RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

Big data literacy for youth: an intervention agenda

Literacia dos big data *para jovens: uma agenda de intervenção*

ANA KUBRUSLY

Interdisciplinary Centre of Social Sciences (CICS.NOVA), Faculty of Social Sciences and Humanities (NOVA FCSH) and Polytechnic Institute of Setubal (IPS), Portugal
a59022@campus.fcsh.unl.pt

LIDIA MARÔPO

Polytechnic Institute of Setubal (IPS) and Interdisciplinary Centre of Social Sciences (CICS.NOVA), Faculty of Social Sciences and Humanities (NOVA FCSH), Portugal
lidia.maropo@ese.ips.pt

SUSANA BATISTA

Interdisciplinary Centre of Social Sciences (CICS.NOVA), Faculty of Social Sciences and Humanities (NOVA FCSH), Portugal
susanabatista@fcsh.unl.pt

Abstract

Children and young people are growing up in increasingly datafied societies, with rapid developments in related digital technologies. In this context, it becomes essential to promote the necessary skills to identify, understand, and critically interact with (big) data, being able to deal with their impacts on society and daily life. Exploring the emergent field of research on youth's big data literacy through a thematic analysis of references selected by a scoping literature review, we identify relevant gaps and trends and propose an intervention agenda. Specifically, we aim to propose a larger understanding of young people's big data literacy and reflect on future paths — including central themes and pedagogical strategies — that can be used to enhance it. We identify four main topics for this agenda: 1) the defiance of big data's mythology and data-driven technologies as neutral and impartial tools; 2) situating the big data phenomenon within the digital economic ecosystem; 3) connecting big data to artificial intelligence's impact on society and 4) working on potential strategies to overcome youth's stances of apathy and indifference.

Keywords

big data literacy | datafication | big data | youth

Resumo

Crianças e jovens estão crescendo em sociedades crescentemente dataficadas, com rápidos desenvolvimentos de tecnologias digitais relacionadas a este fenómeno. Neste contexto, torna-se essencial promover um conjunto de competências para identificar, compreender e interagir criticamente com os *big data*, para que possam lidar com os seus impactos na sociedade e na vida diária. Através de uma *scoping review* da literatura, conduzimos uma análise temática das referências selecionadas, a fim de explorar o campo emergente da investigação sobre a literacia dos *big data* de jovens. Desse modo, identificamos lacunas e tendências relevantes e propomos uma agenda de intervenção. Especificamente, visamos construir uma maior compreensão da literacia dos *big data* de jovens, refletindo sobre os caminhos futuros — incluindo temas centrais e estratégias pedagógicas — que podem ser utilizados para a aprimorar. Nesse sentido, identificamos quatro temas principais para esta agenda: 1) o desafio à mitologia dos *big data* e às tecnologias baseadas em dados como ferramentas neutras e imparciais; 2) situar o fenómeno dos big data no ecossistema económico digital; 3) conectar os big data ao impacto da inteligência artificial na sociedade e 4) trabalhar com estratégias para superar as posições de apatia e indiferença dos jovens face à datafication.

Palavras-chave

literacia dos *big data* | datafication | *big data* | juventude

1. Introduction

Big data can be broadly defined as vast sets of heterogeneous data — in terms of format, origin, and content — that are collected, aggregated, stored, and analyzed constantly, having their origin associated with human actions in different social contexts, such as online shopping, use of social media, GPS systems, intelligent assistants, etc. (Daniel 2019; François, Monteiro, and Allo 2020). According to boyd and Crawford's (2012) classic definition, big data are built on the interaction between technology (the actual computational and algorithmic power that allows the collection and analysis of large data sets), analysis (the possibility to extract patterns from data sets to legitimize social, legal, and economic decisions), and mythology (the prevalent idea that data are neutral, objective, and accurate units of knowledge).

As a social, cultural, economic, and technological phenomenon, big data are closely related to the datafication of society, that is, the transformation of everyday actions (e.g. online shopping, liking posts on social media, using smart home devices or health-tracking apps) into quantifiable data (Mascheroni 2020). This process can be linked to the ideology of dataism, which sustains that everything transformed into data can be treated as "raw material, that can be analyzed and processed into predictive algorithms about future human behavior" (van Dijck 2014, 201). Dataism also presupposes an inherent trust in the institutions that collect, and process data, including the government,

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

academic researchers, or corporate platforms (Mascheroni and Siibak 2021). Hence, we can observe that dataism promotes the mythological aspect of big data identified by boyd and Crawford (2012) since it encourages the perception of data as a means to obtain knowledge that is objective, accurate, and impartial (Mascheroni and Siibak 2021). This notion, nonetheless, has largely been questioned and criticized, as many scholars have identified how algorithms, artificial intelligence (AI), and big data sets can lead to discriminatory and biased outputs (Barassi 2020; Chun 2018; Noble 2018).

The current deeply datafied social scenario has led to the emergence of a new research field, dedicated to the datafication of childhood (Barassi 2020; Livingstone, Stoilova, and Nandagiri 2019; Lupton and Williamson 2017; Mascheroni and Siibak 2021; Montgomery 2015). The contemporary generation of children and adolescents is the first to be datafied since birth and even before that, through pregnancy tracking apps or the sharing of ultrasound scans on social media (Mascheroni 2020; Holloway 2019).

Moreover, children are increasingly subjected to forms of intimate surveillance (Leaver 2017), that is, the use of digital surveillance apparatus by parents to monitor and quantify their children as an expression of a culture of care. In this context, digital surveillance is associated with "good parenting" practices. This cultural discourse is further reinforced by a series of products devoted to surveillance in family contexts such as smart baby wearables that collect and analyze biometric data (e.g. sleep patterns, heart rate), smart toys (e.g. Hello Barbie, Dash and Dot) or parenting apps designed to track a child's development and day-to-day life. As noted by Wilson (2018), these digital tools that quantify children through extensive data collection and analysis promise parents a formula to raise happy and healthy children, a belief that closely relates to the ideology of dataism, in the sense that it proposes "a self-evident relationship between people and data" (van Dijck 2014, 199).

Children are also surveilled in schools, through EdTech tools, such as Google Classroom or Apple Classroom (Livingstone and Pothong 2022). In that sense, today "children are positioned within intense networks of surveillance on the part of parents, healthcare workers, and teachers" (Lupton and Williamson 2017, 782). Nevertheless, it is still unclear how this scenario affects children's present and future (Holloway 2019; Mascheroni and Siibak 2021). Several authors have "raised concerns about the specter of digital 'dossiers' that could follow young people into adulthood, affecting their access to education, employment, healthcare, and financial services" (Montgomery 2015, 268).

As kids grow, they tend to become more autonomous in their use of digital media, but not necessarily less datafied. With time, older children and teenagers start to have more control over their online presence and practices, choosing what to share on social media, which websites and apps to use, and, possibly, employing practices of digital self-tracking (Lupton 2016). However, recent research shows that young people are largely unaware of the big data phenomenon and datafication in general, which could compromise their ability to consciously make choices about their personal data and digital footprint (Dias et al. 2022; Lv, Chen, and Guo 2022; Ponte, Batista, and Baptista

2022). In this respect, a growing consensus arises in terms of the necessity of educating citizens from a young age about big data and associated technologies (e.g. AI), mainly when it comes to their social and cultural consequences (Christozov and Toleva-Stoimenova 2016; Sander 2020a).

Big data literacy (D'Ignazio and Bhargava 2015; François, Monteiro, and Allo 2020) is one of many concepts — such as critical data literacy (Buzato 2017; Hautea, Dasgupta, and Hill 2017; Tygel and Kirsch 2016), critical data education (Pangrazio and Selwyn 2020) or critical algorithmic literacy (Wang et al. 2022) — suggested to refer to this educational process. This emerging field of research gathers little consensus in terms of conceptualization and lacks studies dedicated to the creation of uniform indicators of big data literacy or instruments able to evaluate it (Kubrusly, Batista, and Marôpo, forthcoming). In this article, the concept of big data literacy is broadly defined as the necessary set of skills to identify, understand, and critically interact with big data, as well as being able to deal with its impacts on society and daily life.

In this context, we aim to propose an intervention agenda to improve youth's big data literacy, based on a thematic analysis performed on 20 papers selected by a scoping literature review. With this in mind, we point to relevant gaps and trends in the literature, searching for a larger understanding of young people's current big data literacy and reflecting on future paths — including central themes and pedagogical strategies — that can be used towards enhancing it.

2. Methodology

To develop an agenda for youth's big data literacy we conducted a scoping review of the literature regarding empirical research about youth's thoughts, feelings, and knowledge about (big) data and datafication. This review allowed us to assess the existing literature on those topics in a broad manner, regarding its potential size, scope, and tendencies (Grant and Booth 2009; Munn et al. 2018). Moreover, scoping reviews are particularly adequate when it comes to emerging topics, as they allow us to frame a research topic when it is still unclear what lines of research exist (Yang et al. 2023).

We consulted and tested multiple search terms in three databases — Scopus, Web of Science (All Databases), and ERIC — in February 2023. Generally, searches derived from a combination of youth terms (e.g. adolescent*, child*, student*), education terms (e.g. learn*, school*, educat*), big data terms (e.g. big data, datafication) and literacy terms (e.g. data literac*, critical data literac*, big data literac*). Even though research expressions were written in English, studies in Portuguese or Spanish were also included. Articles, review articles, and book chapters published from 2010 onward found in the databases were imported to Mendeley Reference Manager and included or excluded from the final sample according to uniform inclusion criteria:

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

— The study presents empirical evidence on the big data literacy of children, adolescents, or young adults;
— The study presents empirical evidence on the data literacy of children, adolescents, or young adults, with an emphasis on the digital context;
— The study presents empirical evidence on the artificial intelligence literacy or algorithmic literacy of children, adolescents, or young adults, with a clear connection to big data.

As shown in the inclusion criteria, multiple concepts were considered since, as mentioned before, the concept of big data literacy is not consensual among authors who aim to educate people about big data and their impacts (Kubrusly, Batista, and Marôpo, forthcoming). In the selection process, firstly, collected references were excluded according to the inadequacy of their titles and abstracts following the inclusion criteria. The remaining references were systematically read to assess their suitability. Lastly, we looked for more literature that suited the inclusion criteria in the sample reference lists, which was added to the final sample, at this point including conference proceedings and reports. At last, this process resulted in a selection of 20 final references (Table 1) which were considered in the following analysis.

—
Table 1
Selected references

Reference (author(s), year)	Journal, research project, conference, or book	Type of publication	Location(s)	Methodology
Agesilaou and Kyza (2022)	International Journal of Child-Computer Interaction	Article	Cyprus	QUAL: observation; interviews; educational intervention
Barton et al. (2021)	AERA Open	Article	USA	QUAL: participatory research; surveys; interviews
Bowler et al. (2017)	Proceedings of the Association for Information Science and Technology	Conference proceeding	USA	QUAL: interviews
Chi et al. (2018)	International Conference on Information: Transforming Digital Worlds	Conference proceeding	USA	QUAL: interviews
Hargittai and Marwick (2016)	International Journal of Communication	Article	USA	QUAL: focus groups

Kim et al. (2023)	Education and Information Technologies	Article	USA	QUAL: observations; educational intervention
Kumar et al. (2017)	ACM on Human-Computer Interaction	Conference proceeding	USA	QUAL: interviews
Marín, Carpenter, and Tur (2021)	British Journal of Educational Technology	Article	USA, Germany and Spain	QUAN: survey
Marwick and Hargittai (2019)	Information, Communication & Society	Article	USA	QUAL: focus groups
Pangrazio and Selwyn (2018)	Social Media and Society	Article	Australia	QUAL: participatory workshops
Pangrazio and Selwyn (2020)	Pedagogy, culture and society	Article	Australia	QUAL: participatory workshops
Pronzato and Markham (2023)	Convergence: the international journal of research into new media technologies	Article	Italy	QUAL: autoethnographic diaries; educational intervention
Robertson and Tisdall (2020)	Journal of Media Literacy Education	Article	Scotland	QUAL: consultations
Sander (2020a)	Internet Policy Review	Article	Wales	QUAL: digital content analysis; educational intervention; questionnaire; interviews
Sander (2020b)	Data & Policy	Article	Wales	QUAL: digital content analysis; educational intervention; questionnaire; interviews
Selwyn and Pangrazio (2018)	Big Data and Society	Article	Australia	QUAL: participatory workshops
Stornaiuolo (2020)	Journal of the Learning Sciences	Article	USA	QUAL: social design experiment
Wolff et al. (2018)	The Hackable City: Digital Media and Collaborative City-Making in the Network Society	Book chapter	UK	QUAL: educational intervention; observation
Yates et al. (2020)	Me and My Big Data	Report	UK	QUAN: survey
Zhao (2019)	Proceedings of the Conference on Human Factors in Computing Systems	Conference proceeding	UK	QUAL: focus group

Even though we searched for literature from 2010 onward, all 20 selected references were published between 2016 and 2023, with a peak in 2020, which reinforces the notion of big data literacy as an emerging research topic that is just now being more intensively explored. Only two out of all the references analyzed employed quantitative methods, while the others opted for a variety of qualitative methods, such as interviews (Agesilaou and Kyza 2022; Bowler et al. 2017; Chi et al. 2018), focus groups (Hargittai and Marwick 2016; Zhao 2019), participatory workshops (Selwyn and Pangrazio 2018) and autoethnographic diaries (Pronzato and Markham 2023).

Furthermore, it is possible to notice the prevalence of literature coming from English-speaking countries: 40% of references were studies conducted in the United States of America, 30% in the United Kingdom, and 15% in Australia. In terms of age, none of the references included children younger than five in their sample, and one-quarter of the references included those younger than eleven. The majority of studies — a total of 12 — focused on children and adolescents no older than 18. This is an important aspect to highlight since most of the results explored in this paper come from non-representative samples and, therefore, cannot be generalized.

We conducted a thematic analysis of the studies to identify, analyze, and report patterns (themes) within the studies (Braun and Clarke 2006). This process was divided into two phases. First, we looked at the studies' research questions and objectives, intending to identify trends and gaps in the literature. The idea was that mapping the current research on big data literacy would allow us not only to characterize our sample but also to understand what topics prevail on the scientific agenda and which may be understudied. Secondly, we looked specifically at the results obtained by our sample of references, organizing them into categories drawn from the literature itself to obtain a broader understanding of youth's big data literacy. The results reached were then utilized to support our proposal of an agenda focused on improving youth's big data literacy.

3. Youth and big data literacy: main findings from selected studies

This section is dedicated to the findings reached through the thematic analysis of the sample obtained by the scoping review of the literature.

When it comes to research questions and objectives, it is possible to observe three main tendencies. Most studies tend to explore young people's management of their digital data, including their ability to recognize what counts as data (Bowler et al. 2017; Pangrazio and Selwyn 2018), when and why data about them are collected, and what are the possible consequences attached to the analysis of said data (Marwick and Hargittai 2019; Robertson and Tisdall 2020). Moreover, there is some focus, in this group of studies, on youth's understanding and perceptions of digital data influxes and ecosystems (Agesilaou and Kyza 2022; Yates et al. 2020) and their awareness of the part that digital data play in the economic ecosystem of online platforms (Bowler et al. 2017; Marwick

and Hargittai 2019). Overall, studies centered on personal data management online cover a broad variety of topics, including youth's beliefs, worries, attitudes, feelings, knowledge, and tactics.

Another segment of studies focuses specifically on one aspect of personal data management: online privacy (Hargittai and Marwick 2016; Kumar et al. 2017; Marín, Carpenter, and Tur 2021; Sander 2020a; Zhao 2019). In the sample examined in this paper, studies on digital privacy tended to adopt a more risk-focused approach, evaluating youth's knowledge and beliefs about privacy risks, the value they concede to privacy, how they make decisions about their privacy in digital environments, and what strategies they employ to protect it. Some studies in this category also dedicate themselves to understanding young adults' trust in technology companies or governmental privacy policies (Marín, Carpenter, and Tur 2021) or to the effect of educational tools related to big data on their attitudes and feelings toward their privacy online (Sander 2020a).

The last tendency, and the least popular in the sample considered in this paper, regards youth's skills to employ data — personal or otherwise — to fulfill personal, communitarian, or political interests (Barton et al. 2021; Stornaiuolo 2020).

When it comes to results reached by our selected studies, we were able to draw six different categories from the literature: (3.1) youth's knowledge about data, (3.2) youth's knowledge about data ecosystems, (3.3) youth's knowledge about risks when it comes to data, (3.4) youth's feelings of (dis)comfort regarding several aspects of datafication, (3.5) youth's attitudes and strategies of resistance regarding the less desirable consequences of datafication and (3.6) the impact of educational initiatives. The first three categories are focused on knowledge about data and the big data phenomenon. Categories four and five, on the other hand, reflect youth's feelings and attitudes towards datafication. In turn, the sixth category showcases the successes and shortcomings of educational interventions focused on big data literacy as they were reported by some studies from our sample. Each one of the six categories is detailed below.

3.1. Youth's knowledge about data

Selected studies show that young people associate data with numbers, statistical analysis, facts, hard scientific evidence, and its representations, like Excel sheets, graphs, or percentages (Bowler et al. 2017; Stornauiolo 2020). It is also possible to verify that there is a common confusion among groups of young people — between 11 and 18 years old — who often associate the term "data" in the digital context with the mobile data of their cell phones (Bowler et al. 2017; Pangrazio and Selwyn 2018; Stornaiuolo 2020). Only in Bowler et al. (2017), did part of the young participants express thinking about data in terms of their digital footprints. In this same study, when questioning the participants about the images and metaphors they associated with data, the authors found that the most common images are linked to archives. Parallel to that, some students revealed that they associate data with pop culture references, such as spy movies

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

and the "green things raining down" in The Matrix movie. In general, studies tend to demonstrate that a significant part of young people struggle to define what data are (Pangrazio and Selwyn 2018).

Although it is not clear that young people can understand and define what counts as personal data, the literature shows that they can identify different kinds of data and show distinct attitudes toward each of them. For example, several studies show that young people have negative feelings about the tracking of their geolocation data, considering this practice scary and invasive (Agesilaou and Kyza 2022; Pangrazio and Selwyn 2018; Selwyn and Pangrazio 2018). Young adults, on the other hand, reveal different attitudes toward the protection of their privacy depending on the content of the data involved (Marwick and Hargittai 2019). For example, health information or financial status were considered sensitive, while online shopping patterns, photos, and social media likes were considered more mundane and, hence, more suited for tracking. Teenagers in Pangrazio and Selwyn (2018) expressed similar views, seeming to be unbothered by the tracking and analysis of images (e.g. selfies) or text (e.g. text messages).

Research also shows that children and adolescents can hold some naive conceptions about data — which express the mythological dimension of big data (boyd and Crawford 2012) and dataism (Mascheroni and Siibak 2021; van Dijck 2014) — when it comes to trusting data as neutral and objective sources of knowledge. Kim et al. (2023) recognized a knowledge gap among students aged 11 to 14 regarding the relationship between data and artificial intelligence. Firstly, "some of the students consider that artificial intelligence is already born to be intelligent, accurate and efficient, and therefore does not need to be trained with data" (Kim et al. 2023, 12). On the other hand, participants seemed to devalue the quality of input data in the process of building artificial intelligence programs, valuing above all the amount of data, which demonstrates uncritical conceptions regarding this topic. However, when working with a group of older individuals — between 12 and 24 — Barton et al. (2021) verified an opposite scenario where many young people demonstrated an understanding that data are not neutral units of information, being often guided by interests and expressions of power.

3.2. Youth's knowledge about data ecosystems

It is possible to observe that even when youth are aware of the collection of their data in a digital environment, many do not understand how such collection can affect their lives (Robertson and Tisdall 2020; Sander 2020a). In this sense, there seems to be a significant gap in knowledge regarding the economic ecosystem of the digital environment and how datafication is related to this scenario. The gap that appears more consistently in the literature refers to young people's understanding of data flows and infrastructures in terms of when and by whom data are collected, stored, exchanged, and analyzed. Many young people are unaware that their data can be sold, traded, and

shared between different companies and entities (Bowler et al. 2017; Robertson and Tisdall 2020) or that through a single piece of data, such as an email address, a company could have access to an extensive profile of accumulated data, developed by third parties (Marwick and Hargittai 2019).

There also seems to be little clarity on the part of young populations about their own digital footprints and ownership of their personal data. In a study conducted by Agesilaou and Kyza (2022), with 63 fifth graders, only 9% of them recognized that ownership of their digital data is shared with technology companies through which such data is produced and circulated. In line with this perception, Bowler et al. (2017) found that young people are largely unaware of the extent and durability of their digital footprints, sometimes having the misleading perception that when deleting an account or profile on a digital platform, the data associated with it would also be deleted. Regarding notions about the production and collection of data, there are some controversial results. On the one hand, in Wolff et al. (2018), young people demonstrated difficulties in understanding how data are collected in smart city models and how different kinds of data might be aggregated and used to improve the quality of life. In Pangrazio and Selwyn (2018), some adolescents linked the collection of geolocation data exclusively to Google Maps, an incorrect perception that reveals gaps in knowledge regarding the actors involved in data collection. On the other hand, in Robertson & Tisdall (2020), participants showed awareness of different contexts in which data were collected, such as school or their own homes. Similarly, in Bowler et al. (2017) adolescents were able to associate data creation with different digital environments and identify what types of data were collected and by whom.

Even so, the literature indicates that young people do not associate such practices with an economic business model. For example, even though fifth graders can identify that their data can be seen and used by online gaming companies, they relate such use with the need to verify the proper functioning of the application and the repression of inappropriate content and behavior (Agesilaou and Kyza 2022). In Zhao (2019), few children recognized content personalization and autoplay as platform monetization strategies, on the contrary, most saw such features as indicators of the platform's proper functioning. It is noted, in this sense, that the commercial value of data in the digital medium is either not recognized or simply considered to be unproblematic (Bowler et al. 2017; Marwick and Hargittai 2019). This finding aligns with previous research dedicated to the topic of digital privacy which shows that "commercial privacy is the area that children are least able to comprehend and manage" (Livingstone, Stoilova, and Nandagiri 2019, 4).

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

3.3. Youth's knowledge about risks

In terms of risks linked to personal data and digital footprints, the literature shows that children under 11, especially the younger ones — between 5 and 7 years old — have significant gaps in terms of their knowledge of privacy and digital security (Kumar et al. 2017; Zhao 2019). Among adolescents and young adults, risks are conceptualized mainly in the individual sphere, such as possible conflicts with family members, damage to their online reputation, cyberbullying, or feelings of shame (Agesilaou and Kyza 2022; Hargittai and Marwick 2016; Robertson and Tisdall 2020). Another curious trend is the often-made association of privacy risks with "hackers", "malicious people" or "pedophiles" (Agesilaou and Kyza 2022; Pangrazio and Selwyn 2018; Selwyn and Pangrazio 2018). Nonetheless, individual risks related to the potential impacts of predictive analysis in future opportunities at the professional, educational, and economic levels (Barassi 2020) were not mentioned by participants in selected studies.

These results can be considered particularly worrying, given that they indicate that young people might be unaware of consequences related to the datafication of their experiences beyond their private lives. For instance, discriminatory practices perpetuated by algorithms and AI, which are often a result of the lack of diversity in the data used to train these technologies, may be unknown (Holm and Lippert-Rasmussen 2023; Noble 2018; Perez 2019). Examples of such biased practices can be observed in healthcare access (Obermeyer et al. 2019) and within the judicial system (Angwin et al. 2016). In this sense, big data collective risks seem to be neglected by youth when it comes to its impact on our democracies and public policies.

3.4. Youth's feelings of (dis)comfort regarding datafication

Participants in selected studies manifest some discomfort concerning the commercialization of data among technology companies (Robertson and Tisdall 2020). According to a study conducted with UK citizens, 90% of young participants classified as social and media users between 16 and 24 years of age say it is not acceptable for "companies to sell their personal data to other companies or to influence their opinions using such data" (Yates et al. 2020, 32). However, 38% consider it acceptable for companies to personalize advertisements and use their data to shape the content presented to them. Following this trend, investigations conducted in Australia reveal that adolescents consider the use of their data for advertising, commercial, or content customization to be normal and acceptable (Pangrazio and Selwyn 2018). These results reinforce the previously explored perception that youth are not aware of the monetary business model that is built around their data.

In terms of views on technology companies, the results are controversial. In some contexts, young people reveal that they trust Amazon and Google, including when it

comes to sensitive information such as their home address or credit card number (Marwick and Hargittai 2019). On the other hand, Australian teenagers report distrusting Facebook and Snapchat (Pangrazio and Selwyn 2018), while only 8% of 148 pre-service teachers surveyed in Germany, Spain, and the United States are comfortable with how social media companies use their data. This number drops to 6% when talking about the use of their students' data (Marín, Carpenter, and Tur 2021). Despite expressing these fears regarding their online privacy, young people still identify a series of benefits linked to social networks, such as building and maintaining affective ties (Agesilaou and Kyza 2022; Pangrazio and Selwyn 2018; Sander 2020a), convenience (Marwick and Hargittai 2019; Sander 2020a; Pronzato and Markham 2023), access to personalized content (Marwick and Hargittai 2019), health benefits (Barton et al. 2021; Marwick and Hargittai 2019), and access to information (Barton et al. 2021).

3.5. Youth's attitudes and strategies of resistance

Another notable trend towards attitudes and perceptions about datafication in the digital environment is a shared feeling of apathy and indifference (Chi et al. 2018; Selwyn and Pangrazio 2018). Even when they are aware of datafication's potential damaging impacts, youth consider this scenario to be inevitable, and, therefore, impossible to change or resist (Hargittai and Marwick 2016; Zhao 2019). Thus, young people have low expectations regarding their privacy in the digital environment (Bowler et al. 2017) and sometimes trivialize it, considering that privacy would only be relevant if they "had something to hide" (Marwick and Hargittai 2019, 10). Furthermore, many studies point to a feeling of disempowerment when it comes to structural aspects of the digital sphere. In Pronzato and Markham (2023), for example, university students stated that they felt "addicted" to social networks and that, although dissatisfied with their use of digital media, they considered it impossible to give up.

This perception of inability to resist social networks is relatively transversal in the literature (Hargittai and Marwick 2016; Marwick and Hargittai 2019; Selwyn and Pangrazio 2018). It is also noted that young people seem to have difficulties in idealizing or thinking about resistance strategies and practices in the face of aspects of datafication that they consider negative, limiting themselves to more radical ideas such as the complete rejection of social networks (Pangrazio and Selwyn 2018). Still, the literature gives indications of some of the most popular strategies adopted by young people to protect their privacy, such as reducing activity on social networks (Agesilaou and Kyza 2022), asking adults for help (Agesilaou and Kyza 2022; Kumar et al. 2017), the use of false personal information (e.g. name, date of birth) (Agesilaou and Kyza 2022; Chi et al. 2018; Kumar et al. 2017), and not sharing too much intimate content, such as emotional states or political beliefs (Hargittai and Marwick 2016).

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

In addition to that, some studies were able to identify a few more complex tactics. Agesilaou and Kyza (2022) recognize four criteria considered by children between 11 and 12 years old in their decision-making processes regarding the sharing of personal data: the nature of the data being shared, how the data will be used, who will use the data being shared, and for what purposes will the data be used. Beyond these strategies, Barton et al. (2021) assessed the existence of more critical attitudes among young people in the United States during the COVID-19 pandemic. This group recognized the non-neutrality of large data sets and, therefore, sought to inform themselves using not only official statistical data but rather combining them with personal and contextualized narratives they found on TikTok and YouTube. In this sense, throughout the pandemic, young people built their data networks, tailored to their experiences, particular interests, and needs of their communities, combining various forms and sources of data.

3.6. Impact of educational initiatives

Several of the studies found through our scoping review reflected upon educational initiatives and their impacts on young participants. Even though these pedagogical experiments and workshops were able to enhance youth's knowledge about datafication and big data's impact in general, they were not always successful in overcoming feelings of apathy and impotence regarding these same phenomena (Pangrazio and Selwyn 2018; Pronzato and Markham 2023; Sander 2020b).

An effective strategy to contextualize big data's impacts is materializing data through the collection of personal data (Stornaiuolo 2020), in diverse hypothetical scenarios (Agesilaou & Kyza 2022), or through digital tools that showcase online data collection by various parties (Pangrazio and Selwyn 2020; Selwyn and Pangrazio 2018). This seems to be a fundamental step towards building engagement and awareness about data, making it clear that, despite its volume and complexity, big data are composed of multiple "small" data connected to people's behavior and beliefs (Bhargava et al. 2015).

Several authors in our sample also found that working with students' personal data was a useful strategy for developing big data literacy (Agesilaou and Kyza 2022; Pangrazio and Selwyn 2020; Sander 2020a; Stornaiuolo 2020). Other pedagogical strategies deemed efficient were autoethnographic diaries (Pronzato and Markham 2023), quizzes about privacy and digital practices (Pangrazio and Selwyn 2020), and interactive multimedia tools (Sander 2020a). Overall, critical pedagogical strategies were positively related to the development of big data literacy as an "attempt to move students away from negative consequences, risk management and individualized responsibility towards more collective forms of analytical thinking and critical reflection" (Pangrazio and Selwyn 2020, 444).

4. An agenda for big data literacy

Given the results found in the literature, we propose an agenda for future developments of big data literacy. Here, we choose to highlight four transversal topics within our first five categories, which reflect youth's knowledge, feelings, and attitudes related to the big data phenomenon and its impacts. The sixth category, focused on educational initiatives, helped us to consider shortcomings and positive results obtained by existing pedagogical strategies, which are included in our proposal.

The first topic that we emphasize as relevant across several categories is the defiance of big data's mythology and data-driven technologies as neutral and impartial tools. This appears to be fitting when it comes to youth's knowledge about data, as studies show their limited conception of what counts as data, but also regarding their lack of knowledge on the societal risks of employing biased data and data-based technologies.

The second topic explored in our agenda aims to situate the big data phenomenon within the digital economic ecosystem. This theme is showcased both in the category regarding youth's knowledge of risks, which demonstrates their disregard towards commercial privacy issues, and in the category linked to knowledge about data ecosystems, as results reveal youth seldom recognize data's economic value. Moreover, this matter is also reflected in youth's feelings of discomfort when it comes to commercial practices that include selling their personal data.

Our third topic — connecting big data to AI's impact on society — was drawn from young people's lack of knowledge about risks associated with the employment of AI in diverse societal sectors. This limited view of big data's impacts is also reflected by youth's perceptions, attitudes, and strategies on privacy, as risks and opportunities in this sphere are only conceptualized at an individual level, with a significant lack of awareness about potential collective and societal impacts.

The fourth point in our agenda speaks on potential strategies to overcome youth's stances of apathy and indifference, which were mostly discussed on results about young people's attitudes and strategies regarding the impacts of datafication processes. This topic also comes up as a potential shortcoming of existing educational initiatives discussed in category six, as many of them were able to enhance knowledge, but did not impact youth's attitudes. All four topics are detailed below.

4.1. Defying big data's mythology and data-driven technologies as neutral and impartial tools

As attested in several studies of our sample, mythological and naive perceptions about data are common among young people, who often associate them with hard facts and numbers (Bowler et al. 2017; Stornaiuolo 2020). Consequently, technologies that depend on big data — such as AI — are also understood as neutral and impartial tools

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

(Kim et al. 2023). This scenario seems to be associated with big data's mythological dimension (boyd & Crawford 2012) and the ideology of dataism (van Dijck 2014). Therefore, it is necessary to deconstruct these paradigms to teach a more critical view of the datafication of society.

Previous research on critical data literacy (Claes and Philippette 2020; Hautea, Dasgupta, and Hill 2017) sustains the need to engage with big data as a cultural narrative (Carrington 2018). This line of research highlights as central skills the ability to make critical judgments underlying data-driven innovation (Claes and Philippette 2020), understanding that data collection and analysis are processes that occur in contexts of power (Philip, Schuler-Brown, and Way 2013; Spiranec, Kos, and George 2019), and that data are always shaped by decisions and assumptions that are not always visible (Claes and Philippette 2020; Hautea, Dasgupta, and Hill 2017). In that sense, critical literacies and pedagogies seem to be able to offer relevant contributions to big data literacy (D'Ignazio 2017; D'Ignazio and Bhargava 2015; Pangrazio and Selwyn 2018; Philip, Schuler-Brown, and Way 2013; Sander 2020a; Spiranec, Kos, and George 2019). Moreover, this approach is useful when it comes to aligning big data projects — which are often used in social good-related sectors, such as education — with social good values (D'Ignazio and Bhargava 2015).

To work towards empowering youth to defy big data's mythology, authors point to the adoption of a Freirian approach (Carrington 2018; D'Ignazio 2017; D'Ignazio and Bhargava 2015; Spiranec, Kos, and George 2019), grounding the educational process in the reality and authentic experiences of students. For instance, Stornaiuolo (2020) suggests positioning youth as authors and architects of data, that is, inviting them to take an active role in collecting and analyzing data to tell and interpret stories about their lives. This exercise is also relevant to expand what "counts" as data, "to include personal narratives, art, and everyday activities" (Stornaiuolo 2020, 83). In this context, we recognize that educational activities where youth actively collect, organize, and analyze data can be useful in raising awareness about decisions and assumptions that influence data collection. To this same end, reflexive activities about data as social, political, and cultural objects can help youth reflect on where data comes from and what it represents.

4.2. Situating big data within the digital economic ecosystem

Even though the idea that technology should be employed in favor of social good and humane values is present in several European Union (EU) official documents, such as the report on Civil Law Rules on Robotics (Cath et al. 2017), technological innovation is still overwhelmingly pushed by commercial and private interests (van Dijck 2020). Nonetheless, youth seldom recognize data as potential sources of profit for corporations, digital platforms, and data brokers and struggle particularly when it comes to understanding their commercial privacy, that is, how their personal data are "harvested and used

for business and marketing purposes" (Livingstone, Stoilova, and Nandagiri 2019, 3).

Thus, it is vital to work towards an understanding that data play a double part within the digital economic ecosystem. On the one hand, as a commodity that can be sold to interested third parties; on the other hand, as a way to refine and personalize strategies that "capture" the user's attention, incentivizing constant connection (Bhargava & Velasquez 2021). This second aspect seems to be particularly important to young people, as studies show they often struggle to manage the time they spend online (Jiang 2018; Ponte and Batista 2019).

The lack of understanding about their digital footprint — both in terms of its management and its potential consequences — might also contribute to the limited conception young people show regarding their online privacy. In this sense, it may be useful to contextualize (big) data within youth's day-to-day use of social media, reflecting on what data is collected, how it might be analyzed, by whom, and for what. Introducing social media's business models and discussing practical examples of their strategies to potentialize profit (e.g. autoplay, infinite scrolling, cookies) and what part their personal data plays within this system also seems to be fundamental. Once again, exercises where they reflect on their personal experiences and data could be a way of raising awareness of this topic engagingly and critically.

4.3. Connecting big data to AI's impact on society

As previously mentioned, the perception of data as neutral entities leads youth to believe that data-based technologies, such as AI systems, are also neutral, objective, and truthful (Kim et al. 2023). This belief is particularly worrying as the use of AI in diverse social sectors (e.g. education, healthcare, judicial system) becomes more widespread. Yet, working on defying big data's mythology might be a valuable strategy when it comes to incentivizing a more critical view of AI and its application, raising awareness as to how these technologies can be discriminatory and biased.

Furthermore, it seems to be fundamental that youth understand not only AI's potential individual risks but also risks on a collective level, which seems to be less known. In terms of pedagogical strategies, it could be useful to approximate students from AI, exploring how it is used in their daily lives, on social media, streaming platforms, and other digital spaces. Moreover, science fiction movies, books, and TV shows or documentaries could be used as a starting point to discuss social beliefs about AI and how they relate to current uses of these technologies. Practical examples of AI's applications on a societal level may also be useful to discuss collective risks. Generally, creating a clear connection between big data and AI seems to be a useful strategy for building a critical view on both these issues and improving youth's big data literacy.

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

4.4. Overcoming feelings of apathy

According to Bawden (2008), digital literacy cannot be simply understood as a collection of skills and knowledge, as attitudes and perspectives are also a part of the equation. In this sense, the feelings of apathy towards datafication, that seem to be frequent among youth, are particularly worrying. Moreover, research shows that youth often feel powerless in the face of digital structures (Creswick et al. 2019; Vallejos et al. 2021). These findings align with the popularity of attitudes of digital resignation (Draper and Turow 2019), that is, a pattern of inaction that persists in the face of digital privacy issues, even when privacy is valued by users. In this scenario, while people "feel dissatisfied with the pervasive monitoring that characterizes contemporary digital spaces, they are convinced that such surveillance is inescapable" (Draper and Turow 2019, 2).

With this in mind, we suggest that empowering youth is a fundamental step to improving their big data literacy. To challenge feelings of apathy, the use of hypothetical scenarios where students are invited to take on an active role (e.g. creating a new social media platform, suggesting laws regarding digital privacy, advising the government on the use of AI) could be effective. This premise aligns with previously discussed critical literacy approaches, that focus on empowerment through literacy education (D'Ignazio and Bhargava 2015). In addition, focusing on collective action might also be beneficial, particularly if grounded on already existing local and global initiatives that aim to protect digital rights and promote data justice.

Furthermore, we believe it is always important to balance risks and opportunities when speaking about big data's societal impacts, as too much emphasis on the risks might reinforce feelings of inescapability and attitudes of indifference. Thus, big data literacy initiatives should aim to build "power, not paranoia" (Lewis et al. 2018), steering youth towards active stances.

5. Final considerations

As childhood becomes increasingly datafied (Mascheroni and Siibak 2021), scholars have raised concerns about how this process may impact young people's futures (Montgomery 2015). Nonetheless, youth seem to be largely unaware of the big data phenomenon, which leads researchers and educators to highlight the importance of teaching children about this topic and its social and cultural consequences (Christozov and Toleva-Stoimenova 2016; Sander 2020a). In this respect, research about big data literacy is on the rise, as the scientific community struggles to find methodological and conceptual consensuses (Kubrusly, Batista, and Marôpo, forthcoming). With this in mind, we have conducted a scoping review of the literature, aiming to reach a coherent intervention agenda based on the available empirical evidence, which can be used to inform policy and practice aimed at promoting youth's big data literacy. Through a thematic

analysis of 20 studies, we point to relevant gaps and trends in the literature, searching for a broader picture of young people's current big data literacy.

In this paper, we have reflected on future paths towards the enhancement of youth's big data literacy arriving at an intervention agenda made up of four topics: 1. Defying big data's mythology and data-driven technologies as neutral and impartial tools; 2. Situating big data within the digital economic ecosystem; 3. Connecting big data to AI's impact on society, and 4. Overcoming feelings of apathy. These topics can inform pedagogical resources and activities to be used in formal and informal educational contexts. We conclude that, in a digital world in constant transformation, it is important that efforts to improve youth's digital literacy are not limited to technical and operational skills. On the contrary, there should be an emphasis on developing an informed and critical use of digital technologies and on children's rights in the online context. In that sense, it seems to be beneficial to adopt participatory approaches, involving youth in the process of thinking and implementing activities to promote digital literacy.

Nonetheless, it is important to say that our aim is not restricted to stressing individual responsibilities. Even though digital literacies seem to be a key part of empowering youth online, it is also fundamental to highlight the need for regulation and accountability when it comes to digital platforms and companies. In this regard, we believe initiatives such as the Child Rights by Design principles, created "to inspire innovators to help realize children's rights when designing digital products and services" (Livingstone and Pothong 2023, 7) are needed to involve other actors in the issues of datafication and digitalization of our societies.

RCL — Revista de Comunicação e Linguagens *Journal of Communication and Languages* N.60-61 (2024) ISSN 2183-7198

—

Funding

This work was partially supported by an individual PhD grant by the Portuguese Foundation for Science and Technology (FCT), reference PRT/BD/154767/2023.

—

References

Agesilaou, Andria, and Eleni A. Kyza. 2022. "Whose Data Are They? Elementary School Students' Conceptualization of Data Ownership and Privacy of Personal Digital Data." *International Journal of Child-Computer Interaction* 33: 100462–480. https://doi.org/10.1016/j.ijcci.2022.100462.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. *Pro Publica: Machine Bias*. May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Barassi, Veronica. 2020. "Datafied Times: Surveillance Capitalism, Data Technologies and the Social Construction of Time in Family Life." *New Media & Society* 22 (9): 1545–60. https://doi.org/10.1177/1461444820913573.

Barton, Angela Calabrese, Day Greenberg, Chandler Turner, Devon Riter, Melissa Perez, Tammy Tasker, Denise Jones, Leslie Rupert Herrenkohl, and Elizabeth A. Davis. 2021. "Youth Critical Data Practices in the COVID-19 Multipandemic." *AERA Open* 7: 1–16. https://doi.org/10.1177/23328584211041631.

Bawden, David. 2008. "Origins and Concepts of Digital Literacy." In *Digital Literacies: Concepts, Policies, And Practices*, edited by Colin Lankshear and Michele Knobel, 17-32. New York: Peter Lang.

Bhargava, Vikram R., and Manuel Velasquez. 2020. "Ethics of the Attention Economy: The Problem of Social Media Addiction." *Business Ethics Quarterly* 31 (3): 321–59. https://doi.org/10.1017/beq.2020.32.

Bhargava, Rahuk, Erica Deahl, Emmanuel Letouzé, Amanda Noonan, David Sangokoya, Natalie Shoup, Mark Frohardt, Willam Hoffman, Alex Sandy Pentland, and Alessia Lefébure. 2015. *Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data*. Data Pop Alliance, Internews, MIT Center for Civic Media. https://datapopalliance.org/wp-content/uploads/2015/10/BeyondDataLiteracy_DataPopAlliance_Sept30.pdf.

Bowler, Leanne, Amelia Acker, Wei Jeng, and Yu Chi. 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.

Boyd, Danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15 (5): 662–79. https://doi.org/10.1080/1369118x.2012.678878.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Buzato, E. K. Marcelo. 2017. "Critical Data Literacies: going beyond words to challenge the illusion of a literal world." In *Multimodal meaning-making and authorship in post-graduate language and literacy studies* edited by Walkyria Monte Mór and Nara Hiroko Takaki, 119-42. Campinas: Pontes Editores.

Carrington, Victoria. 2018. "The Changing Landscape of Literacies: Big Data and Algorithms." *Digital Culture & Education* 10 (1): 67-76. https://ueaeprints.uea.ac.uk/id/eprint/68380/1/Carrington.pdf.

Cath, Corinne, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo, and Luciano Floridi. 2017. "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach." *Science and Engineering Ethics* 24: 505-28. https://doi.org/10.1007/s11948-017-9901-7.

Chi, Yu, Wei Jeng, Amelia Acker, and Leanne Bowler. 2018. "Affective, Behavioral, and Cognitive Aspects of Teen Perspectives on Personal Data in Social Media: A Model of Youth Data Literacy." In *Lecture Notes in Computer Science* edited by Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willet, 442–52. Sheffield: Springer Cham.

Christozov, Dimitar, and Stefka Toleva-Stoimenova. 2016. "Big Data Literacy: A New Dimension of Digital Divide, Barriers in Learning via Exploring "Big Data"." In *Strategic Data-Based Wisdom in the Big Data Era*, edited by John Girard, Deanna Klein, and Kristi Berg, 156-171. Hershey: IGI Global.

Chun, Wendy Hui Kyong. 2018. "Queerying Homophily." In *Pattern Discrimination*, edited by Clemens Apprich, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl, 59-97. Lüneburg: meson press.

Claes, Arnaud, and Thibault Philippette. 2020. "Defining a Critical Data Literacy for Recommender Systems: A Media-Grounded Approach." *The Journal of Media Literacy Education,* 12 (3): 17–29. https://doi.org/10.23860/jmle-2020-12-3-3.

Creswick, Helen, Liz Dowthwaite, Ansgar Koene, Elvira Pérez Vallejos, Virginia Portillo, Monica Cano, and Christopher Woodard. 2019. "'... They Don't Really Listen to People.'" *Journal of Information, Communication and Ethics in Society* 17 (2): 167–82. https://doi.org/10.1108/jices-11-2018-0090.

Daniel, Ben Kei. 2019. "Big Data and Data Science: A Critical Review of Issues for Educational Research." *British Journal of Educational Technology* 50 (1): 101–13. https://doi.org/10.1111/bjet.12595.

Dias, Patrícia, Lídia Marôpo, Catarina Delgado, Maria Do Rosário Rodrigues, João Paulo Machado Torres, and Eduarda Ferreira. 2022. "I'm Not Sure How They Make Money': How Tweens and Teenagers Perceive the Business of social media, Influencers and Brands." *Media Education Research Journal* 11 (1): 1-25. https://doi.org/10.5281/zenodo.6702887.

D'Ignazio, Catherine. 2017. "Creative Data Literacy." *Information Design Journal* 23 (1): 6-18. https://doi.org/10.1075/idj.23.1.03dig.

D'Ignazio, Catherine, and Rahul Bhargava. 2015. "Approaches to Building Big Data Literacy". In *Bloomberg Data for Good Exchange, New York, 2015.* MIT Media Lab: https://www.media.mit.edu/publications/approaches-to-building-big-data-literacy/.

Draper, Nora, and Joseph Turow. 2019. "The Corporate Cultivation of Digital Resignation." *New Media & Society* 21 (8): 1824–39. https://doi.org/10.1177/1461444819833331.

François, Karen, Carlos Monteiro and Patrick Allo. 2020. "Big-data Literacy as a New Vocation for Statistical Literacy." *Statistics Education Research Journal* 19 (1): 194-205.

Grant, Maria J., and Andrew Booth. 2009. "A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies." *Health Information and Libraries Journal* 26 (2): 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.

Hargittai, Eszter, and Alice Marwick. 2016. "What Can I Really Do? Explaining the Privacy Paradox with Online Apathy." *International Journal of Communication* 10: 3737-57. https://ijoc.org/index.php/ijoc/article/view/4655.

Hautea, Samantha, Sayamindu Dasgupta, and Benjamin Mako Hill. 2017. "Youth Perspectives on Critical Data Literacies." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, *Denver, 2017.* https://doi.org/10.1145/3025453.3025823.

Holloway, Donell. 2019. "Surveillance Capitalism and Children's Data: The Internet of Toys and Things for Children." *Media International Australia* 170 (1): 27–36. https://doi.org/10.1177/1329878x19828205.

Holm, Sune, and Kasper Lippert-Rasmussen. 2023. "Discrimination, Fairness, and the Use of Algorithms." *Res Publica* 29 (2): 177–83. https://doi.org/10.1007/s11158-023-09587-2.

Jiang, Jingjign. 2018. *How Teens and Parents Navigate Screen Time and Device Distractions.* Pew Research Center. https://www.pewresearch.org/internet/2018/08/22/how-teens-and-parents-navigate-screen-time-and-device-distractions/.

Kim, Keun-Jae, Kyungbin Kwon, Anne Ottenbreit-Leftwich, Haesol Bae, and Krista Glazewski. 2023. "Exploring Middle School Students' Common Naive Conceptions of Artificial Intelligence Concepts and the Evolution of These Ideas." *Education and Information Technologies* 28 (8): 9827–54. https://doi.org/10.1007/s10639-023-11600-3.

Kubrusly, Ana, Susana Batista, and Lidia Marôpo. Forthcoming. "Literacia do *big data*: uma revisão integrativa da literatura e diretrizes para futuras investigações". *Livro de Atas do VI Congresso Literacia, Media e Cidadania*.

Kumar, Priya, Shalmali Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara Clegg, and Jessica Vitak. 2017. "'No Telling Passcodes Out Because They're Private.'" *Proceedings of the ACM on Human-Computer Interaction* 1: 1–21. https://doi.org/10.1145/3134699.

Leaver, Tama. 2017. "Intimate Surveillance: Normalizing Parental Monitoring and Mediation of Infants Online." *Social Media + Society* 3 (2): 1-10. https://doi.org/10.1177/2056305117707192.

Lewis, T., S. P. Gangadhara, M. Saba, and T. Petty. 2018. *Digital defense playbook: Community power tools for reclaiming data*. Detroit: Our Data Bodies. https://www.odbproject.org/wp- content/uploads/2019/03/ODB_DDP_HighRes_Single.pdf.

Livingstone, Sonia, Mariya Stoilova, and Rishita Nandagiri. 2019. *Children's data and privacy online: Growing up in a digital age.* London: London School of Economics and Political Science. https://eprints.lse.ac.uk/101283/1/Livingstone_childrens_data_and_privacy_online_evidence_review_published.pdf.

Livingstone, Sonia, and Kruakae Pothong, eds. 2022. *Education Data Futures: Critical, Regulatory and Practical Reflections*. 5Righsts Foundation.

Livingstone, Sonia, and Kruakae Pothong. 2023. *Child Rights by Design: Guidance for Innovators of Digital Products and Services Used by Children*. 5Rights Foundation. https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf.

Lupton, Deborah. 2016. "You are your data: Self-tracking practices and concepts of data." In *Lifelogging*, edited by Stefan Selke, 61-79. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-13137-1_4.

Lupton, Deborah, and Ben Williamson. 2017. "The Datafied Child: The Dataveillance of Children and Implications for Their Rights." *New Media & Society* 19 (5): 780–94. https://doi.org/10.1177/1461444816686328.

Lv, Xing, Yang Chen, and Weiqi Guo. 2022. "Adolescents' Algorithmic Resistance to Short Video APP's Recommendation: The Dual Mediating Role of Resistance Willingness and Resistance Intention." *Frontiers in Psychology* 13: 1-17. https://doi.org/10.3389/fpsyg.2022.859597.

Marín, Victoria I., Jeffrey P. Carpenter, and Gemma Tur. 2021. "Pre-service Teachers' Perceptions of Social Media Data Privacy Policies." *British Journal of Educational Technology* 52 (2): 519–35. https://doi.org/10.1111/bjet.13035.

Marwick, Alice E., and Eszter Hargittai. 2019. "Nothing to Hide, Nothing to Lose? Incentives and Disincentives to Sharing Information with Institutions Online." *Communication & Society* 22 (12): 1697–1713. https://doi.org/10.1080/1369118x.2018.1450432.

Mascheroni, Giovanna. 2020. "Datafied Childhoods: Contextualising Datafication in Everyday Life." *Current Sociology* 68 (6): 798–813. https://doi.org/10.1177/0011392118807534.

Mascheroni, Giovanna, and Andra Siibak. 2021. *Datafied Childhoods: Data Practices and Imaginaries in Children's Lives.* New York: Peter Lang.

Montgomery, Kathryn. 2015. "Children's Media Culture in a Big Data World." *Journal of Children and Media* 9 (2): 266–71. https://doi.org/10.1080/17482798.2015.1021197.

Munn, Zachary, Micah D J Peters, Cindy Stern, Cătălin Tufănaru, Alexa McArthur, and Edoardo Aromataris. 2018. "Systematic Review or Scoping Review? Guidance for Authors When Choosing between a Systematic or Scoping Review Approach." *BMC Medical Research Methodology* 18 (1). https://doi.org/10.1186/s12874-018-0611-x.

Noble, Safiya, U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–53. https://doi.org/10.1126/science.aax2342.

Pangrazio, Luci, and Neil Selwyn. 2018. "'It's Not Like It's Life or Death or Whatever': Young People's Understandings of Social Media Data." *Social Media and Society* 4 (3): 1-9. https://doi.org/10.1177/2056305118787808.

Pangrazio, Luci, and Neil Selwyn. 2020. "Towards a School-Based 'Critical Data Education." *Pedagogy, Culture and Society* 29 (3): 431–48. https://doi.org/10.1080/14681366.2020.1747527.

Perez, Criado, Carolina. 2019. *Invisible Women: Data Bias in a World Designed for Men*. London: Vintage Publishing.

Philip, Thomas M., Sarah Schuler-Brown, and Winmar Way. 2013. "A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles." *Technology, Knowledge, and Learning* 18 (3): 103–20. https://doi.org/10.1007/s10758-013-9202-4.

Ponte, Cristina, Susana Batista, and Rita Baptista. 2022. *Resultados da 1ª série do questionário ySKILLS (2021) Portugal*. Leuven: ySKILLS. https://zenodo.org/records/6376327.

Ponte, Cristina, and Susana Batista. 2019. *EU Kids Online Portugal: Usos Competências, Riscos e Mediações da Internet Reportados por Crianças e Jovens (9-17 anos)*. EU Kids Online. https://fabricadesites.fcsh.unl.pt/eukidsonline/wp-content/uploads/sites/36/2019/03/RELATÓRIO-FINAL-EU-KIDS-ONLINE.docx.pdf.

Pronzato, Riccardo, and Annette N. Markham. 2023. "Returning to Critical Pedagogy in a World of Datafication." *Convergence* 29 (1): 97–115. https://doi.org/10.1177/13548565221148108.

Robertson, Judy, and E. Kay M. Tisdall. 2020. "The Importance of Consulting Children and Young People about Data Literacy." *The Journal of Media Literacy Education* 12 (3): 58–74. https://doi.org/10.23860/jmle-2020-12-3-6.

Sander, Ina. 2020a. "What is critical big data literacy and how can it be implemented?" *Internet Policy Review* 9 (2): 1–22. https://doi.org/10.14763/2020.2.1479

Sander, Ina. 2020b. "Critical Big Data Literacy Tools—Engaging Citizens and Promoting Empowered Internet Usage." *Data & Policy* 2: 1-10. https://doi.org/10.1017/dap.2020.5.

Selwyn, Neil, and Luci Pangrazio. 2018. "Doing Data Differently? Developing Personal Data Tactics and Strategies amongst Young Mobile Media Users." *Big Data and Society* 5 (1): 1-12. https://doi.org/10.1177/2053951718765021.

Spiranec, Sonja, Denis Kos, and Michael George. 2019. "Searching for Critical Dimensions in Data Literacy." In *Proceedings of the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, June 2019*. https://informationr.net/ir/24-4/colis/colis1922.html.

Stornaiuolo, Amy. 2019. "Authoring Data Stories in a Media Makerspace: Adolescents Developing Critical Data Literacies." *Journal of the Learning Sciences* 29 (1): 81–103. https://doi.org/10.1080/10508406.2019.1689365.

Tygel, Alan Freihof, and Rosana Kirsch. 2016. "Contributions of Paulo Freire for a Critical Data Literacy: A Popular Education Approach." *The Journal of Community Informatics* 12 (3): 108-21. https://doi.org/10.15353/joci.v12i3.3279.

Vallejos, Elvira Pérez, Liz Dowthwaite, Helen Creswich, Virginia Portillo, Ansgar Koene, Marina Jirotka, Amy McCarthy, and Derek McAuley. 2021. "The Impact of Algorithmic Decision-Making Processes on Young People's Well-Being." *Health Informatics Journal* 27 (1). https://doi.org/10.1177/1460458220972750.

van Dijck, José. 2020. "Governing Digital Societies: Private Platforms, Public Values." *Computer Law & Security Review* 36. https://doi.org/10.1016/j.clsr.2019.105377.

van Dijck, José. 2014. "Datafication, Dataism, and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance and Society* 12 (2): 197–208. https://doi.org/10.24908/ss.v12i2.4776.

Wang, Ge, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. "'Don't Make Assumptions About Me!': Understanding Children's Perception of Datafication Online." *Proceedings of the ACM on Human-Computer Interaction* 6: 1–24. https://doi.org/10.1145/3555144.

Wilson, Michèle. 2018. "Raising the Ideal Child? Algorithms, Quantification and Prediction." *Media, Culture & Society* 41 (5): 620–36. https://doi.org/10.1177/0163443718798901.

Wolff, Annika, Daniel Gooch, Jose Cavero, Umar Rashid, and Gerd Kortuem. 2018. "Removing Barriers for Citizen Participation to Urban Innovation." In *The Hackable City: Digital Media and Collaborative City-making in the Network Society,* edited by Michiel de Lange and Martijn de Waal, 153-168. Limerick: Springe.

Yang, Dong, Huanhuan Wang, Ahmed Hosny Saleh Metwally, and Ronghuai Huang. 2023. "Student Engagement during Emergency Remote Teaching: A Scoping Review." *Smart Learning Environments* 10 (1): 1-17. https://doi.org/10.1186/s40561-023-00240-2.

Yates, Simeon, Elinor Carmi, Alicka Pawluczuk, Bridgette Wessels, and Justine Gangneux. 2020. *Me and My Big Data Report 2020: Understanding citizen's data literacies.* Liverpool: University of Liverpool. https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/documents/Me,and,My,Big,Data,Report,alts,3.pdf.

Zhao, Jun, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrook-Childs, Max Van Kleek, and Nigel Shadbolt. 2019. "I make up a silly name': Understanding Children's Perception of Privacy Risks Online." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, May 2019,* 1-13. https://doi.org/10.1145/3290605.3300336.

—
ANA KUBRUSLY
Ana Kubrusly is a PhD-candidate in Communication Sciences at NOVA University of Lisbon. She holds an individual grant funded by the Portuguese Foundation for Science and Technology (FCT), hosted at the Interdisciplinary Centre of Social Sciences (CICS.NOVA) and at the Polytechnic Institute of Setúbal. Her research focuses on children's and adolescents' relationships with digital environments, namely in terms of their literacy, skills, well-being, and digital cultures.
—
ORCID
0000-0001-6681-7193
—
CIÊNCIA ID
E912-5C44-558F
—
Institutional address
NOVA FCSH, Av. de Berna, 26C, 1069-061 Lisboa — Portugal.

—
LIDIA MARÔPO
Lidia Marôpo is an Associate Professor at the Polytechnic Institute of Setubal, Portugal, and a researcher at the Interdisciplinary Centre of Social Sciences at Nova University of Lisbon, Portugal. Her research mainly addresses the relationship between children, young people, and the media. She has participated in international projects and networks such as COST Action Transforming Audiences and ySKILLS. She has published in journals such as the Journal of Children and Media and Communication & Society and publishers such as Palgrave and Routledge.
—
ORCID
0000-0003-4687-7628
—
CIÊNCIA ID
801F-2368-4801
—
SCOPUS ID
56196180100
—
WEB OF SCIENCE ID
AAI-1374-2021
—
Institutional address
Campus do IPS — Estefanilha — 2910-761, Setúbal — Portugal.

—
SUSANA BATISTA
Susana Batista is an assistant professor at NOVA FCSH, where she coordinates the Master in Education Studies, and a researcher at CICS.NOVA. She has been a member of the Portuguese team of the European network EU Kids Online since 2017 and was a researcher for the ySKILLS Project (2020–2023, H2020). She is also a member of CriA.On, a Portuguese digital platform that provides resources on the digital contexts, skills, and rights of children and adolescents.
—
ORCID
0000-0003-2545-4538
—
CIÊNCIA ID
DC13-F045-3A7F
—
SCOPUS ID
55634403800
—
Institutional address
NOVA FCSH, Av. de Berna, 26C, 1069-061 Lisboa — Portugal.

—

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

—

To cite this article

Kubrusly, Ana, Lidia Marôpo, and Susana Batista. 2024. "Big data literacy for youth: an intervention agenda." *Revista de Comunicação e Linguagens* (60-61): 114-137. https://doi.org/10.34619/jvda-wgve.

—

Received Recebido: 2024-02-09 **Accepted** Aceite: 2024-07-31