data about them?

In this study, we implemented a pretest-posttest design. A questionnaire was used to survey students' data awareness (pre-test). Students then attend a data awareness teaching unit (see Section 4.3). Finally, a similar questionnaire was used as post-test. The study focused on young students in K-12 education. This design allows measuring students' data awareness, respective improvements, and effects during the teaching unit. The operationalisation of data awareness and the questionnaire is described in Section 4.4.

4.2. Procedure for the Intervention and Data Collection

The ethics committee of the authors' institution granted approval for this study, especially assessing the study design concerning children's privacy and self-determination and the data protection regulation. Student participation was voluntary and based on informed consent.

The teaching unit involved 147 students (11-13 years old) of seven sixth grade classes from three schools. The intervention took place in a computer science subject that is mandatory in the local curriculum. Prior to the study, students had not been taught about data or AI concepts. One of the authors taught one class, while experienced computer science teachers taught the others. We instructed the teachers on the framework and the intervention beforehand. Afterwards, we discussed the teaching unit and materials with them. The teaching unit comprised four to six lessons of 45 minutes each. Pre-test and post-test data could be matched for 123 students (71 female, 49 male, and three non-binary students).

4.3. Teaching Unit

The teaching unit consists of three parts, drawing on cellular networks' collection and processing of location data. We chose this context to provide an entry point for considering interactions with ddA, without too much technical complexity.

In the **first part**, students learn about the collection of location data and the primary purpose of its use in cellular networks. Students watched a video about cellular networks explaining their main components and functionalities concerning how a call between two mobile phones is established. Using a puzzle as a model for cellular networks, students simulate and explore the networks' functions (see supplementary materials). They reflect on the simulation by asking how the cellular network 'knows' where the receiver of the call is or to which base station the receiver is connected. Students learn that collecting location data (*implicit data collection*) in addition to other *explicitly collected data*, like phone numbers, is necessary to establish calls (*primary purpose*).

The **second part** addresses the questions what it means to collect location data and which insights can be obtained by analysing location data from a person. In groups, students explore publicly available location data from a cellular network provider and characterise the underlying person by creating a profile representing a '*digital doppelgänger*' (*secondary purpose*). For this, we developed a web-application that allows students to analyse the data and make conclusions by filtering the timestamps and visualising the location data on a map (see supplementary materials). For example, when trying to analyse where the person probably has his/her home, they filter the location data by selecting all data with timestamps between 2am and 4am. The respective data are visualised on the map. The students can explore the data points on a map and find some clusters of location data in the selected time range. Based on this they make interpretations where the person probably lives. By exploring several other personal aspects chosen by the students, they create a profile of the person. After this exploration phase, they present and discuss their interpretations in class and reflect and evaluate this interaction context on individual and societal levels.

The concepts of explicit and implicit data collection, primary and secondary data processing purposes, and digital doppelgänger were introduced and used in the unit's example within the first and second parts. The **third part** then concentrates on applying this knowledge to other ddA. Students are asked to identify other everyday interactions with ddA in which location data about them are collected and processed. After brainstorming such situations in class, groups of three students chose an example. In these groups, they analyse the selected ddA regarding the framework's model. Possible ideas include maps services or social media apps. Students then present, discuss, and evaluate their results in class. The discussion is prompted by some questions: Students are asked to discuss the extent to which they (as users) have a choice about what data (especially location data) is collected about them, and whether it is possible to make a change in these situations. We emphasise that students discuss from their personal point of view to ensure no opinions about rules for behaviour are prescribed. Students are intended to reflect on the role of data in the exemplary interactions with ddA and take a position regarding these issues.

This teaching unit serves as a simple introduction to enhance students' data awareness, allowing them to analyse the role of data in interactions with ddA regarding the framework's model. We intend to provide a perspective on ddA, which is useful for students when utilising their knowledge from the lessons to everyday interactions with ddA and reflecting on these interactions.

4.4. Instrument for Data Collection

According to the research questions, we designed a questionnaire to examine whether students can become aware of the role of data within interactions with ddA (RQ1 and RQ3). To assess students' understanding of data collection and processing by ddA, the **first part** of the questionnaire has open-ended items asking students to describe the data collection and processing by a given example of a ddA. We chose the situation of interacting with a messenger app (e.g., sending a message) because it is likely to be a familiar context for students. All the items of the first part refer to this example. The context is not covered within the teaching unit. Therefore, students are not simply asked to repeat what they have learned in class but have to apply their understanding of ddAs' data practices to another situation.

The questionnaire also aims to assess students' engagement in reflecting on data collection and processing during interactions with ddA (RQ2 and RQ3). We propose that reflecting or thinking about data collection and processing when interacting with ddA indicates a higher level of data awareness. Hence, the **second part** of the questionnaire consists of two scales: "reflection on data collection and processing" (3 items, four-point Likert scale) and "try to disclose few personal data" (1 item, four-point Likert scale). These scales assess whether students reflect on these processes and relate learning about ddA to their everyday interactions. As data awareness is not intended to prescribe behaviour like not disclosing personal data, the single-item scale allows to examine whether students think about their interactions. Due to the young age of the participants, we designed these scales as concisely as possible. The four-point Likert scales range from "not applicable" to "fully applicable". The items were developed based on literature and the data awareness framework. Table 4 gives an overview of the scales and their mean, standard deviation, and reliability.

Moreover, we added items asking about gender and ID to compare the pre-test and the post-test.

4.5. Data Analysis

As students' answers to open-ended items were mainly only short phrases and keywords (e.g., "the time of the message"), we decided to conduct only the first steps of qualitative content analysis by focusing on quantifying the coding and analysing the quantified data (Kuckartz 2014). The coding scheme includes deductively defined codes for the framework's facets of the data collection and processing by ddA (see Table 1). We used the code 'unclear' if an answer was ambiguous. While one author coded all qualitative data, another coded 50%. We then held an intercoder session as part of consensual coding to resolve problematic coding and improve intercoder reliability (Kuckartz 2014). Next, we refined the coding scheme, calculated the intercoder reliability, and analysed the quantified data.

We analysed the quantitative data by calculating scales' reliabilities, correlations, and differences between participant groups and between pre- and post-test data. The scale "reflection on data collection and processing" is interval scaled but is not normally distributed according to the Shapiro test. The single item scale "try to disclose few personal data" is ordinarily scaled. The pre- and post-test data are dependent. We calculated Cronbach's alpha to test the reliability of the reflection scale and tested differences between pre- and post-test by using Wilcoxon signed-rank test (two-sided with $\alpha = 0.05$). To examine differences between groups of participants, we used the Mann-Whitney U test (two-sided with $\alpha = 0.05$). We also calculated Spearman correlations (undirected variables) to test correlations. The effect size r of test results was classified as small between 0.1 and 0.3, medium between 0.3 and 0.5, and large above 0.5 (Cohen 1992).

5. Results

5.1. Describing Data Collection and Processing (RQ1)

The initial intercoder reliability was $\kappa = 0.68$. After discussing the coding, disagreements and coding errors were resolved, and the intercoder reliability changed to $\kappa = 1.0$, indicating a perfect strength of agreement (Landis & Koch 1977).

Almost all students (120 of 123) reported using messenger apps, so they are familiar with this interaction. Applying the coding scheme, 94 (76.4%) students described collected data, and 79 (64.2%) named data processing purposes in the pre-test. Afterwards, 99 (80.5%) students described collected data and 97 (78.9%) listed purposes. Table 1 shows code frequencies and examples.

TABLE 1

Results of coding the qualitative data ($N = 123$)

code	example	Number of students in the pre-test	Number of students in the post-test
explicit data collection	sent text/picture/location/...	65 (52.8%)	60 (48.8%)
implicit data collection	date and time of a message	47 (38.2%)	83 (67.5%)
primary purposes	communicate with text messages	14 (11.4%)	46 (37.4%)
secondary purposes	profiling, personalised advertising	66 (53.7%)	61 (49.6%)
unclear	(ambiguous answers)	37 (30.1%)	44 (35.8%)

Note. The numbers of students referring to the respective codes are not disjunctive within pre-test and post-test, as shown in the Tables 2 and 3. (*Source:* Authors' own work)

In the pre-test, 52.8% of students named at least one example of explicitly collected data, while 38.2% identified implicitly collected data. The difference between explicit and implicit data collection is significant in the pre-test ($W = 1116.5,\ z = -1.794,\ p = 0.0392,\ r = -0.1144$). In the post-test, 48.8% provided examples for explicit data collection and 67.5% of students mentioned implicit data collection. Again, the difference between explicit and implicit data collection is significant ($W = 448.0,\ z = -2.6979,\ p = 0.0020,\ r = -0.1720$).

Comparing pre- and post-test data showed almost no differences in naming explicit data collection. However, there are significant differences in the proportion of students who could identify implicitly collected data with a small effect size ($W = 247.5,\ z = -4.2621,\ p < 0.0000,\ r = -0.2717$). This indicates that the intervention positively affected students' ability to recognise and describe implicit data collection by ddA not covered during the intervention.

We developed a model of four levels to analyse and understand the relationship between describing explicit and implicit data collection: Students who described (1) no data collection, (2) explicit data collection only, (3) implicit data collection only, and (4) both types of data collection (see Table 2). In line with the characterisation of these facets in the framework (see Section 3.1), these categories are ordered hierarchically and thus represent an ordinal scale.

TABLE 2
Frequencies of the categories according to data collection descriptions

Data Collection	*Nothing described*	*Only explicit data collection described*	*Only implicit data collection described*	*Explicit and implicit data collection described*
Pre-test	29 (23.6%)	47 (38.2%)	29 (23.6%)	18 (14.6%)
Post-test	24 (19.5%)	16 (13.0%)	39 (31.7%)	44 (35.8%)

(*Source:* Authors' own work)

Students' categorisation based on the descriptions of data collection is significantly different between pre- and post-test with small effect size ($W = 727.5$, $z = -4.3928$, $p < 0.0000$, $r = -0.2801$). Significantly more students could describe both types of data collection after the teaching unit. Consequently, many students developed a broader view of data collection beyond the intervention's example, that is, they developed a conceptual understanding about types of data collection they could apply to this example.

Similarly, we examined students' answers about data processing purposes. In the pre-test, only 11.4% of students could describe at least one primary purpose, but 53.7% named a secondary purpose. Significantly more students could describe secondary purposes than primary purposes with medium effect size ($W = 513.5$, $z = -5.1152$, $p < 0.0000$, $r = -0.3261$). In the post-test, 37.4% of students named primary purposes, and 49.6% described secondary purposes (no significant differences).

Comparing pre- and post-test data showed almost no difference regarding secondary purposes. Nevertheless, the number of students, who described at least one primary purpose, has increased strongly. This pre-post-test difference is significant with small effect size ($W = 164.5$, $z = -4.1079$, $p < 0.0000$, $r = -0.374$). This indicates that more students paid attention to primary data processing purposes by the messenger app after the intervention.

Students' answers on data processing purposes were analysed quantitatively by categorising students into four groups: Students who described (1) no purposes, (2) secondary purposes only, (3) primary purposes only, and (4) both kinds of purposes (see Table 3). According to the characterisation of primary and secondary purposes (see Section 3.1), describing secondary purposes would indicate (regarding this aspect) a higher level of data awareness than describing primary purposes. However, the students in this study only named secondary purposes very nonspecifically (e.g., "personalised advertising"). These answers address a superficial notion of secondary purposes. Therefore, the categories listed above are ordered hierarchically and thus represent an ordinal scale.

Students' categorisation is significantly different between pre- and post-test with small effect size ($W = 452.5$, $z = -4.2883$, $p < 0.0000$, $r = -0.2734$). Significantly more students better understood the kinds of data processing purposes of a messenger app af-

TABLE 3
Frequencies of the categories according to data processing purposes descriptions

Data Processing	*Nothing described*	*Only secondary purposes described*	*Only primary purposes described*	*Primary and secondary purposes described*
Pre-test	44 (35.8%)	65 (52.8%)	13 (10.6%)	1 (0.8%)
Post-test	26 (21.1%)	51 (41.5%)	36 (29.3%)	10 (8.1%)

(*Source:* Authors' own work)

ter the intervention. Hence, many students developed a broader view of data processing purposes beyond the intervention's example. Thus, they developed a conceptual understanding of kinds of data processing purposes they could apply to this example.

Using students' categorisations (see Tables 2 and 3), we examined a relationship between descriptions according to these facets. The categories are moderately correlated ($p = 0.0002, r = 0.3271$). This suggests that students who can describe both types of data collection often also address both kinds of data processing purposes. Therefore, an adequate understanding of data collection relates to understanding data processing purposes of a ddA.

5.2. *Reflection on Data Collection and Processing (RQ2)*

We also examined whether the teaching unit impacts students in reflecting on data collection and processing by ddA. In the pre-test, students were asked whether they reflect on where, how, and why ddA collect and process personal data. In the post-test, they were asked whether they planned to do so more often. Scale's reliability is acceptable in the pre-test ($\alpha = 0.79$) and high in the post-test ($\alpha = 0.88$). The scale values indicate that most students agree on the items in both tests (see Table 4).

TABLE 4
Statistics of the questionnaire's scales

	pre-test				**post-test**			
scales	*mean*	*std*	*mdn*	α	*mean*	*std*	*mdn*	α
reflection of data collection and processing (3 items, e.g., "I often think about the purposes for which data collected about me is processed.")	2.34	0.78	2.33	0.79	2.68	0.93	2.67	0.88
try to disclose few personal data (1 item, "I try to ensure that digital artefacts only collect a few data about me.")	2.96	1.05	3.0	-	3.09	0.97	3.0	-

(*Source:* Authors' own work)

The pre-post-test difference is significant with small effect size ($W = 1497.5$, $z = -3.7688$, $p = 0.0002$, $r = -0.2433$). Consequently, significantly more students planned to reflect on ddAs' data practices after the teaching unit. This suggests that they are paying more attention to that and are probably willing to analyse ddA accordingly in everyday life.

Teachers' observations support this interpretation. They reported that students were excited about exploring the location data. Some even checked their mobile phones to see whether location data was being collected and used by apps and changed privacy settings. Therefore, students were interested in ddAs' data practices for location data and related what they had learned in class to their everyday lives.

In addition, students were asked whether they try to ensure that ddA only collect a few personal data as an exemplary behaviour that addresses whether students reflect on their behaviour. In the pre-test, this item was rated, on average, about three (scale ranges from 1 to 4). The mean value in the post-test is slightly higher than in the pre-test, but the difference is not significant. This indicates that many want to use (and configure) ddA before and after the teaching unit so that they only collect few personal data.

Moreover, we also analysed whether this is related to reflecting on data collection and processing by ddA. The scales of reflecting on data collection and processing by ddA and of trying to disclose few personal data are positively significantly correlated with a small effect in the pre-test ($p = 0.0016,\ r = 0.292$) and with a medium effect in the post-test ($p = 0.0001,\ r = 0.3527$). Thus, when reflecting on data practices, students preferred to disclose less personal data, indicating that reflecting on data practices is related to reflecting one's own behaviour.

5.3. Relationship Between Describing of and Reflecting on Data Collection and Processing (RQ3)

To determine whether describing data collection and processing purposes is related to reflecting on these processes, we divided the students into two groups based on their previous categorisations in Tables 2 and 3 (only post-test data): (1) Those who described both types of data collection (explicit and implicit) and at least a primary purpose of data processing and (2) all other students.

Of the 123 students, 25 students were assigned to the first group (G1), and 98 ended up in the second group (G2) (similar gender distribution). Table 5 gives an overview of the respective statistics of the scale for reflection on data collection and processing. On average, the first group rated the scale with a mean score of 3.07, while the second group rated it with a mean score of 2.58. This difference is significant ($U = 1573.5,\ p = 0.0162$). This indicates that students with a better understanding of ddAs' data practices are more likely to reflect on these processes in everyday life.

TABLE 5

Groups of students and respective scale statistics

Distinction regarding understanding of data collection and processing	**Numbers of students in the group**	**Scale "reflection of data collection and processing" (post-test)** *mean*	*std*	*mdn*
G1: higher level of understanding	25 (20.3%)	3.07	0.91	3.33
G2: lack of understanding	98 (79.7%)	2.58	0.91	2.67

(*Source:* Authors' own work)

6. Discussion

This section briefly discusses the findings of the study according to the research questions.

Developing an understanding about the role of data (RQ1) Related research indicates that students lack awareness and understanding of the data collection and processing by ddA (see Figure 1). However, these studies do not distinguish systematically between different types of data collection and kinds of data processing purposes. While the framework characterises types of data collection and kinds of data processing purposes, the presented study offers more nuanced insights for students' awareness and understanding of data practices.

Before the intervention, students could describe some explicitly collected data but significantly less implicitly collected data, which supports the mentioned results from prior research (see Figure 1). However, after the teaching unit, many students developed a broader view on data collection and identified implicitly collected data or even both types. Regarding the framework's model about the role of data, these findings indicate that many students seem to understand both types of data collection and could apply this conceptual understanding to an everyday life example not covered in class.

Similarly, before the teaching unit, students could imagine some secondary purposes of the ddA's data processing, but mostly unspecific, and only a few primary purposes. This aligns with research which revealed that students struggle with describing ddAs' data processing purposes (see Figure 1). After the intervention, significantly more students described primary purposes or even both kinds, indicating that the intervention enables students to apply a conceptual understanding of data processing purposes to an everyday life example.

Moreover, the results revealed that students' understanding of data collection by ddA relates to their understanding of the data processing purposes. Hence, understanding the data collection could be a prerequisite for understanding and assessing the data processing purposes (and/or vice versa).

Compared to related studies as summarised in Figure 1, the students' initial position on the figure's spectra aligns closely with those from previous research. However, the findings indicate a progression to the right side of the spectra in Figure 1. In consequence, this study contributes to research on how to enable students to comprehend data practices of data-driven technologies and how to measure such an understanding. The findings suggest that many students developed an understanding of ddAs' data practices. Hence, the framework's model about the role of data during interactions with ddA, as taught in the intervention, seems to be understandable and usable to analyse data practices of ddA from everyday life.

Reflecting on data practices of ddA (RQ2) Previous research highlighted that students often struggle to relate learning about data to their everyday lives and feel powerless regarding data practices (e.g., Sander 2020, Hargittai & Marwick 2016, Keen 2020, Lutz et al. 2020). In this study, many students stated that they rather reflect and think about ddAs' data practices in their everyday lives, and the teaching unit seems to encourage them to think about it more often. Additionally, some students now want to pay more attention to sharing less personal data when interacting with ddA, which relates to reflecting on such data collection and processing. These findings align with teachers' observations that some students were curious and wanted to find out immediately after class which apps on their mobile phones were allowed to collect location data. These results indicate that students may feel the need to reflect on their role in these interactions, control data flows, and change their behaviour. That contrasts with findings reported in literature, which suggest that students often did not perceive a need to act regarding data practices or feel powerless regarding the data practices (see Section 2.2). However, the results support observations made by Sander (2020), who also found that students were concerned and interested in learning about data practices.

In summary, the study indicates that students did not appear to resign or 'give up' facing ddAs' data practices. It rather suggest that the intervention may encourage students to reflect on these data practices more often and might let students think about their interactions (i.e. disclosing less personal data). Hence, we assume that the framework can encourage students to engage with data practices in their everyday interactions with ddA. This might mean that the framework could be a first step towards supporting empowerment and agency and countering powerlessness and resignation.

Relation between understanding and reflecting (RQ3) Pangrazio & Selwyn (2018) argue that teaching students to understand the data processes of data systems is an initial step for supporting students to address challenges for their privacy and agency posed by data practices of data systems (e.g., social media). Similarly, as data awareness includes developing a conceptual understanding of data practices during interactions with ddA, we argue that data awareness could support students (in a second step) in reflecting on their role within these interactions and in developing empowerment and agency. However, learning about data and ddA does not necessarily mean that students are also more engaged and reflect on these data practices in everyday life. For example, when teaching about data-driven technologies, Vartiainen et al. (2021) found in their study that even if children learned basic concepts and mechanisms of ML, they did not develop a critical stance towards data practices and data-driven technologies in their everyday lives. As the findings for RQ1 indicate a progression in students' understanding of ddAs' data practices, the question arises as to whether interventions that take such a first step, as suggested by Pangrazio & Selwyn (2018), also support in reflecting on these data practices.

The results indicate that students with a better understanding of data collection and processing by ddA are more likely to reflect on data practices in their everyday lives. In other words, a higher level of data awareness might support students in reflecting on their role in interactions with ddA, and hence, in taking a (critical) stance towards ddA in everyday life. Thus, data awareness might support them in making informed and reflective decisions when interacting with ddA (i.e. interacting more self-determinedly with ddA). That leads to the conjecture that data awareness may initiate the development of self-determination and agency in everyday interactions with ddA, that is, to counteract feelings of powerlessness.

7. Limitations

The study provides valuable insights but also has limitations, which we reflect on in this section. First, the questionnaire limits the findings. The study is primarily focused on examining students' understanding of and reflection on ddA's data collection and processing. A secondary interest was whether the framework also encouraged reflection on the personal role within these interactions. However, given the age of the participants, a concise questionnaire was preferred. While focusing on the primary aims of the study, we decided to use a single-item measure ('try to disclose few personal data', see Section 4.4) to gain preliminary insights into whether students think about their actions, contrasting to resignation and 'given up' attitudes. This item has limitations in terms of validity in measuring students' behaviour, as it focuses on one specific behaviour. In combination with the scale on reflection on data practices (which also correlates with the single item), this item can provide preliminary insights into whether students think about their interactions or are resigned to data practices. However, conjectures drawn from these results, particularly about their actions, need to be examined in future research.

Second, the questionnaire's open-ended items about data practices were analysed to answer the first research question. The data provide significant insights about which types of data collection and kinds of data processing purposes students recognise in the given situation. However, many responses were short (e.g., single words or short phrases) so that interpreting the extent of students' understanding of these facets in detail is limited. Addressing this question could be interesting in future work.

Finally, we chose the example of interacting with mobile networks in the intervention because of its well explainable structure and functionality. This allowed us to focus on introducing and using the framework's perspectives without introducing additional, more complex concepts of data-driven technologies such as ML aspects. Applying the framework to more complex ddA could provide fruitful insights into fostering data awareness throughout secondary education.

8. Conclusions and Future Work

We have introduced the data awareness framework, which combines perspectives on teaching about data and data-driven technologies according to data and AI literacies. It provides a model for the role of data within interactions with data-driven digital artefacts (ddA) (see Figure 2). This model is meant to be used by students to shift their focus on relevant aspects during their daily interactions with ddA. The framework's goal is to enable students to become aware of their role and the role of data within interactions with ddA, reflect on these interactions, and make informed and self-determined actions. For initial insights about data awareness, we present an empirical study which focuses on students' understanding of and reflections on data practices of ddA.

The study's findings indicate that the framework can support students in developing a conceptual understanding of and a nuanced perspective on data practices of ddA. This leads to the preliminary suggestion that the model taught is comprehensible for students and usable to analyse ddA regarding their data practices. The results also indicate that students were encouraged to reflect on ddAs' data collection and processing and want to reduce the amount of data they disclose when interacting with ddA. Thus, the framework might support students to reflect on their role during interactions with ddA. Taken together, these findings suggest that the framework could support students to be more informed and reflective when interacting with ddA, so that students could be more self-determined when interacting with ddA in everyday life.

In future work, research is needed to explore what specific impact data awareness has on long time behavioural changes. The framework is intended to support students in reflecting on their role within these interactions (including their actions), identifying and assessing possible actions, and making informed and reflected decisions about whether or how to interact with ddA. To avoid prescribing behaviour – which would limit students' agency – our intervention consciously ends with discussing possible actions. However, the study indicates that students think about their behaviour (i.e. about disclosing less personal data). Exploring these effects on students' behaviour would be interesting to understand the relationship between awareness and agency.

Moreover, future work on data awareness could benefit from adopting critical perspectives, as used in critical data literacies (e.g., Sander 2020, Pangrazio & Selwyn 2019). Future implementations of the framework could extend the perspective on individual interactions on a societal level by considering impacts of large-scale data practices. Then, societal and ethical issues could be critically discussed and reflected, e.g., influences on electoral processes (e.g., Bond et al. 2012, Lazer et al. 2018) or asymmetries of information and power (e.g., West 2019, Brunton & Nissenbaum 2015). This can also include discussing design decisions of ddA, like regarding predictive privacy (e.g., Mühlhoff 2021).

Overall, this paper contributes to research on data literacies by focusing on supporting students in understanding and becoming aware of data practices of ddA, as already emphasised by other researchers (e.g., Pangrazio & Selwyn 2019, Acker & Bowler 2018, Gebre 2022). While approaches for teaching data and ddA typically focus on enabling students to handle and interpret data or design ML applications, the presented framework contributes to this research with an alternative perspective. It proposes an approach of teaching a model of the role of data in interactions with ddA (see Figure 2) that appears to be understandable and usable as a lens for analysing ddA and reflecting on daily interactions with them. Based on the findings, we assume that the approach can combine learning about data and ddA with encouraging students to reflect on their own role in a datafied world. Therefore, we believe that the data awareness framework contributes to research on data and AI literacies with an approach that may enable students to understand ddA and reflect on their own role within the interactions with ddA, that is, to become aware of their role and the role of data in these interactions. With regard to Pangrazio & Selwyn (2018), our study suggests that the framework may support students in addressing the challenges to their privacy and agency posed by data practices of ddA.

References

Acker, A. & Bowler, L. (2018), Youth Data Literacy: Teen Perspectives on Data Created with Social Media and Mobile Devices, *in* 'Proceedings of the 51st Hawaii International Conference on System Sciences', Hawaii, p. 10.

Acquisti, A. & Grossklags, J. (2005), 'Privacy and rationality in individual decision making', *IEEE Security and Privacy Magazine* **3**(1), 26–33.

Barnes, S. B. (2006), 'A privacy paradox: Social networking in the United States', *First Monday* **11**(9).

Bilstrup, K.-E. K., Kaspersen, M. H., Lunding, M. S., Schaper, M.-M., Van Mechelen, M., Tamashiro, M. A., Smith, R. C., Iversen, O. S. & Petersen, M. G. (2022), Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data, *in* 'Proceedings of the 21st Annual ACM Interaction Design and

Children Conference', IDC '22, Association for Computing Machinery, New York, NY, USA, pp. 225–236.

Bode, M. & Kristensen, D. B. (2016), The digital doppelgänger within: A study on self-tracking and the quantified self movement, *in* R. Canniford & D. Bajde, eds, 'Assembling Consumption: Researching Actors, Networks and Markets', Routledge, Oxon, United Kingdom, pp. 119–135.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E. & Fowler, J. H. (2012), 'A 61-million-person experiment in social influence and political mobilization', *Nature* **489**(7415), 295–298.

Bowler, L., Acker, A., Jeng, W. & Chi, Y. (2017), "'It lives all around us": Aspects of data literacy in teen's lives', *Proceedings of the Association for Information Science and Technology* **54**(1), 27–35.

Brodsky, J. E., Zomberg, D., Powers, K. L. & Brooks, P. J. (2020), 'Assessing and fostering college students' algorithm awareness across online contexts', *Journal of Media Literacy Education* **12**(3), 43–57.

Brunton, F. & Nissenbaum, H. F. (2015), *Obfuscation: A User's Guide for Privacy and Protest*, MIT Press, Cambridge, Massachusetts.

Bucher, T. (2017), 'The algorithmic imaginary: Exploring the ordinary affects of facebook algorithms', *Information, Communication & Society* **20**(1), 30–44.

Burrell, J. (2016), 'How the machine 'thinks': Understanding opacity in machine learning algorithms', *Big Data & Society* **3**(1).

Calzada Prado, J. & Marzal, M. Á. (2013), 'Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents', *Libri* **63**(2).

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B. & Barro, S. (2023), 'AI literacy in K-12: A systematic literature review', *International Journal of STEM Education* **10**(1), 29.

Cheney-Lippold, J. (2017), *We Are Data: Algorithms and The Making of Our Digital Selves*, NYU Press, New York.

Cohen, J. (1992), 'A power primer', *Psychological Bulletin* **112**(1), 155–159.

Denning, P. J. & Denning, D. E. (2020), 'Dilemmas of artificial intelligence', *Communications of the ACM* **63**(3), 22–24.

D'Ignazio, C. (2017), 'Creative data literacy: Bridging the gap between the data-haves and data-have nots', *Information Design Journal* **23**(1), 6–18.

D'Ignazio, C. & Bhargava, R. (2015), Approaches to Building Big Data Literacy, *in* 'Bloomberg Data for Good Exchange Conference', New York.

Dinev, T. & Hart, P. (2006), 'An Extended Privacy Calculus Model for E-Commerce Transactions', *Information Systems Research* **17**(1), 61–80.

Dogruel, L. (2021), What is Algorithm Literacy?: A Conceptualization and Challenges Regarding its Empirical Measurement, *in* M. Taddicken & C. Schumann, eds, 'Algorithms and Communication', Vol. 9 of *Digital Communication Research*, Freie Universität Berlin, Berlin, pp. 67–93.

Dowthwaite, L., Creswick, H., Portillo, V., Zhao, J., Patel, M., Vallejos, E. P., Koene, A. & Jirotka, M. (2020), "It's your private information. it's your life.": Young people's views of personal data use by online technologies, *in* 'Proceedings of the Interaction Design and Children Conference', ACM, London United Kingdom, pp. 121–134.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K. & Sandvig, C. (2015), "I always assumed that I wasn't really that close to [her]": Reasoning about Invisible Algorithms in News Feeds, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', ACM, Seoul Republic of Korea, pp. 153–162.

Gabriele, S. & Chiasson, S. (2020), Understanding Fitness Tracker Users' Security and Privacy Knowledge, Attitudes and Behaviours, *in* 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', ACM, Honolulu HI USA, pp. 1–12.

Gebre, E. (2018), 'Young Adults' Understanding and Use of Data: Insights for Fostering Secondary School Students' Data Literacy', *Canadian Journal of Science, Mathematics and Technology Education* **18**(4), 330–341.

Gebre, E. (2022), 'Conceptions and perspectives of data literacy in secondary education', *British Journal of Educational Technology* **53**(5), 1080–1095.

Goray, C. & Schoenebeck, S. (2022), 'Youths' perceptions of data collection in online advertising and social media', *Proceedings of the ACM on Human-Computer Interaction* **6**(CSCW2), 1–27.

Hargittai, E. & Marwick, A. (2016), "'What Can I Really Do?" Explaining the Privacy Paradox with Online Apathy', *International Journal of Communication* **10**, 3737–3757.

Keen, C. (2020), 'Apathy, convenience or irrelevance? Identifying conceptual barriers to safeguarding children's data privacy', *New Media & Society* **24**(1).

Kitchin, R. (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*, SAGE Publications, Los Angeles, California.

Kramer, A. D. I., Guillory, J. E. & Hancock, J. T. (2014), 'Experimental evidence of massive-scale emotional contagion through social networks', *Proceedings of the National Academy of Sciences* **111**(24), 8788–8790.

Kuckartz, U. (2014), *Qualitative Text Analysis: A Guide to Methods, Practice & Using Software*, SAGE, Los Angeles.

Landis, J. R. & Koch, G. G. (1977), 'The measurement of observer agreement for categorical data', *Biometrics* **33**(1), 159–174.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J. & Zittrain, J. L. (2018), 'The science of fake news', *Science* **359**(6380), 1094–1096.

Livingstone, S., Stoilova, M. & Nandagiri, R. (2019), *Children's Data and Privacy Online: Growing up in a Digital Age. An Evidence Review*, London School of Economics and Political Science, London.

Long, D. & Magerko, B. (2020), What is AI Literacy? Competencies and Design Considerations, *in* 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', ACM, Honolulu HI USA, pp. 1–16.

Lutz, C., Hoffmann, C. P. & Ranzini, G. (2020), 'Data capitalism and the user: An exploration of privacy cynicism in Germany', *New Media & Society* **22**(7).

Marques, L. S., Gresse Von Wangenheim, C. & Hauck, J. C. R. (2020), 'Teaching Machine Learning in School: A Systematic Mapping of the State of the Art', *Informatics in Education* pp. 283–321.

Mayer-Schönberger, V. & Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston.

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

Mühlhoff, R. (2021), 'Predictive privacy: Towards an applied ethics of data analytics', *Ethics and Information Technology* **23**(4), 675–690.

OECD (2014), 'Summary of the oecd privacy expert roundtable: Protecting privacy in a data-driven economy: Taking stock of current thinking'. (accessed 14 November 2023).
URL: *https://one.oecd.org/document/DSTI/ICCP/REG(2014)3/en/pdf*

Pangrazio, L. & Sefton-Green, J. (2020), 'The social utility of 'data literacy'', *Learning, Media and Technology* **45**(2), 208–220.

Pangrazio, L. & Selwyn, N. (2018), ''It's Not Like It's Life or Death or Whatever'': Young People's Understandings of Social Media Data', *Social Media + Society* **4**(3), 1–9.

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

Pangrazio, L. & Selwyn, N. (2020), 'Towards a school-based 'critical data education'', *Pedagogy, Culture & Society* **29**(3), 431–448.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., Roberts, M. E., Shariff, A., Tenenbaum, J. B. & Wellman, M. (2019), 'Machine behaviour', *Nature* **568**(7753), 477–486.

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B. & Ali-Hassan, H. (2015), 'Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report'. (accessed 14 November 2023).
URL: *http://dx.doi.org/10.13140/RG.2.1.1922.5044*

Sander, I. (2020), 'What is critical big data literacy and how can it be implemented?', *Internet Policy Review* **9**(2).

Schulte, C. & Budde, L. (2018), A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education, *in* 'Proceedings of the 18th Koli Calling International Conference on Computing Education Research', ACM, Koli, Finland, pp. 1–10.

Shane, S. & Wakabayashi, D. (2018), ''The Business of War': Google Employees Protest Work for the Pentagon', *The New York Times* . (accessed 14 November 2023).
URL: *https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html*

Sun, K., Sugatan, C., Afnan, T., Simon, H., Gelman, S. A., Radesky, J. & Schaub, F. (2021), "They See You're a Girl if You Pick a Pink Robot with a Skirt": A Qualitative Study of How Children Conceptualize Data Processing and Digital Privacy Risks, *in* 'Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems', CHI '21, Association for Computing Machinery, New York, NY, USA, pp. 1–34.

Susser, D., Roessler, B. & Nissenbaum, H. (2019), 'Technology, autonomy, and manipulation', *Internet Policy Review* **8**(2), 1–22.

Tedre, M., Vartiainen, H., Kahila, J., Toivonen, T., Jormanainen, I. & Valtonen, T. (2020), Machine learning introduces new perspectives to data agency in k-12 computing education, *in* '2020 IEEE Frontiers in Education Conference (FIE)', IEEE, Uppsala, Sweden, pp. 1–8.

Touretzky, D., Gardner-McCune, C., Martin, F. & Seehorn, D. (2019), 'Envisioning AI for K-12: What Should Every Child Know about AI?', *Proceedings of the AAAI Conference on Artificial Intelligence* **33**, 9795–9799.

Tufekci, Z. (2014), 'Engineering the public: Big data, surveillance and computational politics', *First Monday* **19**(7).

Vartiainen, H., Tedre, M., Kahila, J. & Valtonen, T. (2020), 'Tensions and trade-offs of participatory learning in the age of machine learning', *Educational Media International* **57**(4), 285–298.

Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M. & Valtonen, T. (2021), 'Machine learning for middle schoolers: Learning through data-driven design', *International Journal of Child-Computer Interaction* **29**, 100281.

West, S. M. (2019), 'Data Capitalism: Redefining the Logics of Surveillance and Privacy', *Business & Society* **58**(1), 20–41.

Williams, M., Nurse, J. R. C. & Creese, S. (2017), Privacy is the Boring Bit: User Perceptions and Behaviour in the Internet-of-Things, *in* '2017 15th Annual Conference on Privacy, Security and Trust (PST)', IEEE, Calgary, AB.

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U. & Kortuem, G. (2016), 'Creating an Understanding of Data Literacy for a Data-driven Society', *The Journal of Community Informatics* **12**(3).

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, first edition edn, PublicAffairs, New York.