Informatics in Education, 2024, Vol. 23, No. 3, 593–624

DOI: 10.15388/infedu.2024.19

Empowering Students for the Data-Driven World: A Qualitative Study of the Relevance of Learning about Data-Driven Technologies

Lukas HÖPER, Carsten SCHULTE
Computing Education, Department of Computer Science, Paderborn University, Germany
e-mail: lukas.hoeper@uni-paderborn.de, carsten.schulte@uni-paderborn.de

Received: December 2023

Abstract. In K-12 computing education, there is a need to identify and teach concepts that are relevant to understanding machine learning technologies. Studies of teaching approaches often evaluate whether students have learned the concepts. However, scant research has examined whether such concepts support understanding digital artefacts from everyday life and developing agency in a digital world. This paper presents a qualitative study that explores students' perspectives on the relevance of learning concepts of data-driven technologies for navigating the digital world. The underlying approach of the study is data awareness, which aims to support students in understanding and reflecting on such technologies to develop agency in a data-driven world. This approach teaches students an explanatory model encompassing several concepts of the role of data in data-driven technologies. We developed an intervention and conducted retrospective interviews with students. Findings from the analysis of the interviews indicate that students can analyse and understand data-driven technologies from their everyday lives according to the central role of data. In addition, students' answers revealed four areas of how learning about data-driven technologies becomes relevant to them. The paper concludes with a preliminary model suggesting how computing education can make concepts of data-driven technologies meaningful for students to understand and navigate the digital world.

Keywords: K-12, computing education, data awareness, AI literacy, machine learning, digitalisation, datafication.

1. Introduction

The increasing digitalisation highlights the need for computing education for all, leading to a growing emphasis on computing education in schools in various countries. The range of topics covered in computer science education (CSE) has also expanded, with a prominent additional focus on artificial intelligence (AI) and machine learning (ML) (e.g., Tedre *et al.*,2021; Shapiro *et al.*, 2018; Rizvi *et al.*, 2023; Sentence and Waite, 2022). The argument for this need for computing education for all seems obvious: The digital world functions and exists through the field of computer science, as it enables the development of digital systems and technologies. Therefore, the typical argument is that knowledge and skills in computer science are essential to understand and act independently in the digital world, to gain professional qualifications and to participate in society (e.g., Caspersen *et al.*, 2022, p. 2).

But what does this mean conceptually? We have illustrated the typical argument structure in Fig. 1 and give a brief explanation here: Due to advances in digitalisation and datafication (Mayer-Schönberger and Cukier, 2013), new technologies and conceptual ideas are emerging that are also transforming concepts that need to be taught in computing education, such as those related to ML (e.g., Tedre *et al.*, 2021; Sentence and Waite, 2022). At the same time, these concepts are already finding their way into actual products readily available in people's everyday lives. Hence, there is an ongoing educational need to identify specific concepts that are relevant to understanding digital technologies in the digital world. Subsequently, the question is how to teach these concepts effectively. Empirical (evaluative) research in CSE often focuses on examining the effectiveness of these approaches, that is, whether students actually learn the concepts taught. Currently, many researchers are developing approaches for teaching ML concepts and examining whether students have learned these concepts (e.g., Vartiainen *et al.*, 2021; Höper and Schulte, 2023; Martins *et al.*, 2023; Hitron *et al.*, 2019).

However, scant research examines whether learning data and ML concepts helps students understand the data-driven technologies they encounter in everyday life, that is, making use of the concepts learned in CS classes. It is an ongoing question of how to support students in engaging with these technologies in their everyday lives and develop agency in interacting with digital artefacts to navigate the digital world better. If this step is neglected, it raises the question of whether there is a gap in the above argument: When students have learned concepts, can they really relate them meaningfully

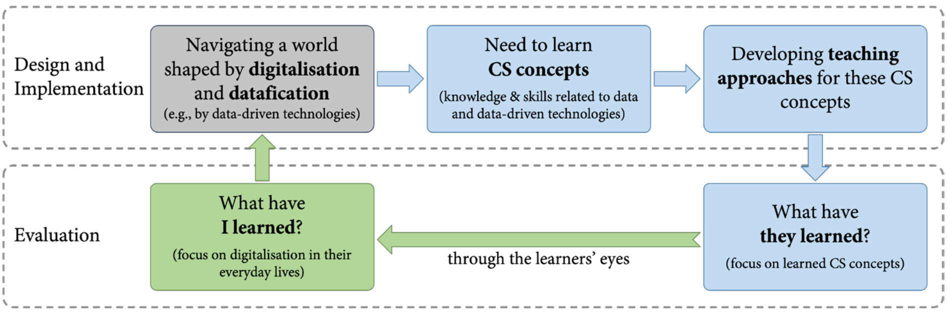

Fig. 1. Overview of typical research steps and questions in Computer Science Education Research (CSER). *Blue flow commonly addressed:* Comprehending digitalisation requires specific CS-related knowledge and skills; respective teaching approaches are developed and evaluated regarding whether students effectively learn the concepts. *Green flow as underrepresented steps in CSER:* Examine whether students relate and apply the acquired knowledge and skills (concepts) to their everyday lives, which should support them in navigating the digital and datafied world, by focusing on the students' views on these concepts ("through the learners' eyes").

to their everyday lives and experiences in the digital world? In particular, this is important as this would be a prerequisite for developing agency in the digital world based on the concepts. This question is underlined by recent studies that report that students struggle to relate and apply learned concepts about data and data-driven technologies to their everyday lives and to reflect on everyday situations with data-driven technologies (e.g., Bowler *et al.*, 2017; Gebre, 2018; Livingstone *et al.*, 2019; Vartiainen *et al.*, 2021). For example, when evaluating workshops with children, Vartiainen *et al.* (2021) found that students learned about ML concepts and mechanisms but did not critically reflect on them in their everyday lives. The urgency is further underlined by findings that even teachers struggle to see computing education as relevant, meaningful and important (Mertala, 2021).

This paper presents an exploratory, qualitative study that addresses these questions, particularly "through the learners' eyes". The *data awareness framework* forms the basis of this study. It is intended to explicitly link learning about the role of data in data-driven digital artefacts with students' everyday experiences (Höper and Schulte, 2023). This study aims to examine whether students have learned concepts of data-driven technologies as characterised in this framework and explore, through the learners' eyes, the relevance of these concepts for navigating the digital world. With a particular focus on the second step (see green flow in Fig. 1), we analysed semi-structured interviews with school students after participating in a data awareness intervention by conducting a qualitative content analysis.

The paper is therefore structured as follows: After providing an overview of the background to this study in Section 2, we briefly introduce the data awareness framework, its teaching approach and the concepts it encompasses (see Section 3). We then present in Section 4 the method and design of the study, including an intervention that implements the framework. Next, we present the results, which include sections on the concepts learned (particularly in terms of their application to other contexts) and, perhaps most importantly, students' perspectives on the perceived relevance of what they have learned (see Section 5). We then discuss the findings and synthesise them into a preliminary model of how to support students in relating the concepts of data-driven technologies they have learned to their everyday lives (see Section 6) and discuss the study's limitations (see Section 7). The model can suggest how such concepts can become relevant and useful to students' development of agency and empowerment in a digital and data-driven world. Finally, in Section 8, we conclude the paper with reflections on the need for such evaluations and whether demonstrating that students have attained the intended disciplinary and conceptual learning objectives is sufficient or whether additional steps are required.

2. Background

In this section, we briefly discuss the context and research on the aforementioned issues, which highlight the need to make learning about concepts of data and data-driven technologies relevant to students' everyday lives, as this is the focus of this paper.

2.1. *Students Lack Understanding of Data-Driven Technologies and Feel Powerless*

In recent years, more and more technologies and applications in our daily lives are using data-driven techniques, especially ML methods. Examples include search engines, social media applications and streaming services. These artefacts use data models generated by collecting and processing massive amounts of data. Thus, data has an essential role in these artefacts (see, Sculley *et al.*, 2015). We call them therefore data-driven digital artefacts (hereafter abbreviated as ddA). Their data practices allow for large-scale tracking and profiling of peoples' lives (Tedre *et al.*, 2020; Zuboff, 2019; Pangrazio and Selwyn, 2019), as described with the term of datafication by Mayer-Schönberger and Cukier (2013). Through such data practices, they have the potential to influence individuals and societies (Tufekci, 2014; Rahwan *et al.*, 2019).

However, the data collection and processing by ddA and their influences are often not transparent and apparent to students. Recent research indicates that students lack awareness and understanding of the data collection and processing by ddA, and are not aware of where, how and why data about them is collected and processed (e.g., Bowler *et al.*, 2017; Bucher, 2017; Goray and Schoenebeck, 2022; Pangrazio and Selwyn, 2019, 2020; Tedre *et al.*, 2020) (for an overview, see, Höper and Schulte, 2023). Moreover, even when students are taught about data and data-driven technologies, studies report that they struggle to relate their understanding to their everyday lives and reflect on it accordingly (e.g., Bowler *et al.*, 2017; Gebre, 2018; Livingstone *et al.*, 2019; Vartiainen *et al.*, 2021). Hence, we argue for the need to support and encourage students to relate their learning to their everyday lives, as shown in the green flow in Fig. 1 and similarly argued by others (e.g., Gebre, 2018; Pangrazio and Selwyn, 2019; Bilstrup *et al.*, 2022).

In addition, some studies report that people tend to develop feelings of apathy, resignation, powerlessness or lack of control over the data practices of institutions and providers of ddA (e.g., Hargittai and Marwick, 2016; Sander, 2020; Keen, 2020; Lutz *et al.*, 2020). Even if they are concerned about the collection and processing of data about them, they feel powerless to do anything about it (Dowthwaite *et al.*, 2020; Bilstrup *et al.*, 2022) and fall into a state of surrender when using ddA (Sander, 2020). Possible reasons for this are discussed in the literature. For example, it could be due to a lack of knowledge and skills to control personal data, an inability to take appropriate actions or a belief that privacy-protective behaviours are useless (Lutz *et al.*, 2020; Hargittai and Marwick, 2016).

These findings indicate a gap between learning about data and ddA and the question "What have I learned?" concerning the digitalisation and datafication that students encounter in their everyday lives (see Fig. 1). This highlights the need to support students in learning concepts about data and ddA and (!) to make them appropriately usable and meaningful for navigating the digital world and overcoming powerlessness in everyday life. Similarly, other researchers argue for supporting students in perceiving their role in a data-driven society and relating learning about data and ddA to their everyday lives to engage with the ddA they interact with in everyday life, reflect on their interactions, and make informed and self-determined decisions about the role of technology in their lives (Gebre, 2018; Pangrazio and Selwyn, 2019; Dindler *et al.*, 2020; Bilstrup *et al.*, 2022).

2.2. *Approaches to Teach Concepts of Data-Driven Technologies*

Based on the findings discussed above, approaches are needed that (effectively) support students in developing an understanding of specific concepts about ddA while simultaneously empowering them to relate these concepts to their everyday lives and make them usable and meaningful for navigating the digital world.

A wide range of approaches have been developed and discussed in recent years, such as data literacy, AI literacy, and data agency. These approaches include different aspects of teaching about data and ddA. While data literacy is about understanding and handling data (i.e. reading, interpreting, collecting, analysing, etc.) (e.g., Ridsdale *et al.*, 2015; Wolff *et al.*, 2016; Gebre, 2022), AI literacy aims at supporting students in understanding, recognising and developing AI applications (e.g., Long and Magerko, 2020; Casal-Otero *et al.*, 2023), and data agency focuses on enabling students to become makers and producers in the digital world in terms of designing ML applications (Tedre *et al.*, 2020).

From an empowerment perspective and as an extension of computational thinking, Dindler *et al.* (2020) describe the idea of computational empowerment. This encompasses constructing and deconstructing digital artefacts, which refers to students' different roles when engaging with technology: (1) students design and construct digital artefacts that others can use, and (2) students engage with digital artefacts others have designed. This leads to the need for students to learn to shape and design digital technologies (i.e., learn to program) and analyse and reflect on digital artefacts (e.g., according to their impact on individuals and society). Dindler *et al.* (2020) argue that fostering students' empowerment in a digital world requires a balance between constructing and deconstructing activities.

From this perspective, school initiatives on data literacy, AI literacy and data agency in computing education often focus on enabling students to handle data and design digital artefacts such as ML-based applications. Approaches that support students in analysing, understanding and reflecting on data-driven technologies and their role in students' everyday lives are rare. As outlined in the following section, the data awareness framework addresses this gap and starts by supporting students in learning concepts about ddA that are useful for analysing, understanding and reflecting on ddA from their everyday lives. In doing so, it aims to support them in engaging with such artefacts in a more informed and empowered way to develop self-determination in navigating the data-driven world.

3. Data Awareness Framework as the Approach of this Study

In this study, we draw on the data awareness framework, which is presented in this section.

3.1. *Basic Concepts of Data-Driven Digital Artefacts in the Light of Data Awareness*

The data awareness framework includes an educationally designed model for interaction systems with ddA; we call it an explanatory model. This model encompasses concepts of the role of data in ddA, mainly focusing on the collection and processing of data. We assume these concepts are relevant in supporting school students in relating their learning about ddA to their everyday lives, reflecting on their interactions with ddA, overcoming feelings of powerlessness, and developing agency in a digital world. Below, we describe these concepts in more detail, as summarised in Fig. 2.

Types of data collection. During interactions with ddA, data collection is conducted in different ways. In literature, various types of data collection are described (e.g., Pangrazio and Selwyn, 2019; Livingstone *et al.*, 2019; OECD, 2014). In summary, data can be distinguished as provided data that the user actively creates; observed data that is gathered through observation and recording, of which the user is not necessarily aware; derived data that is generated by directly processing existing data; and inferred data that is generated by probability-based processing (OECD, 2014). Thus, in interactions with ddA, personal data is collected when users intend to provide data, but also through observation and tracking or generation during data processing. Based on this, in the model, we distinguish between *explicit data collection*, which refers to data that users intentionally provide through their actions when using a ddA (provided data), and *implicit data collection*, which refers to data that is collected through observation, tracking and data processing alongside the user's action, so that users are rather not aware of it (observed, derived and inferred data).

Kinds of data processing purposes. When ddA collect data explicitly and implicitly during interactions, the data processing is motivated by several purposes. Some data is used to provide features and generate outputs during the interaction, which is often apparent from the user's perspective. In the framework, these purposes are conceptualised as *primary purposes* and are usually described from the user's perspective. However,

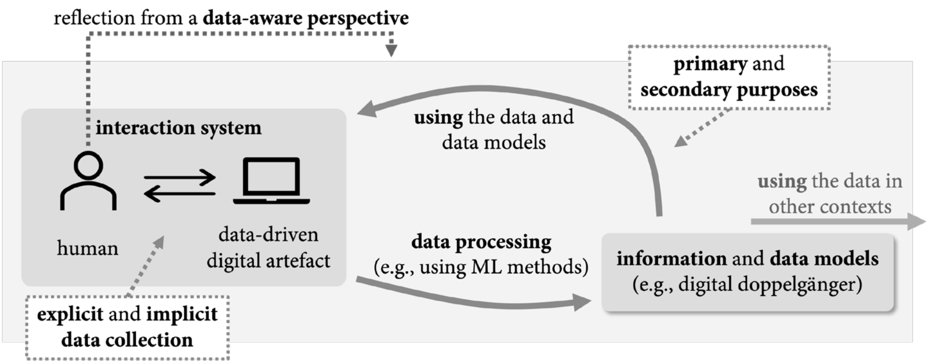

Fig. 2. Explanatory model of interactions between humans and ddA with a focus on the role of data within these interaction systems as an overview of the concepts of the data awareness framework (see, Höper and Schulte, 2023).

from the provider's perspective, data processing often serves additional purposes beyond the immediate generation of output. For example, the collected data is used and processed to predict users' future behaviour in order to adapt and develop the features of the ddA accordingly (e.g., Mühlhoff, 2021; Tufekci, 2014; West, 2019). These predictions can be used, for example, to make recommendations or, at another level, to influence users' behaviour and emotions (see for examples, Kramer *et al.*, 2014; Bond *et al.*, 2012; Zuboff, 2019). In the data awareness framework, we have described them as *secondary purposes*. They are typically not readily apparent to users when using the ddA. It is often challenging to identify secondary purposes in detail as they are sometimes deliberately obscured by the ddA provider or overshadowed by primary purposes (e.g., Brunton and Nissenbaum, 2015; Burrell, 2016; Zuboff, 2019).

Construction of data models about users. In addition, using different data-driven methods, ddA create different data models, such as ML models for generating outputs or models about users that are generated through user modelling techniques. In this paper, we focus on *data models* about users. For these models, several conceptions are described in the literature. For example, Bode and Kristensen (2016) characterise them as digital doppelgänger, or Kitchin (2014, pp. 166-168) describe them as data footprints and shadows. Such a model can be continuously refined during the interaction (Bode and Kristensen, 2016). However, it is always limited to proxies and can never be understood as a 'copy' of the user. It consists, for example, of explicitly and implicitly collected data about the user, such as past and present actions. It may also include predictions about preferences or future behaviour (Kitchin, 2014; Zuboff, 2019). By processing collected data, such a data model could also include sensible data, even if not provided by the user (Goray and Schoenebeck, 2022; Mühlhoff, 2021). For example, predictive analytics methods can be used to predict sensitive personal information by processing a lot of data about other users or using specific predictive models (Mühlhoff, 2021).

3.2. *Developing Data Awareness in Everyday Interactions with ddA*

Based on the aim to enable students to understand and reflect on the ddA they interact with, we have described the previous explanatory model. This model is intended to be an analytical lens on ddA, with a particular focus on the role of data due to its essential role. Thus, the embedded concepts are used for the following definition:

Data awareness is defined as being aware of the explicit and implicit data collection, the primary and secondary purposes, and the role of data models about oneself, as well as one's role during interactions with ddA.

In the framework, we use the term 'awareness' to emphasise the idea that it is about using the proposed model as a lens to shift the focus on the role of data in ddA, that is, to direct attention to the role of data rather than to the immediate goals of why one is using the ddA. Furthermore, the connotation of awareness is intended to highlight that it does not primarily focus on competencies (e.g., for analysing data or developing ML applications) but involves critical reflection on one's interactions with ddA and understanding of one's role in a digital world. This is similar to approaches for including critical perspectives and dimensions in computing education, such as in the pedagogy of multiliteracies as described by Mertala (2021).

The overarching rationale of data awareness is to support students in developing empowerment or agency in a digital world. Notably, the framework is not intended to teach particular attitudes towards digital technologies or habits of use but rather to support them in making informed and reflected decisions by themselves. Therefore, the framework follows the meaning of education (or Bildung) as supporting the transformation of students' perspectives on the world, on themselves and on their behaviour, as discussed, for example, by Schulte and Budde (2018) in their theoretical discussion of the meaning of Bildung. Accordingly, the data awareness framework aims to provide students with the previously described explanatory model (see Fig. 2) in terms of a lens on ddA to support them in understanding such interaction systems, reflecting on the role of ddA, and reflecting on their role and their behaviour. The framework's model is designed from a user perspective on ddA, which is intended to support bridging the gap between learning concepts in class and experiences with ddA in everyday life, as shown in the green flow in Fig. 1. In doing so, this approach aims to enable students to apply this model as an analytical lens during their everyday interactions with ddA to uncover the role of data and then make informed and self-determined decisions, supporting the development of agency and empowerment in navigating the data-driven world.

We use a context-based approach to support this relationship between learning concepts of ddA according to the explanatory model and students' everyday lives. This approach has a long tradition in science education (see for example, Gilbert, 2006; Bennett *et al.*, 2007), but it has also been applied in computing education (e.g., Nijenhuis-Voogt *et al.*, 2021). The idea is to use contexts or situations from students' everyday lives in which the learning of concepts is embedded. It aims to support students in relating learning experiences to everyday life. Accordingly, the approach for the data awareness framework is to embed learning the framework's concepts in consideration of everyday situations of interaction with a ddA. In this way, students subsequently learn the different aspects of the explanatory model through a cyclical process of (a) considering an exemplary ddA, (b) decontextualising and learning a concept from the explanatory model, and (c) recontextualising the concept to apply it to the exemplary context. Section 4.3 describes an exemplary implementation of this approach as part of this study.

However, as discussed earlier, there are several related approaches to teaching about data and ddA that are part of computing education in schools, such as data literacy, AI literacy and data agency (e.g., Gebre, 2022; Long and Magerko, 2020; Tedre *et al.*, 2020). How does data awareness relate to these approaches? The relationship could be explained using the different perspectives of empowered engagement with digital technologies as described earlier (e.g., Dindler *et al.*, 2020), that is, designing digital technologies and analysing and reflecting on digital technologies. Most approaches to teaching and learning about data and ddA (i.e. regarding the literacies mentioned) are concerned with teaching knowledge and skills to enable students to carry out data projects or develop ML applications (e.g., Druga and Ko, 2021; Hitron *et al.*, 2019; Vartiainen *et al.*, 2021). Somewhat contrasting, data awareness addresses students from their everyday perspective on ddA and focuses on analytical and reflective considerations of ddA to support in developing a conceptual understanding of these technologies (e.g., the internal workings regarding data practices). Thus, data awareness is part of the intersection of the different literacies related to data and ddA but also suggests new perspectives for teaching and learning about data and ddA.

4. Method

In the following sections, we present the empirical study of the learned concepts of ddA and, perhaps most importantly, students' perspectives on the relevance of these concepts to their everyday lives.

4.1. *Research Questions*

This study focuses on the question 'What have I learned?' as perceived by the learners when they are introduced to concepts about data-driven technologies (see green flow in Fig. 1). As it is unclear how students relate learning about data and ddA to their everyday lives or what they perceive to be relevant to them, we needed an open and explorative approach. Therefore, we chose a qualitative study to explore students' thoughts on these questions (i.e. examine the relevance "through their eyes"). Thus, the study aims to identify how learning the previously described explanatory model could become relevant for navigating the data-driven world. This could facilitate the development of theoretical conjectures about making learning about data-driven technologies relevant to students' everyday lives and inform future developments of respective computing education approaches. However, to make appropriate interpretations, insights into students' understanding of the concepts taught (i.e. concepts of the framework's model) are also needed. This study, therefore, addresses two research questions:

- **RQ1.** To what extent do students have an understanding of the role of data within interactions with ddA, especially according to the concepts embedded in the data awareness framework? ("What have they learned?" in Fig. 1)
- **RQ2.** What relevance do students perceive for everyday life when learning about the concepts of data and data-driven technologies as characterised in the data awareness framework? ("What have I learned?" in Fig. 1)

4.2. *Procedure and Participants*

We developed a teaching unit for this study that implements the data awareness framework (details are described in Section 4.3). This intervention is for secondary school computing education in grades 8 to 10. It deals with recommendation systems as exemplary technology embedded in many ddA that students use in everyday life. We recruited two teachers from collaborations in prior projects who conducted the teaching unit in two 8th-grade computing classes as their regular teachers. One of the authors instructed the teachers about the intervention, all the materials and the study design. A total of 58 students (13–15 years old) participated in this intervention. It consisted of eight lessons (45 minutes) over four weeks. During this time, we held weekly meetings with the teachers to present and discuss the steps of the teaching unit and its materials. The teachers reported on their experiences and observations of the lessons. We used these reports to check how the lessons fit the data awareness framework and to get ideas for revising the teaching unit.

After the teaching unit (i.e. after the last lesson), we conducted individual interviews with some of the students who had participated in the intervention. One of the authors conducted the interviews based on written informed consent from the students and their parents. Participation was voluntary. The retrospective interviews were semi-structured (details are described in Section 4.4). Consecutively, six students (five males and one female) were interviewed, three from each class. The selection of students from the classes was random, so we had a convenience sample. According to the teachers, the six students had different achievement levels in computing. The teachers described some of them as usually highly motivated in computing but others as less motivated. The six interviews were audio recorded, which provides the data for the analysis presented in this paper.

4.3. *Teaching Unit about Recommendation Systems*

The teaching unit focuses on the role of data when using streaming services and other everyday contexts with recommendation systems. The intervention implements the core idea of the data awareness framework of teaching students the explanatory model, which they use to reconstruct and reflect on recommendation systems. It consists of four parts, summarised in Fig. 4 and described below.

Given that students have experiences from everyday interactions with ddA, we use such an exemplary situation as an entry point for the intervention in the *first part*. Choosing an everyday situation of an interaction with a ddA allows students to build learning on their prior experiences with ddA. We have chosen the example of a movie streaming service and used it to raise the question of how such a service generates personalised recommendations. Thus, this introduction draws on students' previous experiences and perspectives from their interactions with streaming services. Working in pairs, students experience the process of recommendation systems by switching to the perspective of a streaming service provider: First, they write down three recommendations for each other without telling each other. Second, they ask each other two questions (i.e. collect information about each other). Third, they then refine their previously written recommendations. Fourth, they share their movie recommendations with each other and evaluate the initial and refined recommendations regarding personal fit. Finally, the students reflect on this process and imagine what personal data would be useful to find personalised movie recommendations. This task introduces students to the core idea of a recommendation system, which covers the primary purpose of providing personalised recommendations. The students discuss their ideas in class based on their reflections and ideas for valuable personal data. In this context, the concepts of *explicit and implicit data collection* are introduced. Students then map their ideas for useful data onto explicitly and implicitly collected data. In this part, students learn about the underlying idea of recommendation systems and what personal data might be involved in this process.

The *second part* delves deeper into reconstructing the inner workings of a movie recommendation system. Students are given a prepared Jupyter Notebook. At the beginning of that, students interact with a movie recommendation system we have developed to allow students to look under the hood. For this recommendation system, we have implemented a Python module to generate individual movie recommendations based on rating data from real people. Fig. 3 shows the application, which is given at the beginning of the Jupyter Notebook, where students rate movies and get personalised recommendations. Through step-by-step exploration, students reconstruct the process of generating the recommendations they get in this application. They gain insight into data collection processes (e.g., collecting movie ratings and tracking user behaviour) and data processing to generate personalised movie recommendations. The Jupyter Notebook contains tasks where students explore how to use rating data to identify relevant movies to recommend. They reconstruct the idea of k-nearest-neighbour as an example of an ML technique used for collaborative filtering. In doing so, students learn basics of developing an ML model and using it to generate movie recommendations based on the data of the identified similar users. In this process, the *concept of data models about users* is introduced, which students contextualise to describe the data collection and the role of data models about users in the collaborative filtering process. Such a data model

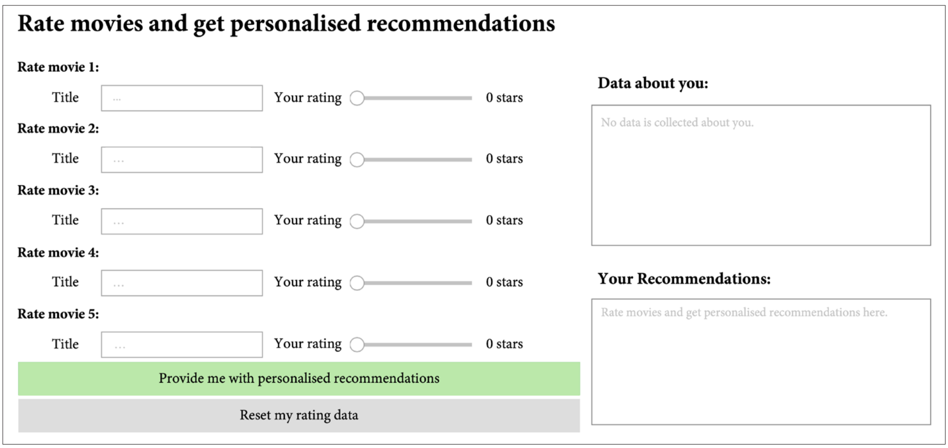

Fig. 3. Students are given a prepared Jupyter Notebook that allows them to reconstruct the internal workings of recommendation systems. This application is shown at the beginning of the Jupyter Notebook, where students receive personalised recommendations based on their ratings. In the following tasks of the Jupyter Notebook, students examine how the recommendations given in this application are generated.

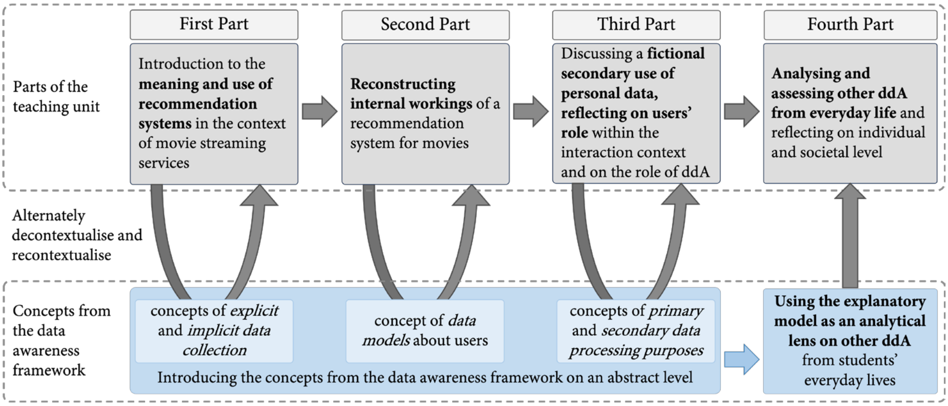

Fig. 4. Overview of the teaching unit's parts and their relation to the approach of the data awareness framework and the concepts as characterised within the explanatory model (see Fig. 2).

contains data about the user and predictions about the user's interests based on the ratings of similar users.

After having learned about the data collection and the primary purpose of processing the data to generate personalised recommendations, the *third part* deals with secondary data processing purposes of streaming services. Thus, this part addresses the question of the secondary use of the collected data (e.g., rating data) and considers the influences of individuals and societies. Therefore, students held a panel discussion on an exemplary secondary purpose of a personalised paywall based on a fictional recommendation system's predictions of users' interests in movies. They discuss this secondary purpose from different perspectives and reflect on the responsible use of such data-driven technologies. In doing so, they also reflect on their role in such interaction systems and explore various aspects of the influence of streaming services (e.g., regarding user behaviour concerning filter bubbles). During this part, the *concepts of primary and secondary purposes* are introduced and distinguished while evaluating the data collection and processing by the ddA.

In the *fourth part*, the students brainstorm about ddA from their everyday lives that use recommendation systems. Working in groups, students select one of these ddA and analyse it according to the concept of the framework that they have learned in the previous parts of the intervention: They analyse what data is collected, how it is processed, for what purposes it is processed and imagine what the data models about the users might look like. In doing so, students apply the conceptual knowledge and skills to other ddA and gain experience in considering ddA in this way (i.e. analysing them regarding the role of data and reflecting on their role in such an interaction). This part could include, for example, examining specific apps that students use on their mobile devices. Students discuss possible advantages and disadvantages after analysing and evaluating the data collection and processing in these different examples. This part should encourage students to apply the concepts they have learned and relate them to their everyday lives. This part aims to support students in developing an understanding of how to make sense of the inner workings of such technologies and make informed decisions about their interactions with ddA. It also intends to encourage critical thinking and reflection.

4.4. *Interviews for Data Collection*

At the end of the teaching unit, we conducted semi-structured interviews to ask students about the concepts they had learned ('What have they learned?' in Fig. 1; RQ1). In addition, and probably more importantly, we wanted to ask students about the relevance of what they had learned from the teaching unit to gain insights into the relevance of the intervention from their perspective ('What have I learned?' in Fig. 1; RQ2). We chose a semi-structured format to be able to re-ask questions during the interviews if the students' answers did not fit the questions and to ask follow-up questions to get more detailed insights into their thoughts. This allowed us to talk about students' experiences and perceptions. Although such a semi-structured format could limit the validity of the interviews, we chose this format because of the explorative approach for RQ2. The interview guideline, therefore, consists of four parts. The interview questions are reported in the Appendix A. Overall, the students were asked:

1. To describe the teaching unit and assess exciting and important aspects.
2. To describe recommendation systems.
3. To apply their knowledge to another context by explaining the role of data in the situation of using a search engine.
4. To say something else about the teaching unit or the interview.

In part three of the interviews, we chose the context of a search engine because it was likely to be familiar to the students and had not been covered in the intervention before. To examine what students have learned about the framework's concepts, we focus on parts two and three while referring to parts one and four to explore the perceived relevance of learning the explanatory model through the learners' eyes. One of the authors conducted the interviews according to the interview guidelines. They lasted between 13 min. and 24 min. 31 sec. with an average length of about 16 min. 14 sec. We transcribed the interviews in full as verbatim transcripts for the analysis, leaving out filler words and pauses.

4.5. *Data Analysis*

We analysed the interview data using a thematic qualitative content analysis (see: Kuckartz, 2014, pp. 69–88) to assess students' understanding about ddA (RQ1) and to explore the relevance of learning about ddA from students point of view (RQ2).

Data analysis regarding RQ1. Regarding the first research question, we aimed to examine whether students have learned the concepts of the explanatory model. Hence, the coding scheme for this analysis is defined deductively based on the framework, resulting in five code categories: (a) explicit data collection, (b) implicit data collection, (c) primary data processing purposes, (d) secondary data processing purposes, and (e) data models about users (see Table 1). We have developed the coding manual for the code categories (a) to (d) in a previous study (see, Höper and Schulte, 2023). One of the authors coded all relevant segments from the interviews (these are described in the previous section). Another researcher received the coded data and the coding manual, assessed the coding and negotiated the results accordingly. The coding was then discussed in data sessions to develop a consensual understanding of the code categories (Kuckartz, 2014). This process contributed to the reliability (or dependability) of the coding (see for a dependability audit, Akkerman *et al.*, 2008). During this analysis, we examined whether the students could apply the concepts to the interview example and explain these concepts in more general terms. This process of assessing the coding in an audit approach and discussing the results in data sessions supports the quality of the coding process. This process contributes to the reliability and validity of the data analysis and the interpretations (Akkerman *et al.*, 2008; Kuckartz, 2014, p. 74).

Data analysis regarding RQ2. We adopted an inductive approach to constructing code categories to explore the relevance students perceived in learning about data and ddA from the intervention (RQ2). This allowed us to explore students' perspectives on how and why learning about ddA was relevant to them. The analysis followed the steps of thematic qualitative content analysis as described by Kuckartz (2014, p. 70). First, we identified the interview segments that were relevant to this research question (mainly two of the four parts of the interviews as described in the previous section). In particular, these were answers in which students described relationships to their everyday lives or assessed the topic of the teaching unit. We then inductively constructed a coding scheme which resulted in four code categories, as shown in Table 2. The interview transcripts were coded accordingly and analysed from a topic-oriented perspective (Kuckartz, 2014, pp. 66–67). The coding and analysis were primarily carried out by one of the authors, while all authors discussed the codes, coding and interpretations in several sessions throughout the process. Analogous to the data analysis for RQ1, we conducted a dependability audit and data sessions among the authors, including assessing the coding and negotiating the results (see, Akkerman *et al.*, 2008; Kuckartz, 2014). According to Akkerman *et al.* (2008), such a process can be helpful in iterative data analyses, such as the explorative approach we used for RQ2. A method of auditing the data collection, data analysis and data interpretations and the corresponding discussions in the data sessions contributes to the reliability and validity of the data analysis (see, Akkerman *et al.*, 2008). During this process, we paid particular attention to the interpretations of the data, which we discussed during the data sessions, partly also with other researchers.

5. Results

This section reports the study's results on students' understanding of the framework's concepts (RQ1) and the perceived relevance of learning about ddA (RQ2).

5.1. *Results on Students' Understanding: What have they Learned?*

We report the results structured according to the concepts mentioned in the data awareness framework in the following. These results are summarised in Table 1 according to the code categories.[1]

Explicit data collection. All students described several examples of explicitly collected data from the teaching unit and their everyday lives. For example, they mentioned "*the search term*" (S1, pos. 30) collected by search engines or "*bank account information*" (S4, pos. 32) collected by online shops. When characterising explicit data collection, they referred to data that is "*obvious*" (S5, pos. 55), "*one provides oneself* " (S2, pos. 6) or "*what you enter yourself*" (S1, pos. 30).

Implicit data collection. Similar to explicit data collection, all students mentioned implicit data collection in the search engine example from the interview and other everyday examples. They associated implicit data collection with personal preferences and clicks on search results (e.g., "*you then click on the Nike ones means perhaps that you are interested in Nike*" (S3, pos. 24)). In addition, they perceived pervasive data collection during their actions (e.g., "*when you think about the fact that almost everywhere, no matter what you are doing, data is collected*" (S1, pos. 22)). While some students struggled

Table 1

Findings according to code categories regarding RQ1

Categories	Findings	Examples
Explicit data collection	Students described several examples; some described it in general terms	"*the search term*" (S1, pos. 30); "*things one provides oneself*" (S2, pos. 6)
Implicit data collection	Students described several examples; some described it in general terms	"*what you like*" (S2, pos.6); "*collected in the background*" (S4, pos. 2; similar by S5)
Primary data processing purposes	Students described several examples; explain mostly only within a context	"*recommending similar search terms*" (S1, pos. 36)
Secondary data processing purposes	Students described several examples; they perceived ddA as restrictive and prescribing for users' actions and behaviour	"*if you get advertisements or something, then only the things that interest you and not the ones that are actually also important*" (S6, pos. 22)
Data models about users	Students recognised the construction of data models about the users; they described them as relating to a user and characterising the user; some recognised relations between and aggregation of different data models	"*if you think about it, how much of your own data is stored somewhere*" (S1, pos. 24)

Note. The categories refer to the taught concepts (see Section 4.3) that are the focus of RQ1. Other concepts are also addressed in the teaching unit, e.g., about developing ML models or concerning recommendation systems. However, these concepts were not the focus of this data analysis for RQ1 and were, therefore, not coded.

[1] We have carefully translated all referenced student answers from German into English.

to articulate the concept of implicit data collection, some were able to grasp it, such as "*collected in the background*" (S4, pos. 2; similarly by S5). It may be more challenging to perceive implicitly collected data, especially to characterise implicit data collection in general terms, compared to explicit data collection.

Primary data processing purposes. All students mentioned primary purposes for using and processing collected data and gave examples related to the intervention and the search engine example from the interview. They named purposes such as "*recommending similar search terms*" (S1, pos. 36) or providing personalised recommendations (e.g., "*everything is recommended that would fit to you*" (S2, pos. 14)). While most students focused on specific contexts, one student took a broader perspective. He described primary purposes concerning the features of ddA and the users' perspective on interactions with ddA: "*Primary purposes are the sheer purpose of what it is sold as, so it is officially sold, for example, Google as a search engine. That is the primary purpose. [...] so at least for us as users*" (S5, pos. 69–71).

Secondary data processing purposes. Most of the students gave examples of secondary data processing purposes in the exemplary situation from the interview and other everyday examples. For example, participants referred to personalised advertising or targeting. One student perceived personalised recommendations as restrictive in one's world view (e.g., "*if you don't get anything from the other topics, so not the important ones or from the world, or if you get advertisements or something, then only the things that interest you and not the ones that are actually also important*" (S6, pos. 22)). The students also mentioned the convenience of not having to search oneself. While some of them assessed these benefits positively, one student expressed concerns about potential risks (e.g., "*on the one hand it's quite cool, but on the other hand you also have something like shopping addiction risks*" (S2, pos. 16)). Overall, the students demonstrated an understanding of both primary and secondary purposes. However, their responses were notably more detailed regarding general perceptions of secondary purposes than primary purposes.

Data models about users. Although we have not directly asked about data models about users, some of the answers provide insights into students' perceptions of this concept. They mentioned that ddA have collections of personal data about users (e.g., "*if you think about it, how much of your own data is stored somewhere*" (S1, pos. 24); similarly by S2 and S8) and perceived data models as related to the person and characterising the person. Some students described the aggregation of data models about different people (e.g., "*they look at what they might enjoy, based on what other users have enjoyed*" (S1, pos. 6) or "*they look at what you like and what others have liked [...], where there are parallels to others*" (S2, pos. 6)), which is related to the idea of collaborative filtering as addressed in the teaching unit's example. Furthermore, some students recognised the use of data models about the users for different purposes and in other contexts. For example, they recognised that the collection of data from past interactions could influence future interactions, which is an essential aspect of such data models (e.g., "*perhaps, if you have this now, you can suggest a similar search term again next time*" (S1, pos. 36)).

5.2. *Results Regarding the Perceived Relevance and Value: What have I Learned?*

Having examined students' understanding of the concepts in the framework, for RQ2, we analysed the data to uncover what had been learned through the eyes of the learners, that is, to identify students' perspectives on the value and relevance of learning about the concepts. Using thematic qualitative content analysis, we inductively found four categories for the relevance of learning the concepts (see Fig. 5). As an introduction to these findings, we use the following exemplary dialogue from one interview (S1, pos. 9–14) to illustrate the four categories:

> *Researcher:* Do you think it [recommendation systems] is an important topic to learn about?
>
> *Student:* I think it's important that one just understand more, these are everyday things, almost everything on the internet has a recommendation system so that one knows what's happening and why it's happening. So that one knows for oneself what one is doing.
>
> *Researcher:* So would you say it was interesting and exciting to see how it actually works?
>
> *Student:* Yes.
>
> *Researcher:* Do you see things differently now, for example, when you use apps on your mobile phone in your everyday life?
>
> *Student:* So now you pay more attention to where it might be used, where people are looking to see what I like and what I can get recommended. So maybe you think about that more now than you did before.

The student in this dialogue mentioned four categories we interpreted as perceived relevance in learning about ddA. He saw a relation to his everyday life and found it interesting and exciting to know about the inner workings of data-driven technologies, that is, to open the black box of such digital artefacts. Then, he described learning about

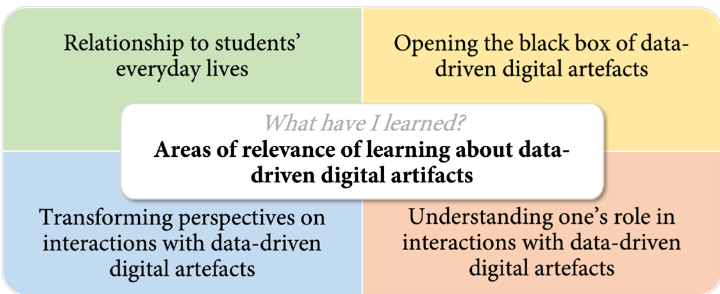

Fig. 5. Students described that learning about ddA during the teaching unit has relevance and value according to four aspects. (*Note:* Fig. 6 builds on these aspects and summarises the results in more detail, thus presenting a preliminary model as one of the main results of this study.)

his interactions with ddA and explained that he had gained a different perspective on interaction systems with such digital artefacts and was thinking about them differently. In summary, the student reported four aspects of why learning about ddA in the intervention was relevant to him, which could also be found in other interviews. Fig. 5 provides an overview of these areas, which are based on the inductively generated code categories (see Table 2) and are described in more detail below.

5.2.1. *Relationship to Students' Everyday Lives*

All participants made connections between their everyday lives and learning about ddA in the teaching unit. They related the concepts learned to their own experiences and provided examples from their everyday lives. For instance, when describing the perceived relevance of the teaching unit, one student explained that it was about "*understand[ing] more about these everyday things, almost everything on the internet has a recommendation system*" (S1, pos. 10). Another student described this connection as "*you can also relate that to yourself* " (S6, pos. 6). Other students also recognised that the topic of the teaching unit was related to their everyday lives (e.g., "*how data then affect the real world, what impact they can have*" (S3, pos. 4)). This indicates that the students were able to make relations between the concepts learned in the teaching unit and their own everyday experiences and thus found them personally meaningful. The fact that the students related the topic of ddA to their everyday lives suggests that the perspective

Table 2

Findings according to code categories regarding RQ2

Code Category	Description	Examples
Relationship to students' every-day lives	Students recognised ddA in their every-day lives and could relate learning about ddA to their everyday experiences	"*you can also relate that to yourself*" (S6, pos. 6); "*that one understands more about these everyday things*" (S1, pos. 10)
Opening the black box of ddA	Students perceived understanding the inner workings of ddA and their data practices as (a) interesting and surprising and (b) as important and useful	"*because one can notice that a lot of us have wondered now about all the data that are collected [...] and what they are used for*" (S5, pos. 37); "*I think it's important that one just understand more [...] so that one knows what's happening and why it's happening*" (S1, pos. 10)
Understanding one's role in interactions with ddA	Students described that what they have learned about ddA supports them (a) to understand and assess their actions, (b) to form opinions about ddA and identify possible actions accordingly, and (c) to act according to their intentions	"*that you then also know more about what you are actually doing*" (S1, pos. 16); "*useful to know [...] because then you can deal with it better*" (S5, pos. 37); "*important to know for people for whom this is a problem, who don't want it but don't know it and always click and act and do everything*" (S2, pos. 10)
Transforming perspectives on the interactions with ddA	Students reported that they have gained other perspectives on ddA and their interactions with ddA, which have encouraged them to think differently about such situations	"*I actually pay a bit more attention to it than I did before the teaching unit*" (S3, pos. 10); "*So maybe you think about that more now than you did before*" (S1, pos. 14)

provided by the data awareness framework facilitated their engagement with the subject matter. By linking the theoretical concepts to their personal experiences, the students grasped the significance of ddA in their daily interactions and identified situations in which they interact with ddA.

5.2.2. *Opening the Black Box of Data-Driven Digital Artefacts*

During the interviews, students mentioned that they found it relevant to learn about the details of ddA and the role of data, that is, to open the black box of data-driven technologies, which covers two aspects: (1) it is interesting and surprising, and (2) it is important and useful.

(1) *It is interesting and surprising to open the black box.* Some students found it interesting to learn about ddA and data practices, for example, to understand how personal information can be collected and how outputs can be personalised. For instance, one student described this process of considering technical aspects of ddA as exciting ("*I definitely found it exciting how it works with these algorithms, that they can track and aggregate everything from so many people and how it all works*" (S2, pos. 4)). Another participant mentioned that it was interesting and useful for everyday life ("*I also find it interesting to know about how you can find out about a person and I also find it interesting with this personalisation related to Netflix, because you can always use it*" (S6, pos. 12)). In addition, some students reported being surprised about the data collection and processing methods used by ddA. One student reported his observation about the class: "*because one can notice that a lot of us have wondered now about all the data that are collected [...] and what they are used for*" (S5, pos. 37). Another student's response supports this when he states: "*I found it quite interesting with the recommendation systems because one didn't really think so much about it that so much is always tracked about you*" (S3, pos. 4). Later, he mentioned: "*Normally, when I call someone, I call someone, so now I think a bit of this and that is happening. I find that a bit impressive*" (S3, pos. 36). This suggests that the students valued examining ddA, especially regarding their inner workings (i.e. opening the black box of these ddA and learning about the technical details behind the user interface).

(2) *It is important and useful to open the black box.* Most participants emphasised the importance of gaining knowledge about ddA and their data practices. They further recognised that it is important to understand the role of these technologies in everyday life (i.e. their impact). For instance, one student said: "*I think it's important that one just understand more [...] so that one knows what's happening and why it's happening*" (S1, pos. 10). When talking about the topic of the teaching unit, another student replied: "*For me personally, it's important what you can find out on the internet, what's dangerous, so I think that's the most important thing for everyone*" (S6, pos. 12). Later she added: "*that you know what happens to the data, where it goes, and I think you have to learn that, so this unit. So I think that's really important*" (S6, pos. 14). Thus, she emphasised the importance of learning about the technical aspects. Similarly, some students grasped the usefulness of learning about the concepts of ddA, particularly in understanding the real-life consequences of ddA. For example, when talking about recommendation sys-

tems and the relation of the topic to his everyday life, one student said: "*I also found it very interesting how data can have an impact in real life, what effects it can have*" (S3, pos. 4; similarly described by S1). In summary, students emphasised the relevance of opening the black box according to the framework's concepts to understand ddA and their data practices.

5.2.3. *Understanding one's Role in Interactions with Data-Driven Digital Artefacts*

In addition, the students mentioned that adopting the data awareness perspective on ddA and the role of data in their daily interactions with ddA is relevant. They emphasised the usefulness of learning about ddA and their data practices in understanding these processes within their daily interactions with ddA. For example, one student expressed the value of this perspective by stating: "*I could also learn something from it that I can use in my everyday life*" (S6, pos. 6). During the category-based analysis of the interview data, we have explored these links to everyday actions in more detail and inductively identified three aspects in which learning about ddA is relevant to students' everyday interactions with ddA: The students mentioned that it supports (1) understanding and assessing one's actions, (2) assessing ddA and forming opinions about their role in one's everyday life and identifying possible actions, and (3) weighing up possible actions and acting according to one's intentions. The following paragraphs describe these three aspects in more detail.

(1) *Understanding and assessing one's own actions.* The participants recognised that adopting the data awareness perspective helped them understand the meaning of their actions when interacting with ddA. For example, one participant said: "*you then also know more about what you are actually doing*" (S1, pos. 16). Another student reported a change in his perspective on his actions (e.g., "*Through the teaching unit, I personally became more aware of what it means to click on accept*" (S3, pos. 4)). Another student emphasised the transformative effect of data awareness on his understanding of the consequences of his actions, as he stated: "*who don't want it but don't know it and always click and act and do everything*" (S2, pos. 10). These statements indicate that the data awareness perspective facilitates a deeper understanding of one's actions and, therefore, could provide a basis for informed decision-making.

(2) *Forming opinions about ddA and identifying possible actions.* Several participants mentioned that the data awareness perspective encouraged them to think about ddA, form opinions about ddA, and support them in identifying possible actions. When talking about the value and importance of learning about ddA and their inner workings, one student said: "*that you could also realise [...] whether it might bother you when that happens*" (S1, pos. 16). This indicates that a more detailed understanding of ddA could help to assess one's interactions with ddA and to form opinions about these ddA. In addition, the students recognised a need to adopt the data awareness perspective, think about their actions (e.g., when entering personal data), and consider alternative courses of action more often. For example, one student highlighted the lack of awareness among others and expressed the importance of considering actions related to data collection (e.g., "*Because many, even in the 5th and 6th grade range, who have Netflix, for example, are not*

aware that so much is tracked and that one should not always simply click on 'accept', perhaps one should also click on 'decline'. And I think that one could take a look at it more often" (S3, pos. 8)).

(3) *Acting according to one's intentions.* The data awareness perspective could help students break out of prescribed interactions and act more agentically. One student mentioned that understanding ddA enables them to deal with its impact and make choices that are in line with their goals: "*it's quite useful to know when you're thinking about it, because then you can deal with it better, then it would also help some people to sit less at the computer or something like that*" (S5, pos. 37). Students also emphasised that without an understanding of ddA and their data practices, individuals may not be able to act according to their intentions and may unknowingly follow prescribed actions (e.g., "*I think it's particularly exciting and important to know for people for whom it's a problem, who don't want it but don't know it and always click and do everything*" (S2, pos. 10)). In addition, some students reported that they were more mindful of their actions and considered alternatives more often (e.g., "*I don't always click 'accept' immediately, but if you click 'decline' and you can still use the app, I sometimes click 'decline' too*" (S3, pos. 10)).

5.3. *Transforming Perspectives on Interactions with Data-Driven Digital Artefacts*

Furthermore, the analysis of the interview data revealed that the students had gained other perspectives on interaction systems with ddA from their everyday life and their behaviour in such situations. Four of the six students mentioned that their perspectives on their interactions with ddA had changed and that the teaching unit was like an eye-opener. For example, one student reported that he thought about the data practices and the ddA more often ("*So maybe you think about that more now than you did before*" (S1, pos. 14)). When describing his opinion about the intervention, one student mentioned thinking about data differently: "*I thought the teaching unit was very good because now you can better engage with what data actually is*" (S3, pos. 36; similarly by S6). In addition, this student also described a change in the perspective on one's actions ("*I pay a bit more attention to it than I did before the teaching unit. I don't always click 'accept' immediately, but if you click 'decline' and you can still use the app, sometimes I click 'decline' too*" (S3, pos. 10)). Moreover, when describing the teaching unit, one student said:

> *"I found it quite interesting with the recommendation systems because one didn't really think so much about it that so much is always tracked about you. Usually, I go to Netflix and just click on accept. Through the teaching unit, I personally became more aware of what it means to click on accept."* (S3, pos. 4; similarly by S2)

This indicates that the students were encouraged to change their perspectives when interacting with ddA in their everyday lives, that is, to reflect on their view of their everyday interactions with ddA.

6. Discussion

In this section, we discuss the study's findings according to the research questions and recent research.

6.1. *Overview of Findings*

Findings about the learning outcomes regarding the concepts (RQ1)
The study results revealed that most students could identify the role of data in interactions with ddA according to the concepts outlined in the data awareness framework. Students could describe explicit and implicit data collection and identify primary and secondary data processing purposes according to a ddA not discussed in the previous teaching unit. In addition, most students had an idea of data models about users. Thus, they have nuanced perspectives on ddA and their data practices and are likely to understand the role of data in interactions with ddA. Therefore, the results indicate that the students understood the concepts of the explanatory model, suggesting that it is comprehensible and usable by school students.

Findings about the perceived relevance of learning about the concepts (RQ2)
The results indicate that the students found learning about ddA according to the data awareness framework relevant across four different areas (see Fig. 5). Firstly, students could relate the concepts they learned to everyday situations and their experiences from interacting with ddA. They emphasised that the concepts about ddA were relevant because they allowed them to understand the inner workings and impacts of the digital artefacts they know from everyday life. Secondly, the students found that opening the black box of ddA and understanding the technical aspects was valuable in itself. They highlighted that they found it interesting, surprising, important and useful to learn about ddA according to the framework's concepts, that is, to engage with ddA and open the black box to understand ddA, especially from a technological perspective. Thirdly, students mentioned that they found the data awareness perspective on ddA (i.e. analysing ddA according to the model in Fig. 2 and reflecting on the interaction systems) relevant for understanding their role in interactions with ddA. Students described understanding ddA and data practices as beneficial for informed decision-making. They emphasised that it helps them to understand and assess such interactions with ddA, form opinions about the role of ddA in their everyday lives, reflect on their actions, identify alternative actions, and choose actions. They described it as supporting them in breaking out of prescribed behaviour and instead acting according to their intentions for how to interact with ddA. Finally, students mentioned that their perspectives on ddA, their interactions with ddA and the digital world evolved during the intervention. Thus, engaging with ddA during the teaching unit not only opened their eyes but also encouraged them to think more often or differently about their interactions with ddA. Overall, the findings indicate that learning about data-driven technologies, as the data awareness framework suggests, appears meaningful to students and has value and relevance to their everyday lives. They mentioned different aspects of how learning about ddA becomes relevant.

We have summarised these findings as a preliminary model of the relevance of learning about ddA for developing agency in the digital world, a model that can potentially inform and shape educational practices (see Fig. 6).

At the end of the analysis, we noticed that some students mentioned the four areas of the preliminary model in a specific order. For example, this could be observed in the dialogue quoted at the beginning of Section 5.2. The student began by talking about the relationship to everyday life, then described understanding the inner workings of the ddA, then mentioned his interactions with ddA and finally explained a change in perspective. Similar lines of argumentation can also be found in the interviews with two other students. Thus, we have included this order in the preliminary model (see Fig. 6), but we assume other argumentation orders are also possible.

6.2. *Discussion of Findings*

In the following, we delve into these findings, interpret them theoretically, and discuss links with prior studies and related research.

Understanding data-driven digital artefacts. Previous research has indicated that students struggle to understand the data practices of ddA, as discussed in Section 2.1. In addition, the asymmetry between users and ddA and the opacity of ddA and their data practices may further challenge students' understanding of ddA (e.g., Burrell, 2016; Zuboff, 2019; Brunton and Nissenbaum, 2015; Denning and Denning, 2020). Thus, recent research has highlighted that students often lack awareness and understanding of the role of data in their interactions with ddA. In contrast, our study suggests that, following the data awareness intervention, participants were able to identify data practices in interactions with ddA. Many students could describe the collection and processing of data by a ddA from a nuanced perspective and understood the role of data in an exem-

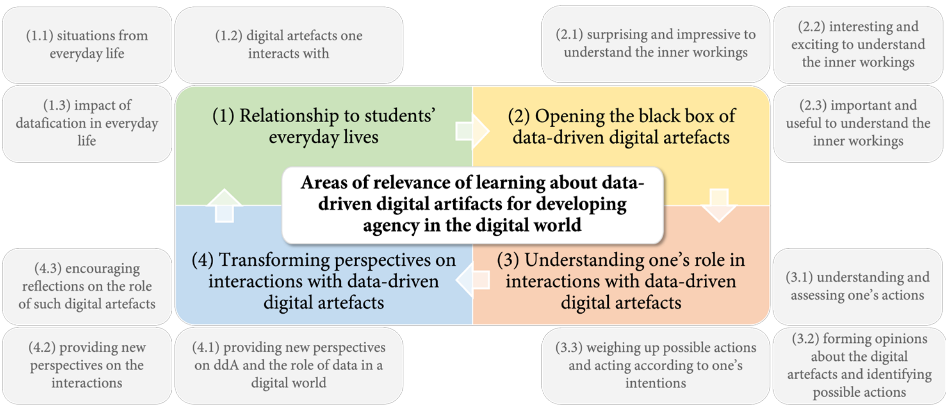

Fig. 6. This preliminary model was derived from the results of the second research question. The model summarises the results for the relevance that students perceived in learning about ddA during the data awareness intervention. For each of the four areas, the study revealed results summarised as three aspects.

plary situation. Therefore, the framework's approach of analysing ddA according to the explanatory model and reflecting on interactions with ddA seems to enable students to identify the concepts and analyse given ddA accordingly. This indicates that learning the explanatory model can empower students to understand ddA and their data practices in everyday life.

Relevance of learning concepts of data-driven digital artefacts. Some studies report that teaching about data or data-driven technologies could be effective in terms of gaining knowledge and skills, but students often struggle to relate this knowledge about data and data-driven technologies to their everyday lives (Livingstone *et al.*, 2019; Bowler *et al.*, 2017; Gebre, 2018; Vartiainen *et al.*, 2021). For example, studies report that students perceive data as impersonal and unrelated to their personal lives (e.g., Gebre, 2018; Bowler *et al.*, 2017). However, supporting students to develop a personal and meaningful relationship with the content when learning about ddA is important, as studies have reported positive impacts on learning ML when using personal data (Register and Ko, 2020). Similarly, Bilstrup *et al.* (2022) argue for engaging students with individual and critical perspectives when learning about data literacies, for example, by supporting emotional engagement through the teaching examples. However, while the intervention introduces an explanatory model about ddA and enables students to analyse and reflect on ddA accordingly, the findings indicate that the students could relate the framework's concepts to their everyday experiences and found it relevant to understand them (see Fig. 6).

Regarding peoples' feelings of powerlessness and resignation, prior research indicates that people often feel unable to take control over and counteract the data collection and processing by ddA, as discussed in Section 2.1. Such powerlessness may be due to a lack of knowledge or skills about controlling personal data, users' perceived limitations in taking appropriate action and their beliefs that privacy-protective behaviours are ineffective and useless (Lutz *et al.*, 2020; Hargittai and Marwick, 2016). Moreover, ddAs' tracking and profiling practices can influence users' attitudes, emotions and behaviours (rather not apparent for the people) (e.g., West, 2019; Tufekci, 2014; Susser *et al.*, 2019; Kramer *et al.*, 2014; Zuboff, 2019). Some students may fear social exclusion if they act differently, and quitting interactions with ddA may have further disadvantages (Pangrazio and Selwyn, 2020). In addition, students may not perceive the impact of ddA on their self-determination, which hinders their motivation to take action (Keen, 2020). In summary, there are several challenges to empowering students to make informed and self-determined decisions about interacting with such technologies. In the interviews, participants attributed significance to understanding and reflecting on their role in everyday interactions with ddA (see Fig. 6). This indicates that students may have adopted a data awareness perspective in everyday interactions, potentially mitigating feelings of powerlessness and lack of control. Hence, data awareness may empower students to engage in self-determined interactions with ddA, enabling them to act according to their intentions and effectively navigate the challenges associated with self-determination and developing their digital selves. Additionally, students mentioned that the intervention transformed their perspectives on their interactions with ddA (see Fig. 6). Thus, the students may be encouraged to reflect on the role of ddA in their lives and their role in the digital world.

Empowerment and everyday agency in a data-driven world. One of the main aims of K-12 computing education (for all) is to cultivate students' agency and empowerment, which includes preparing them for independent and informed participation in a digital and data-driven society. In this study, we have explored the contribution of data awareness to this goal by interviewing students and reconstructing their view on the learning outcomes in terms of their relevance to students' everyday needs, that is, supporting the development of agency in a digital world. The concept of agency is discussed from various perspectives across different disciplines (for example, see for overviews, Eteläpelto *et al.*, 2013; Biesta and Tedder, 2007; Emirbayer and Mische, 1998). Giddens (1984), for instance, associates agentic actions with individuals' intentions, while Bandura (2001) defines agency as the intentional capacity to make things happen through one's actions. In the context of the digital world, Couldry (2014) characterises agency as reflective actions and sense-making processes to navigate and act in the world. Hence, in a digital and data-driven environment, agency in everyday life involves understanding and reflecting on one's actions and making intentional decisions. Since data practices shape interactions with ddA, which necessitates understanding ddA and their data practices within this interaction context, fostering agency becomes crucial (Tufekci, 2014; Susser *et al.*, 2019; Schulte and Budde, 2018). In the interviews, the students reported that they felt supported in understanding, identifying and performing actions according to their intentions and that their perspectives on ddA had changed (see Fig. 6). This indicates that their agency may be enhanced. Our findings also indicate that students were empowered to analyse, understand and reflect on an exemplary ddA. Thus, the students can engage with ddA in the sense of being able to analyse and reflect on the ddA according to the framework's explanatory model so that their empowerment may be fostered (see for empowerment idea, Dindler *et al.*, 2020). Therefore, developing data awareness may facilitate students' development of empowerment and agency when interacting with ddA in their everyday lives.

7. Limitations

The study has several limitations. Firstly, the sample was a convenience sample with a small number of participants, which may limit the results, particularly regarding the generalisability of the findings. However, the results demonstrate different dimensions of relevance that students perceive when learning about ddA. Secondly, we could have asked more about specific aspects during the interviews. For example, it would have been interesting to know what other possible actions students had in mind when they talked about identifying possible actions. In addition, when students described their actual behaviour or intentions to act, we cannot be sure whether participants behave this way or reflect on it more often, even if they told us so. Thirdly, when obtaining students' answers in interviews, participants may respond politely and supportively, leading to a social desirability bias. However, to reduce these risks, we intended to ask more implicitly about the perceived relevance of learning about ddA and the role of data (e.g., what the teaching unit was about). Finally, due to the study's design, we do not have systematic insights into students' prior knowledge, so the causes of the intervention on students' perceptions and understanding are unclear. However, as we were interested in exploring the relevance that students perceived in learning about ddA, the focus was on students' retrospective views. To be able to make adequate statements about this, we also examined the extent to which they understand ddA and the role of data as characterised in the data awareness framework. In line with the research questions, we did not intend to make interpretations about the changes due to the intervention and evaluate the framework's effectiveness; instead, we focused on exploring learners' views on the explanatory model addressed. However, the study offers ideas and conjectures about the effects of the data awareness framework, which need to be evaluated in further research.

8. Conclusions and Outlook

The study presented here addresses the need for research in K-12 computing education that aims to develop approaches for teaching concepts that are relevant to understanding data-driven digital artefacts (ddA) and thus navigating in a digital and data-driven world. The study's approach follows the data awareness framework, which embeds concepts about data-driven technologies in an explanatory model by focusing on the role of data (see Fig. 2). Its idea is to teach this model and enable students to use it as a lens for analysing and reflecting on ddA. The study includes a teaching unit that encourages students to examine the inner workings of recommendation systems used in streaming services, to reflect on such ddA and their data practices, and to use the concepts learned to explore other ddA they use in their everyday lives. The concepts embedded in the framework's model are not meant to be "the" data or ML concepts from the CS discipline. Instead, they are intended to support students in finding an explanation for the ddA they interact with. The framework aims to make these concepts meaningful and useful in students' daily lives to develop agency and empowerment in navigating the digital and data-driven world.

The exploratory study indicates that, after our intervention, students have understood the concepts of ddA (see blue flow in Fig. 7) and have recognised that learning these concepts is relevant (see green flow in Fig. 7). We have inductively generated categories of the relevance of learning about ddA that students have mentioned (see Table 2), which stretches over four areas: (1) relationship to students' everyday lives, (2) opening the black box of ddA, (3) understanding one's role in interactions with ddA, and (4) transforming perspectives on interactions with ddA (see overview in Fig. 6). This indicates that the approach encourages students to engage with data-driven technologies and to develop a perspective that allows them to understand and reflect on ddA that they find useful and relevant to their everyday lives. Hence, the framework may foster students' engagement in making sense of data-driven technologies and their inner workings, becoming more informed and reflective. Thus, the study provides preliminary insights that data awareness may support students in developing empowerment (see for empowerment idea, Dindler *et al.*, 2020). According to the discussed meaning of agency

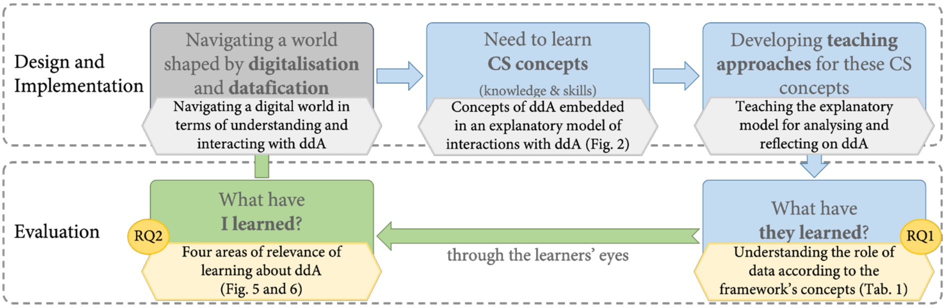

Fig. 7. The main parts of this study and its findings can be located in the typical research steps and questions in CSER for evaluating approaches to teaching CS-related concepts, as previously shown in Fig. 1.

(e.g., Couldry, 2014), the study indicates that data awareness may support students in overcoming powerlessness and developing an agency in a digital world. Therefore, our framework could be an example of how K-12 computing education can make concepts of data-driven technologies meaningful to students' experiences and usable for navigating a data-driven world.

Based on the findings for RQ2 (see Table 2), we have created a preliminary model as shown in Fig. 6. It describes areas in which learning about ddA could become relevant for students in developing agency in the digital world. This model needs to be evaluated in future research and may be extended to other areas. However, this model contributes to computing education research and has the potential to reveal areas to be addressed in the development and evaluation of teaching approaches. We believe that this model can guide how computing education for all can teach computational concepts meaningfully, making them relevant and useful for students in everyday interactions with digital artefacts.

The study also uncovers new questions for future research. For example, it indicates that students found using the data awareness perspective relevant to analysing, understanding, and assessing ddA. However, whether students will use this perspective in their daily interactions with ddA remains to be seen. Therefore, it would be interesting to explore in more detail which role the explanatory model of the framework and the respective concepts play in students' daily lives or whether they engage with ddA accordingly. Moreover, concerning the different roles when engaging with digital technologies as described by Dindler *et al.* (2020), data awareness may support bridging the gap between analysing, understanding and reflecting ddA and shaping the digital world. Based on the findings, students can probably engage with ddA, delving into the inner workings of ddA and developing opinions on technological developments in the digital world. Thus, data awareness could be a step towards supporting students in reimaging the technological developments of ddA. Therefore, it would be fruitful to explore how data awareness can be the basis for engaging in designing and shaping digital artefacts, that is, switching to a designer perspective (see, Fischer, 2002; Schulte and Budde, 2018). For example, future implementations of the framework could include tasks for reimagining ddA after having reconstructed and reflected on the role of data in a given ddA.

The study's methodological approach raises an interesting question: Could it be that there was no need to explicitly inquire about the aspects depicted in the green flow in Fig. 1? Interestingly, the relevance expressed by the students, as derived from the interview data, closely mirrors the general educational rationale for teaching data and data-driven technologies in K-12 computing education, as outlined in the upper left corner of Fig. 1. This seems to suggest that if an intervention is thoughtfully constructed and designed, it may be sufficient to concentrate on empirical evaluations of the components as highlighted in blue in Fig. 1. However, as we have seen in the related research, it is evident that this alignment between students' perceived value and the educational rationale is not always observed as they struggle to relate and apply these concepts to their everyday lives. Therefore, it is interesting to make the educational rationale more explicit in computing education research and to examine more specifically the relationships shown in Fig. 1. In the future, it would be interesting and probably necessary to explore approaches to measure the impact or value of teaching computer science (to all). Hence, future research could take more systematic approaches to evaluate relations as observed in this study.

Based on this study, we believe teaching explanatory models about digital technologies in K-12 computing education could be beneficial. The findings suggest that learning an explanatory model about data-driven technologies allows students to use it as a lens for digital artefacts to understand the inner workings. Moreover, it supports them in understanding and reflecting on their role in interactions with such technologies and provides new perspectives on the digital world. Thus, teaching explanatory models may help students understand the digital technologies they use daily and further support fostering agency and empowerment in navigating the data-driven world.

References

Akkerman, S., Admiraal, W., Brekelmans, M., Oost, H. (2008). Auditing Quality of Research in Social Sciences. *Quality & Quantity*, 42(2), 257–274. `https://doi.org/10.1007/s11135-006-9044-4`

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology*, 52(1), 1–26. `https://doi.org/10.1146/annurev.psych.52.1.1`

Bennett, J., Lubben, F., Hogarth, S. (2007). Bringing Science to Life: A Synthesis of the Research Evidence on the Effects of Context-based and STS Approaches to Science Teaching. *Science Education*, 91(3), 347–370. `https://doi.org/10.1002/sce.20186`

Biesta, G., Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults*, 39(2), 132–149. `https://doi.org/10.1080/02660830.2007.11661545`

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. `https://doi.org/10.1145/3501712.3530783`

Bode, M., Kristensen, D.B. (2016). The Digital Doppelgänger within: A Study on Self-Tracking and the Quantified Self Movement. In: Canniford, R., Bajde, D. (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.

https://doi.org/10.1038/nature11421
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. https://doi.org/10.1002/pra2.2017.14505401004
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. https://doi.org/10.1080/1369118X.2016.1154086
Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1). https://doi.org/10.1177/2053951715622512
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. https://doi.org/10.1186/s40594-023-00418-7
Caspersen, M.E., Diethelm, I., Gal-Ezer, J., McGettrick, A., Nardelli, E., Passey, D., Rovan, B., Webb, M. (2022). *Informatics Reference Framework for School*. National Science Foundation. https://doi.org/10.1145/3592625
Couldry, N. (2014). Inaugural: A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World. *The Sociological Review*, 62(4), 880–897. https://doi.org/10.1111/1467-954X.12158
Denning, P.J., Denning, D.E. (2020). Dilemmas of Artificial Intelligence. *Comm. of the ACM*, 63(3), 22–24. https://doi.org/10.1145/3379920
Dindler, C., Smith, R., Iversen, O.S. (2020). Computational Empowerment: Participatory Design in Education. *CoDesign*, 16(1), 66–80. https://doi.org/10.1080/15710882.2020.1722173
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. https://doi.org/10.1145/3392063.3394410.
Druga, S., Ko, A.J. (2021). How Do Children’s Perceptions of Machine Intelligence Change When Training and Coding Smart Programs? In: *Interaction Design and Children*. ACM, Athens Greece, pp. 49–61. https://doi.org/10.1145/3459990.3460712
Emirbayer, M., Mische, A. (1998). What Is Agency? *American Journal of Sociology*, 103(4), 962–1023. https://doi.org/10.1086/231294
Eteläpelto, A., Vähäsantanen, K., Hökkä, P., Paloniemi, S. (2013). What is agency? Conceptualizing professional agency at work. *Educational research review*, 10, 45–65.
Fischer, G. (2002). Beyond “Couch Potatoes”: From Consumers to Designers and Active Contributors. *First Monday*. https://doi.org/10.5210/fm.v7i12.1010
Gebre, E. (2022). Conceptions and Perspectives of Data Literacy in Secondary Education. *British Journal of Educational Technology*, 53(5), 1080–1095. https://doi.org/10.1111/bjet.13246
Gebre, E.H. (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. https://doi.org/10.1007/s42330-018-0034-z
Giddens, A. (1984). *The constitution of society: outline of the theory of structuration*. University of California Press.
Gilbert, J.K. (2006). On the Nature of “Context” in Chemical Education. *International Journal of Science Education*, 28(9), 957–976. https://doi.org/10.1080/09500690600702470
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. https://doi.org/10.1145/3555576
Hargittai, E., Marwick, A. (2016). “What Can I Really Do?” Explaining the Privacy Paradox with Online Apathy. *International Journal of Communication*, 10, 3737–3757.
Hitron, T., Orlev, Y., Wald, I., Shamir, A., Erel, H., Zuckerman, O. (2019). Can Children Understand Machine Learning Concepts?: The Effect of Uncovering Black Boxes. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, pp. 1–11. 10/ghnn97.
Höper, L., Schulte, C. (2023). The Data Awareness Framework as Part of Data Literacies in K-12 Education. *Information and Learning Sciences*. https://doi.org/10.1108/ILS-06-2023-0075
Keen, C. (2020). Apathy, Convenience or Irrelevance? Identifying Conceptual Barriers to Safeguarding Children’s Data Privacy. *New Media & Society*, 24(1). https://doi.org/10.1177/1461444820960068
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. https://doi.org/10.1073/pnas.1320040111
Kuckartz, U. (2014). *Qualitative Text Analysis: A Guide to Methods, Practice & Using Software*. SAGE, Los Angeles.
Livingstone, S., Stoilova, M., Nandagiri, R. (2019). *Children'sData and Privacy Online: GrowingupinaDigital 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, pp. 1–16. https://doi.org/10.1145/3313831.3376727
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). https://doi.org/10.1177/1461444820912544
Martins, R.M., Von Wangenheim, C.G., Rauber, M.F., Hauck, J.C. (2023). Machine Learning for All!—Introducing Machine Learning in Middle and High School. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00325-y
Mayer-Schönberger, V., Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, Boston.
Mertala, P. (2021). The Pedagogy of Multiliteracies as a Code Breaker: A Suggestion for a Transversal Approach to Computing Education in Basic Education. *British Journal of Educational Technology*, 52(6), 2227–2241. https://doi.org/10.1111/bjet.13125
Mühlhoff, R. (2021). Predictive Privacy: Towards an Applied Ethics of Data Analytics. *Ethics and Information Technology*, 23(4), 675–690. https://doi.org/10.1007/s10676-021-09606-x
Nijenhuis-Voogt, J., Bayram-Jacobs, D., Meijer, P.C., Barendsen, E. (2021). Omnipresent yet Elusive: Teachers' Views on Contexts for Teaching Algorithms in Secondary Education. *Computer Science Education*, 31(1), 30–59. https://doi.org/10.1080/08993408.2020.1783149
OECD (2014). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. OECD. https://doi.org/10.1787/9789264196391-en
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. https://doi.org/10.1177/1461444818799523
Pangrazio, L., Selwyn, N. (2020). Towards a School-Based 'Critical Data Education'. *Pedagogy, Culture & Society*, 29(3), 431–448. https://doi.org/10.1080/14681366.2020.1747527
Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., *et al.*(2019). Machine Behaviour. *Nature*, 568(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y
Register, Y., Ko, A.J. (2020). Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In: *Proceedings of the 2020 ACM Conference on International Computing Education Research*. ACM, Virtual Event New Zealand, pp. 67–78. https://doi.org/10.1145/3372782.3406252
Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., … *et al.* (2015). Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. SSHRC.
Rizvi, S., Waite, J., Sentance, S. (2023). Artificial Intelligence Teaching and Learning in K-12 from 2019 to 2022: A Systematic Literature Review. *Computers andEducation: Artificial Intelligence*, 4, 100145. https://doi.org/10.1016/j.caeai.2023.100145
Sander, I. (2020). What Is Critical Big Data Literacy and How Can It Be Implemented? *Internet Policy Review*, 9(2). https://doi.org/10.14763/2020.2.1479
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. https://doi.org/10.1145/3279720.3279733
Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. In: *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems* (Vol. 2), pp. 2503–2511.
Sentence, S., Waite, J. (2022). Perspectives on AI and Data Science Education. In: *AI, Data Science, and Young People. Understanding Computing Education* Proceedings of the Raspberry Pi Foundation Research Seminars. rpf.io/seminar-proceedings-vol-3-sentance-waite

Shapiro, R.B., Fiebrink, R., Norvig, P. (2018). How Machine Learning Impacts the Undergraduate Computing Curriculum. *Communications of the ACM*, 61(11), 27–29. https://doi.org/10.1145/3277567

Susser, D., Roessler, B., Nissenbaum, H. (2019). Technology, Autonomy, and Manipulation. *Internet Policy Review*, 8(2), 1–22. https://doi.org/10.14763/2019.2.1410

Tedre, M., Denning, P., Toivonen, T. (2021). CT 2.0. In: *21st Koli Calling International Conference on Computing Education Research*. ACM, Joensuu Finland, pp. 1–8. https://doi.org/10.1145/3488042.3488053

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. https://doi.org/10.1109/FIE44824.2020.9274138

Tufekci, Z. (2014). Engineering the Public: Big Data, Surveillance and Computational Politics. *First Monday*, 19(7). https://doi.org/10.5210/fm.v19i7.4901

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. https://doi.org/10.1016/j.ijcci.2021.100281

West, S.M. (2019). Data Capitalism: Redefining the Logics of Surveillance and Privacy. *Business & Society*, 58(1), 20–41. https://doi.org/10.1177/0007650317718185

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). https://doi.org/10.15353/joci.v12i3.3275

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

L. Höper is a PhD student in Computing Education Research at Paderborn University, Germany. His main research interest is empirical research on teaching and learning processes in K-12 computing education. In his dissertation, he develops and evaluates the data awareness framework. Since 2020, he has been working on data awareness and other topics related to AI and data science education in schools in the ProDaBi project.

C. Schulte is a professor for Computing Education Research at Paderborn University, Germany. His work and research interests are the philosophy of computing education, artificial intelligence in education, and empirical research on teaching and learning processes (including eye movement research). Since 2017, he has been working together with Didactics of Mathematics (Paderborn University) on the ProDaBi project, in which data science and artificial intelligence are prepared as teaching topics. He is also a PI in the collaborative research centre 'Constructing Explainability' on explainable AI.

Appendix A. Questions from the interviews

1. *Description of the teaching unit:*
    a) What was the teaching unit about?
    b) What was the most exciting part of the teaching unit?
2. *Description of recommendation systems:*
    a) What can you tell me about recommendation systems? (If a student cannot answer the questions: What is a recommendation system used for? How does it work?)
    b) Do you think recommendation systems are an important topic to learn about?
    c) What do you think everyone should learn about recommendation systems?
    d) In which everyday situations do you find recommendation systems?
    e) What do you think about the use of recommendation systems?
3. *Application of the knowledge in another context:*
    a) How would you describe the terms explicitly and implicitly collected data?
    b) Description of the given context: Imagine the following situation. You are wondering about something and want to look it up online. For example, you could use a search engine on your mobile phone or computer. You then enter a search term and click on "search". This will bring up a page of search results.
    c) Describe this process in your own words, particularly where you recognise data collection in this scenario.
    d) Usually, the data is not just collected; what is the data you just mentioned used for? (If students cannot answer the question: Do you remember the terms primary and secondary data processing purposes?)
4. *Possible closing statement:*
    a) Is there anything else you would like to say about the teaching unit or the interview?

Note. The interviews were held in German. Therefore, we have translated the questions from the interview guideline from German into English.