1.   Introduction

1.1   Why Critical Data Education is Urgently Needed

How Digital and Data Technologies Affect Our Lives

Data-based technologies are a part of all our lives. In the United Kingdom, almost all (94%) adults are connected to the internet, and, on average, they spend nearly four hours online each day (Ofcom 2022). Both during this active use of the internet – using search engines, social media or online shopping sites – and during the time that people spend doing other activities – such as walking down the street, driving, working, shopping, exercising or using their smart home appliances – they encounter and disclose data to countless data-based systems. Most digital technologies today function by collecting, processing, analysing and outputting data with the help of algorithms. These data-based technologies increasingly exert influence over our lives and our societies, for example when data that is collected about individuals feed numerous predictive and automated decision-making systems, which are used to guide decisions such as who is considered for a new job position, who is likely to commit a felony, or whose children are at risk of being neglected (see e.g., O'Neil 2016; Kitchin 2021; Dencik et al. 2022).

In light of these ubiquitous influences of data technologies, scholars speak of the "datafication" of our societies. This neologism describes a "profound transformation in how society is ordered, decisions are made, and citizens are monitored through 'big data'" (Hintz et al. 2019, p. 2f). Key aspects of the datafication of our societies include the massive production of data which track human activity and behaviour; rapid technological advances in how data is being processed; and the social, economic, cultural and political consequences of these data technologies (Pangrazio and Selwyn 2020, p. 4). Besides hopes for financial savings, more efficiency and less subjective decision-making, many risks and controversial or even harmful impacts of these systems have been identified. As data and algorithms can never be truly neutral or "raw" (Dalton and Thatcher 2014, no page number) their use can lead to biased decisions and increased discrimination (O'Neil 2016; Redden 2018b). Moreover, these systems can come with risks for citizens' privacy; potentially increased surveillance; a 'chilling effect' on free speech; new opacities due to data technolo-

gies' 'black box nature'; new power imbalances; and risks for democratic societies (e.g., Kennedy and Moss 2015; Zuboff 2015; O'Neil 2016; Hintz and Milan 2018; Redden 2018b).

The Knowledge Gap

In addition to these risks, research on citizens' knowledge of how digital and data technologies function and how they affect our lives and our societies has repeatedly shown a "major understanding gap" (Doteveryone 2018, p. 5). Studies have demonstrated that the large majority of British people (83%) are not aware that information shared about them by others is collected or that free apps make money from data (70%) (Miller et al. 2018, pp. 17; 21). Two years after this initial study, the British citizens' understanding "remains shallow", especially with regards to technology companies' business models and the way they use data (Miller et al. 2020, p. 15). Even in 2022, the majority of respondents from a large international study did not know why Google and Facebook are free of charge to use or that they earn money from advertisers (Akman 2022, p. 22). However, studies also repeatedly show that people are concerned about the use of their data, and that the claim that most people do not mind disclosing their data because they feel like they have 'nothing to hide' does not hold true. Research from recent years has highlighted that the public is highly concerned about certain data uses and that they want more knowledge about the way their data is used; more control over how much data they share; the ability to enact their data rights; and better control and regulation of these systems along with more ethical and responsible data technologies (Grzymek and Puntschuh 2019; Miller et al. 2020; Kennedy et al. 2021b; Ada Lovelace Institute 2022).

Responses to Challenges around Datafication

In light of the risks related to data technologies, citizens' understanding gap and the clear concerns of the public, many scholars argue that *change is needed*. Several solutions have been suggested, including regulatory, tactical and educational responses. However, regulatory approaches come with many hurdles and the difficult balance of protecting citizens while ensuring the freedom to connect and share with others online (Pangrazio and Sefton-Green 2020, p. 212). Tactical responses, on the other hand, such as blocking or anonymisation tools, compromise users' internet experiences, require sophisticated technical skills and ongoing maintenance, and shift the responsibility of data protection to the individual (ibid.). Therefore, educational responses to datafication have been posited as "the *most plausible and successful strategy to combat the challenges of datafication*" (ibid., emphasis added). While better regulation of data technologies is still urgently required, critical education can be seen as a key component or even *prerequisite of legal and tactical responses* to datafication (ibid., p. 218). Moreover, education is generally often regarded as a preparation for future members of society (Wringe 2012, p. 2). In line with this, my study argues that an un-

derstanding of how data systems affect our lives and societies and the ability to make enlightened choices about one's technology use are essential prerequisites for participating in today's societies in an informed and empowered manner and are further necessary for establishing regulatory solutions that correspond with citizens' attitudes to these systems.

As a response to the urgent need for more education about data, many new educational approaches have emerged in recent years. Besides new definitions for *media literacy*, *digital literacy* and *data literacy*, a variety of further literacies has been suggested. These include information literacy, statistical literacy, coding literacy, data visualisation literacy, internet literacy, new media literacy, big data literacy or data infrastructure literacy (Philip et al. 2013; Frank et al. 2016; Gray et al. 2018; Hug 2019). In light of this wave of new literacies, some scholars even speak of the "literacification of everything" (Hug 2019, p. 146). Particularly the term *data literacy* has rapidly gained attention and popularity in recent years, both in academia and in mainstream debates and policy interventions. Today, data literacy is considered a key "future skill" (Schüller et al. 2019) that is fostered by numerous large European Union funded research projects,[1] various new projects and start-ups,[2] educational initiatives and foundations,[3] countless online courses[4] as well as by new curricula and university courses.[5] Despite this increasing interest in data literacy, there is a tremendous imbalance in the data literacy field. The overwhelming majority of data literacy projects and initiatives takes a practical, skills-based perspective on data literacy, usually aiming to foster people's ability to read, work with, analyse and argue with data (e.g., QlikTech International AB 2018; Schüller et al. 2021). More research on ways to increase critical understanding of datafication is urgently needed.

Why Technical Skills Are Not Enough

While learning how to use data can be important for today's and future citizens, such skills-based approaches often neglect to foster *understanding* of how the data technologies around us function, including their limitations, and *reflection* on how they affect our lives and societies. Although a small number of practical data literacy approaches include such questions (see chapter 2.2.2), the vast majority focusses

---

1   E.g., DigComp (EU Science Hub – European Commission [no date]), DATALIT (Taibi [no date]), the Data Literate Project (Urbšienė et al. [no date]), or DaLi (Wasson [no date]).

2   E.g., Data Literacy – Learn the Language of Data (Data Literacy, LLC [no date]), The Data Literacy Project (QlikTech International AB [no date]).

3   E.g., Data Literacy Education by the German Stifterverband ([no date]), "Data Literacy" by the Valhalla Foundation ([no date]).

4   E.g., The Data Literacy Course (Maugain 2022), Data Literacy Foundations (Rochester Institute of Technology [no date]), Data Literacy for All (Tableau Software LLC [no date]).

5   E.g., Data Literacy Certificate (Derntl [no date]), Strengthening Data Literacy across the Curriculum (Louie [no date]).

on *using* data for one's individual purposes. However, data usage skills alone cannot solve the risks and challenges of datafication outlined above. Moreover, recent research even suggests that teaching skills alone can in fact be *harmful*, particularly for children and young people. A 2021 evidence review examined 34 studies with 12- to 17-year-olds and found that "greater digital skills are linked, directly or indirectly, to more exposure to online risks" (Livingstone et al. 2021a, p. 21). Thus, the authors conclude that "teaching or promoting technical skills alone emerges as a problematic strategy" (ibid.). This is a worrying, yet crucial finding in light of the strong emphasis on technical skills in IT education and digital and data literacy approaches. A recent evidence review by the Ada Lovelace Institute adds to this picture, arguing that public concerns around data should be taken seriously, and that "simply raising awareness about the benefits of data will not increase public trust" (Ada Lovelace Institute 2022, p. 3).

Thus, my study argues that what is needed are not merely better skills or more awareness, but *critical education about datafication*. Citizens need to understand how the data-driven technologies that they encounter on a daily basis function, how these technologies collect and use their data, which implications these systems entail, and how citizens can enact their rights and better protect their data. Such critical literacy is essential for "democratic participation and social inclusion in the digital age" (Polizzi 2020b, p. 2). However, critical reflection of digital technologies is not something that people can learn on their own. Rather, research shows that in order to be critical of the digital, people need support, and to be prompted to critically reflect (Pangrazio and Selwyn 2020, p. 16).

In academic research, this perspective is examined by a small but fast-growing research field on *critical data literacy*, which has suggested educational approaches such as Big Data Literacy (D'Ignazio and Bhargava 2015), data infrastructure literacy (Gray et al. 2018), personal data literacies (Pangrazio and Selwyn 2019), data citizenship (Carmi et al. 2020), or my own previous suggestion of critical big data literacy (Sander 2020a; Sander 2020c). As argued by Pangrazio and Sefton-Green and demonstrated in chapter 2.2 of this study, however, a more complete theorisation of critical data literacy is required (2020, p. 208). Only few studies thoroughly review and analyse existing conceptual suggestions, and even fewer connect critical data literacy approaches to traditional educational concepts that this new and emerging field could learn from. Moreover, implementing the abstract objectives of an academic literacy concept into practice is challenging and there is a lack of practical models for educators (ibid., p. 211; 215). Thus, both an in-depth theoretical framework for critical education about datafication and research on ways to foster such education in practice are urgently needed. This study aims to contribute to both of these gaps in research.

1.2   Goals and Purpose of This Study

Background of the Study

It is not only these gaps in academic research that shaped this study, but also my previous experience in the field. In the years before starting this study, I investigated how online data literacy resources affect university students' privacy attitudes and their actions, and how they perceive these resources (research conducted in 2018/19: Sander 2020a); organised an international working group on critical data literacy together with the German education scholar Dr Harald Gapski in 2019; and, based on this working group, co-founded the "Critical Big Data and Algorithmic Literacy Network" – an international and interdisciplinary network for scholars and practitioners that Harald Gapski and I continue to co-coordinate. In the following years, I was involved in developing a database for critical data literacy resources and organising several events on critical data literacy, such as workshops and a debate series, both in the context of this network and other critical data literacy initiatives that I am involved in. These experiences equally demonstrated how much research and practice in terms of critical education about data already takes place, but also how many gaps remain and how under-researched some highly relevant literacy practices still are. In particular, the experiences of practitioners of critical data education and the connections of critical data literacy to traditional education scholarship urgently need more research.

The Study's Novel Approach to Researching Critical Data Literacy

My study's approach to addressing the gaps in research outlined above is novel in two regards: First, it takes a *holistic perspective* to developing a new, theoretically grounded framework for *critical datafication literacy* by combining an extensive theoretical analysis with empirical research and insights from practice. Already the frameworks' terminology highlights that this literacy takes a particular perspective to education about data: it aims for critical, systemic understanding of datafication and its implications, empowering learners to become active and enlightened citizens in datafied societies (see chapters 3.4; 7).[6] To develop such a novel and well-grounded literacy framework, not only existing digital and data literacy concepts were reviewed and analysed, but the study further examined what can be learnt from more traditional approaches in education research and from practitioners of critical data education. Three relevant educational approaches were analysed in detail and the unique characteristics of each of these well-established approaches were

---

6   In this study, the term "critical data literacy" is used as a generalising descriptor of different academic and practical approaches to critical education about data technologies. The term "critical datafication literacy" is used when referring to the specific literacy framework that is developed throughout the study.

identified and analysed in terms of what insights they offer for a more grounded theoretical framework for critical datafication literacy. This theoretical analysis not only considered educational approaches from English-speaking parts of the world, but my bilingualism also allowed me to analyse core theoretical concepts from the German academic discourse. Education research is a strong research field with a long history in Germany, which has produced a number of influential and internationally renowned theoretical concepts, for example "Bildung" (e.g., von Humboldt 1986). These concepts offer numerous points of connection to critical data literacy concepts that have hardly been explored so far.

The holistic nature of my study's theoretical framework is not limited to making new theoretical connections, but the study also aims to learn from practitioners of critical data education. This highlights the second regard in which my study takes a novel approach: it *closely interconnects theory and practice*. It does this in a number of ways. To begin, the study included a collaboration with the non-governmental organisation (NGO) Privacy International.[7] With over 30 years of experience in raising awareness and educating the public about risks surrounding digital technologies, Privacy International (PI) constituted an ideal collaboration partner to complement my own academic knowledge. Knowledge exchange with the NGO took place in several stages of the research, with the NGO acting as an advisor and providing important feedback on different methodological decisions.

Moreover, the theoretical and empirical research of the study are closely interconnected. The study empirically examined ways in which critical education about data already takes place, including the goals and strategies that practitioners follow. These findings offered valuable insights for the development of an in-depth framework for critical datafication literacy, as they reflected practitioners' year-long experiences in fostering critical understanding of data technologies. Furthermore, these findings on literacy practices could in turn help implement the critical datafication literacy framework developed in my study into practice. As Pangrazio and Sefton-Green highlight, the translation of abstract objectives of an academic literacy concept into "a practical model that can be operationalised by educators" is challenging (2020, p. 215). Learning about the strategies and approaches that are already used to educate about data provides a meaningful starting point for developing such practical models. To further support educators and interconnect theory and practice, the final stage of my study consisted of a knowledge mobilisation project conducted in

---

7   This study was funded by the Economic and Social Research Council Wales as part of the "Collaborative Doctoral Training Partnership" pathway. This pathway supports collaborative research projects between doctoral candidates and non-academic organisations, in this case the non-governmental organisation Privacy International. This collaboration was thus embedded in the project from the start, but how the collaboration was realised could be determined by the project partners.

close cooperation with Privacy International. In this project, the theoretical and conceptual findings on what critical datafication literacy should entail as well as the empirical findings on resources and strategies to foster critical education about datafication were mobilised for educators. Thus, my study takes the interconnection of theory and practice seriously by not only calling for more literacy education, but also providing educators with the information and the resources they need to foster critical datafication literacy.

Goals of the Study

Critical education about datafication can be fostered in a number of contexts, including in the classroom or in textbooks, but also for example in youth social work, in museums and exhibitions or through documentary movies. Pangrazio and Sefton-Green differentiate between formal data literacy pedagogies, personal pedagogies of data and "folk pedagogies of data" (2020, p. 217). The latter describe "a range of open, public programmes" that raise awareness and critical understandings of data "through creativity, visualisation and/or interactivity" (ibid.).

One format that fits into this category and that also constitutes one of the longest-standing and most established approaches to critical data education is what I would term *online critical data literacy resources*. This term describes web-based resources of various formats, such as websites, short videos, interactive tools, online courses and many others, that aim to raise awareness and critically educate about datafication. Online resources were used to educate about data as early as 2006,[8] and have developed into a popular approach with countless examples since then. They provide innovative, creative, diverse and often entertaining approaches to raising critical understanding of data. Nevertheless, these critical data literacy efforts have been examined in very few studies so far (e.g., Milioni and Papa 2019; Sander 2020c; Alegre 2021), and a comprehensive overview of such resources and research into their goals, educational strategies, and their creators' considerations are lacking. Thus, we know little about the perspective of their creators – one of the most important communities who currently implement critical education about datafication into practice – nor about the contexts in which these resources are used and how they are perceived.

---

8   For example the interactive website "Panopti.com" from 2006, which used animated short videos and texts to highlight surveillance in a citizen's daily life and informed about risks and ways to protect one's data. Unfortunately, the website is not available anymore.

My study attempts to address these gaps in empirical and theoretical research and aims to:

1) *Develop an in-depth theoretical framework for critical datafication literacy*, which not only takes existing critical data literacy concepts into consideration, but also learns from established education scholarship and from practitioners of critical data education;

2) *Investigate critical data literacy practices* – examining online educational resources about datafication and how they are created and used by educators as one way to teach about data;

3) *Mobilise the new knowledge for practitioners* of critical data literacy together with the NGO Privacy International.

The Purpose of the Research

The broader purpose of my study is to contribute to ongoing debates on what critical data education should entail and what such literacy education could look like in practice. The development of a framework for critical datafication literacy, which constitutes the core of my study, aims to provide a more complete theorisation of such literacy by taking theoretical insights and practical experiences into consideration, and to make a stronger call for *critical and reflexive* approaches to data literacy, which are often overlooked by large data literacy projects, agendas, policy initiatives or curricula such as the ones referred to above. Moreover, the study provides findings on strategies, approaches, formats and design options to implement this literacy into practice. The findings mostly relate to online critical educational resources as these are the key research objects of the study. However, it is likely that the identified strategies and approaches can be adapted to different contexts of critical data literacy education. These practical suggestions contribute to the highlighted gaps in data literacy approaches and practical models for educators. Through the knowledge mobilisation project that took place after the research and analysis were completed, Privacy International and I further aimed to directly support educators in teaching about topics around data by providing information on datafication, advice on how to find the right educational resource for a specific group of learners, and suggestions on how to create their own learning material.

1.3   Methodological Approach and Research Questions

In its first, theoretical stage, my study examined the research question "How can an extended critical datafication literacy be conceptualised?". Here, research on datafication and its implications for society was reviewed along with existing (critical) data literacy conceptualisations (chapter two). This was complemented by an in-depth

analysis of a small number of relevant theoretical approaches from traditional education scholarship (chapter three). Three educational approaches were selected as most suitable: education about (digital) media, including media literacy and digital literacy concepts; the German approach of "Bildung", including "politische Bildung"; and the Freirean "critical pedagogy". Key texts for each approach were selected, analysed in detail and key insights transferred to the conceptualisation of critical datafication literacy. Chapter 3.4 then presents a preliminary framework for this literacy based on the study's theoretical findings.

The second stage of this study comprised of different empirical approaches that aimed to investigate online critical data literacy resources from several angles. After discussing the applied methodological approaches in chapter four, chapter five addresses three research questions:

1) What is the range, shape and focus of *online critical data literacy resources*?
2) What are the goals, strategies, and experiences of *creators* of online critical data literacy resources?
3) What are *educators'* experiences with the topic of datafication and what do they need and wish for from a critical data literacy resource?

A quantitative content analysis of 75 resources was conducted in order to address the first research question and to provide an overview and initial analysis of existing online critical data literacy resources (chapter 5.1). After this, ten resources were selected out of those examined in the content analysis, and expert interviews with the creators of these resources were conducted to learn about their goals and strategies in critically educating about data (chapter 5.2). Finally, a qualitative survey with the second key community of critical data literacy practitioners – educators who are interested in teaching about digital and data technologies – was conducted (chapter 5.3). The goal here was to better understand the 57 participants' experiences and approaches in teaching about digital and data technologies, including whether or not they already use the priorly examined online resources; and to ascertain what educators' needs and wishes for critical data literacy resources are. Chapter six then interweaves the findings of all three empirical approaches and discusses them in relation to the academic literature. Finally, chapter seven presents the new critical datafication literacy framework that is developed throughout the study, reflects on the study's limitations and offers concluding remarks.

After the study's theoretical and empirical analyses were completed, the knowledge mobilisation project took place. Together with the NGO Privacy International, the knowledge developed in my study and insights from PI's practical experience were mobilised to support educators in their teaching about data. As a result, a learning resource was created that aims to provide educators with everything they need – understanding, critical perspectives and educational material – to foster

critical education about datafication. While this book highlights research findings that have influenced the creation process of this learning resource, the development of the resource itself was not part of the research and is thus not included in the following chapters. A report on the creation process and the outcome of the knowledge mobilisation project can instead be found in appendix IX.