



## The data subject and the myth of the 'black box' data communication and critical data literacy as a resistant practice to platform exploitation

Dennis Nguyen & Bjorn Beijnon

**To cite this article:** Dennis Nguyen & Bjorn Beijnon (2024) The data subject and the myth of the 'black box' data communication and critical data literacy as a resistant practice to platform exploitation, *Information, Communication & Society*, 27:2, 333-349, DOI: [10.1080/1369118X.2023.2205504](https://doi.org/10.1080/1369118X.2023.2205504)

**To link to this article:** <https://doi.org/10.1080/1369118X.2023.2205504>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 26 Apr 2023.



Submit your article to this journal [↗](#)



Article views: 2689



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

# The data subject and the myth of the ‘black box’ data communication and critical data literacy as a resistant practice to platform exploitation

Dennis Nguyen<sup>a</sup> and Bjorn Beijnon<sup>b</sup>

<sup>a</sup>Utrecht University, Utrecht, Netherlands; <sup>b</sup>University of Amsterdam, Amsterdam, Netherlands

## ABSTRACT

This conceptual paper explores the role of communication around data practices of Big Tech companies. By critiquing communication practices, we argue that Big Tech platforms shape users into data subjects through framing, influencing behaviour, and the black-boxing of algorithms. We approach communication about data from three perspectives: (1) current data communication constructs reductive data identities for users and contributes to the colonization of daily routines; (2) by strategically deploying the *black box* metaphor, tech companies try to legitimize abuses of power in datafication processes; (3) the logic in which communication is mediated through the interfaces of Big Tech platforms is normalizing this subjectification. We argue that critical data literacy can foster individual resilience and allows users to resist exploitative practices, but this depends on transparent communication. The opposite seems standard among tech companies that obfuscate their data practices. Current commercial appropriations of data ethics need to be critically assessed against the background of increasing competition in the digital economy.

## ARTICLE HISTORY

Received 1 September 2021  
Accepted 23 March 2023

## KEYWORDS

Data communication; data practices; Big Tech; data identities; critical data literacy

## Introduction

In this conceptual paper, we explore the role of communication around data practices of Big Tech companies. The aim is to critique how they shape users into data subjects through framing, influencing behaviour, and the black-boxing of algorithms. We consider digital platforms as means for controlling users’ subjectivity and argue that data subjects are kept in the dark about how their data identities serve economic and governance-related objectives via layers of mystification about data practices. The idea of an unexplainable, automated black box is a smokescreen that deflects from fundamental questions about the ethical implications of prevailing data practices and the quasi-permanent experimentation with data-driven designs on users. Instead of informing, creating clarity, and building trustworthiness, data-driven organizations apply communication

**CONTACT** Dennis Nguyen  [d.nguyen1@uu.nl](mailto:d.nguyen1@uu.nl)  Muntstraat 2A 3512 EV Utrecht, The Netherlands

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

strategies that serve plausible deniability and potentially undermine users' critical data literacy. We consider critical data literacy as a combination of conceptual knowledge and practical skills relevant for opinion formation about datafication processes and for establishing individual autonomy in digital society. Conceptually, critical data literacy covers awareness for and understanding of societal, political, and economic ramifications of current data practices in the public and private sectors. Practically, it includes capabilities for recognizing how digital media affordances datafy interactions, skills in reading and interpreting data as a means of communication and, where feasible, skills for protecting and managing one's own data. A deficit of critical data literacy exposes citizens to diverse risks and contributes to a gradual dissolution of autonomy (Carmi et al., 2020). Communication about data practices is an important factor in the development of critical data literacy. Only if organizations provide transparent information in ways accessible to different social groups, individuals may develop awareness, understanding, and eventually form opinions.

We argue that communication and data practices are closely interwoven, and we distinguish between two dimensions. First, there is communication *about* data practices. This concerns publicly accessible information about data-driven systems (products/services), terms and conditions, privacy statements, advertising, marketing, and public relations, including how tech companies participate in public discourses. Second, there is communication *through* data practices that are mediated via user-centric interfaces. Explicitly and implicitly, interaction possibilities for users embedded within digital interfaces translate monitored behaviour into data streams.

We aim to contribute to the theorization of datafication as a complex communication process that establishes power hierarchies in the making of meaning and construction of datafied identities in surveillance cultures (Lyon, 2017). The use of data is embedded in narratives that assign meaning to stakeholders, subjects, and their context-specific relationships (Dourish & Gómez Cruz, 2018). Datafication processes and narratives around them share similarities with the one-sided discursive construction of cultural-social identities in the pre-Internet age through gazing (Saïd, 1979), framing (Entman, 1993) and other forms of mediated representation serving domination (Butler, 1990). Big Tech companies establish a data gaze that casts users in a position of dependence.

While tech companies may not aim to differentiate themselves from their users as past Western colonial powers intended to when they described their non-Western subjects, similarities emerge in the ways the data subject is formed through data framing and data narratives that aim for maximizing economic exploitation and control (Couldry & Mejias, 2020). We propose that framing processes through data practices are manifestations of power exertion that need to be in focus of critical analysis. Datafication is communication, and previous research on subjectification through discourse helps with revealing how it operates and causes harm in society. The main goal is to connect framing to critical data literacy through scrutinizing multi-layered communication within and around digital interfaces. We argue that the complex role of communication is not always fully appreciated in research, and aim to expand existing theoretical frameworks by underlining its relevance for empirically investigating current trends. If 'data and data systems have never been and never will be asocial' (Lee & Cook, 2020, p. 8), then we need to unpack how data practices enforce social orders through communication.

This requires a birds' eye view. We explore communication and data practices of Big Tech companies through three angles. While there are numerous studies on each one of these, few consider their intersections in describing and critiquing the communicative logic and manifestation of ethically questionable data practices. The first part outlines how companies determine users' subjectivity, assign data identities, and grant limited possibilities for reclaiming and counter-framing themselves in datafication processes. Users' agency is not fully absent but tightly controlled on most platforms regarding data recognition and data representation, i.e., how they are subjected to datafication and represented as collections of data points. The second part examines communication practices and how the *black box* metaphor obscures Big Tech companies' failure to provide users with a clear understanding of how they affect them. Communication about data practices is a multidimensional process that involves different interrelated communicative means that exchange information with users about and through data-driven systems. The third part discusses how communication around data practices connects to critical data literacy, which we propose as a crucial element for building awareness and forming resistance. Overall, this conceptual paper aims to provide a new understanding of datafication from an interdisciplinary angle to demonstrate the value and necessity of investing in critical data literacy amongst the users of data-driven technology.

## Datafication and subjectification

### *Data practices as means of control and value extraction*

There is an ongoing public debate on the deceptive practices of Big Tech companies that fail to create transparency about their use of personal data. Various scandals illustrated risks that come with un- or under-regulated and ethically questionable data practices, such as privacy invasion and surveillance, algorithmic manipulation, data biases, and discrimination. One notorious example is audience targeting by Cambridge Analytica that attempted to influence voting behaviour of Facebook users during the US 2016 presidential elections and UK Brexit referendum. Other scandals involve the removal of competitor products by Amazon, the collection of unencrypted Wi-Fi data by Alphabet through Google Street View cars, or the audio recording of users via Apple's virtual assistant Siri. The tech discourse is marked by tensions and contradictions between tech-enthusiasm and calls for a more ethical direction in the digital transformation (Jørgensen, 2021).

In Western countries, a handful of tech companies symbolize the dominance of data-driven business models that are prone to bearing these risks: the 'Big Five', consisting of Apple, Amazon, Meta/Facebook, Alphabet/Google, and Microsoft. These organizations acquired quasi-monopolistic positions in the production and distribution of software and hardware. This dominance has provided them with control over datafication processes to which users are exposed daily. It is a self-reinforcing cycle. Datafication catalyses more datafication, as more accurate insights allow for more effective design interventions that collect more valuable data. Where the Big Five have not gained direct control, their ideological impact resonates in dominant narratives of data-driven progress across private and public domains (Jørgensen, 2021).

Although there is growing public awareness of the powerful position of Big Tech, societies might fail to grasp the socio-cultural implications that their products and

services have on people's lives. We are unable to foresee the unprecedented implications of control over our subjectivity that Big Tech companies have acquired. Through digital platforms, Big Tech determines who its data subjects are. Companies assign data identities to users that are expandable by connecting them with other commercially available data about the same individuals. Constant tracking and storing of user data allow for targeting them with specific commercial offerings that fit their (growing) data profiles.

The goal of Big Tech companies' constant data gathering is to capture their users in their perceived entirety, i.e., their continuous attention, behaviours, attitudes, social interactions, and leisure activities. Big Tech platforms turned attention into a commodity and try to keep users engaged 24/7 for virtually everything, i.e., socializing, working, health, education, entertainment, and consumption. Furthermore, platform users seek the attention of other users by means of selfies, messages, and constant updates (Marwick, 2015). Consumers have transformed into users that automatically opt-in with their attention to Big Tech's services (Williams, 2018). For example, users can set their alarm on their Android phone, adjust the temperature in their house with their Nest Thermostat, check their Nest security cameras around their house, cast their favourite show on their Chromecasts, and ask their Nest Audio to play music while checking their daily schedule. All these smart devices and services are datafying users' interactions within their digital media ecology. They are collecting and storing these data for the same company, in this case Alphabet.

Big Tech companies have become part of users' daily routines and habits while they took ownership over data that are not theirs, or did not even exist, in the first place. The omnipresence of these companies normalizes specific data practices that are beneficial for them but not necessarily for users. By purchasing and interacting with these products, users generate data but are not able to view the complete recorded data captured by interfaces that serve company interests. Big Tech companies are gatekeepers to data representations that they create for their users and retain full control over what parts they want to make accessible through a format that is economically attractive. They made conventional that users have ownership of the products that they buy, but they never own the data that are produced through them.

Data practices are shaped through interactions that users have with the affordances of platforms, their infrastructures, and interfaces. Users are involved with the synergetic way in which data practices are initiated, incorporated, and normalized. They are responsible for their own modes of datafication. For example, the success of Amazon's webshop is that developers changed its interface to users' experiences and demands by adding 'direct buy' buttons and prioritizing products on its webpage that are recommended to users based on data analysis. Such a synergy provides users agency to steer the interface to products that they are interested in, but it encloses this agency to the controlled, datafied space built by these companies.

Concerning data ownership, a common saying is: 'when you are not paying for the product, you are the product'. It is even worse than that. The following analogy shall clarify this: Imagine owning huge acres of land, all filled up with trees. Winter is coming, so you decide to chop up some of the trees on your land. To save time and energy, you decide to buy a woodchopper machine. The woodchopper makes your life easier, and you have more time to focus on other things while the job is completed without any physical human labour. After a while, you decide that enough trees have been chopped and

you want to collect your wood. When you arrive at your land, you see that all your wood has already been collected by the company that sold you the woodchopper. The company states that by purchasing their product, they are entitled to all the chopped wood. This analogy demonstrates that we have not only become the product but that we have become the soil on which data is harvested and transformed into economic value.

How did users come to accept this situation? The ownership of data by Big Tech companies has become conventional due to the public legitimacy that they acquired through offering efficacy, efficiency, and entertainment with their products and services (Horne & Przepiorka, 2021). Big Tech companies take great care not to reveal how their data practices create exploitative relationships in their self-portrayal while establishing their dominance for diverse daily activities in the lives of billions.

Under a layer of positive associations that promote narratives of value, efficiency, optimization, and perpetual improvement, Big Tech obscures datafication processes and how they claim users' data without resistance. They position themselves as inevitable, irreplaceable, and without alternatives. The perception and framing of technology companies only recently started to turn more critical in public discourses, which was triggered by several high-profile scandals. However, that did not yet drastically change how users assess the cost/risk-benefit balance when using data-driven services. Users may not be ignorant of Big Tech's selfish goals, but they may see little room for circumventing them.

### ***Subjugation and subjectification***

Through their conventional presence in various data practices, Big Tech companies shape their users' subjectivity. In building a postmodern theory of power, Foucault (1975/1977) argued that people's subjectivity is culturally formed through disciplining practices that transform individuals into specific kinds of law-abiding citizens. Foucault's discussion of Bentham's Panopticon is often used to draw parallels between current surveillance cultures in which citizens adapt to self-disciplining governmentality based on the idea that they might be watched. What Foucault did not foresee – but Deleuze (1992) did – is that disciplining happens nowadays through datafication. Our disciplinary societies have transformed into societies of control, in which ownership and control over data and information steer the subjectivity of citizens.

Conceptually, controlling practices of Big Tech companies should be understood through two positions of domination over users. First, *subjugation* as an act of making individuals subordinate to the domination of a power force (Foucault, 2003). Subjugation is exercised by means of control and dependence. For example, Apple controls and presents applications via its App Store. Users are subjugated to Apple's control over the store, and their dependence on it to find applications. Second, there is *subjectification*, which produces subjectivity by tying itself to the identity of individuals, so that it becomes part of their conscience and self-knowledge (Foucault, 2003). When individuals are subjectified, they are cultivated in how to abide by specific norms that through constant exposure become normalized and invisible. An example is again Apple's digital ecosystem. A range of Apple products and services can be automatically synchronized for optimizing user experience via diverse data streams, such as Apple's iPod, iPad, iPhone, iMac, Apple Watches, AirPods, HomePods, Apple TV and accompanying software (e.g.,

iCloud). The ways in which users make themselves dependent on Apple's offerings allow the company to steer users by means of their products and services, as they have become so connected with users' identities. It is through these two positions of domination that individuals submit to the powers that are exercised over them.

These two post-modern concepts demonstrate that this exercise of power is not just a controlling practice of repression by societally dominant entities. Power is merely a medium of change. This means that for both users and platform owners, power is not a thing, but a facility. Power 'traverses and produces things, it induces pleasure, forms knowledge, produces discourse. It needs to be considered as a productive network which runs through the whole social body, much more than as a negative instance whose function is repression' (Foucault, 1980, p. 119). Both users and Big Tech companies can use power to generate social change. Oppositional power to domination is often coined as a resistant entity, but as power is a transformative capacity, resistance is just a classification of power and not a different subsistence. Resistant practices by platform users, such as rewiring smart speakers and using ad blockers are categorically lesser forms of power compared to the dominant practices by Big Tech companies, but they are not powerless. Since Big Tech companies own the platforms and data, the distribution of the transformative capacity of power as a medium becomes asymmetrical. Big Tech companies seem to have users entrapped in a state of domination.

The cynical twist in this use of power is that Big Tech companies give users the illusion of freedom through providing spaces of possibilities, although these are formatted spaces of participation. Exemplary is Meta's upcoming Metaverse, which is a blend of augmented and virtual reality. Meta datafies users' bodies, movements, and perceptions so that any interaction in the Metaverse is experienced as an act out of users' own agency. However, considering the Metaverse as a platform of surveillance shows that the individual becomes a dividual that is scattered over the platform through datafication (Deleuze, 1992). This digital surveillance through real-time monitoring in ever finer granularity creates a *data gaze* that constructs data subjects who are reduced to economically exploitable metrics. The nudging and monitoring of users through data operationalizes digital surveillance that establishes subjugation and subjectification. A platform such as the 'Metaverse' might be experienced as a free space for resistance by users, but the asymmetric distribution of power in this space governed by Meta ensures that domination overshadows freedom.

The data gaze of Big Tech reduces users to quantifiable bundles of data points. To some extent, data practices display similarities to *Orientalizing* (Saïd, 1979), as they colonize their users through data. From a post-colonial perspective, institutionalized forms of power do not only attempt to own subjugated people in material terms (e.g., land, resources, labour force), but try to determine their cultural identities, social configurations, and histories. This shaping of cultural identities is embedded within how people and their practices are represented. Such a post-colonial perspective is a useful lens in critiquing and challenging dominant power asymmetries in the context of digital surveillance and datafication, as it provides theoretical foundation for the evaluation of ethical implications of datafication; it opens a way of examining data and the act of datafication as a tool of power and the shaping of subjectivity (Schie et al., 2020). A post-colonial lens emphasizes the value of a kind of user literacy that involves how data practices become normalized in users' ecologies. Through the design



of interaction possibilities and connected sensors, Big Tech companies decide what parts of a complex social, cultural, and physical reality are worth taking note of while creating invasive, expansive, yet reductive representations of users in the language of data. Data are not just exploitable metrics but are communication- and framing devices that take selected parts of social reality and embed these in narratives that organizations use to make decisions over target groups. This puts these organizations in a powerful position of determining the meaning of data in data narratives and what consequences these have on the present and future. Users have little to say about how they are looked at and represented in data narratives. It is through these practices that companies cluster, categorize, and label them with algorithmic identities (Cheney-Lippold, 2011). This is not to claim that current power imbalances in affluent digital societies are equal to the horrific plight of colonized peoples throughout history. The point that we want to make is that the principles at work in data discourses reflect in essence the same one-sided identity formation and control through framing practices by powerful organizations over targeted populations.

While users' agency is not fully devoid of control, the design of data-driven systems determines to a substantial extent what their data representations look like. Automated systems draw from user input in data-driven feedback loops that strive for optimization, usually with the goal to increase economic returns. They are responsive to user input to be able to observe, learn, and manipulate. In addition, users are subjectified into perfect attentional consumers willing to provide information about themselves when their attention is captured (Williams, 2018). For example, the longer YouTube can hold user attention by recommending yet another dog video, the more behavioural data users supply to Alphabet. It is through the design of devices and platforms that Big Tech companies determine how attention is captured, what data are collected, and what sort of subjects users are shaped into.

Although states of domination on each of these platforms might indicate that users are merely subjugated, they are most certainly subjectified through the entanglement of all these platforms in users' daily routines. Within users' personal media ecology, these companies datafy users by connecting themselves to their identity. Contemporary surveillance capitalism leaves little space for people to reject the presence of Big Tech in their lives due to their dependence on its services. However, critical data literacy can enable individuals – in their dual role as citizens and users – to recognize possibilities for critically engaging with the omnipresence of Big Tech in different ways, such as questioning their dominance, re-appropriating digital media affordances for their own goals, as well as individual and collective resistance through politicizing data practices (e.g., demanding regulation, boycotting brands (Gray et al., 2018)). The conceptual issues that are laid out in the above theories about power demonstrate the necessity of critical data literacy in questioning power relations that get normalized over time.

## **Data communication and 'black-boxing'**

### ***Communication about data practices***

To most users, datafication processes through and behind Big Tech's interfaces appear invisible and the inner workings of algorithms are perceived as hidden within a *black*



*box* (Pasquale, 2015). Information around applications and platforms seems opaque, imprecise, unclear, distracting, confusing, and difficult to understand. Claims about unexplainable artificial intelligence (AI) even from experts contribute to an aura of vagueness and secrecy around data practices and their consequences. Ideally, clear communication informs users about the exact purpose and extent of data collection, data analysis, data retention, and data sharing (Jia & Ruan, 2020). Users should be able to understand what they sign up for through informational content such as app store pages, consent forms, pop-ups, terms and conditions, and privacy policies. These interventions, even if demanded by law, hardly increase transparency and accessibility (Oeldorf-Hirsch & Obar, 2019). Content and formats are often too long and complex or entirely absent, so users cannot make well-evaluated decisions over whether they comply with the terms presented to them. Users usually learn very little about how the data gaze creates a representation of them that serves specific organizational objectives.

Communication can create awareness, counter misconceptions, inform about benefits and risks, and build trust. However, Big Tech companies approach communication mostly through a Public Relations (PR) lens. They use keywords and make promises but eventually fail to inform about data practices openly and effectively. Admittedly, full transparency about data practices is not in the interest of data-driven companies, since secrecy is integral to the business models that enabled rapid growth and reaping enormous profits.

Communication about data practices is a multidimensional process that happens at different stages and moments of engagement with a digital system. It starts with the initial download and installation, through the sign-up and setup of a system (e.g., profile creation), to the eventual usage and sign-off or deletion of the application. Communication includes informational content (e.g., terms and conditions, PR, advertising) but is also deeply embedded within the constant interaction with data-driven products and services in users' media ecologies.

### ***Algorithmic transparency and the myth of the black box***

One solution that is suggested for rectifying the power imbalance between users and Big Tech is more transparency about algorithmic decision-making. Companies tend to keep their algorithmic designs secret, in effect black-boxing their mechanisms of control. This does not necessarily mean that algorithms are truly impossible to understand. Indeed, the algorithmic code itself is not publicly accessible, but users see how their input affects the outcomes within the interface (Ytre-Arne & Moe, 2021). More important than mathematical formulas translated into lines of code are the social contexts in which algorithms have an effect and how they are shaped by actions outside of the technical (Amoore, 2020). In line with Latour (2005), Bucher (2018) proposes a technography of algorithms for studying them as assemblages that occur in a specific time and space. Observing when algorithms are happening as an event through the interaction of inputs could help with charting the epistemological process to the output. Users can discover and make use of algorithmic affordances without advanced technical knowledge about machine learning and predictive statistics. Algorithmic functioning can be experienced through the way users make use of them. For example, through watching videos on YouTube and

examining changes in recommendations, users discover the machine learning practices behind the user interface, without interacting with algorithms on a technical level. In this way, audiences could develop the critical understanding of algorithms needed to judge them and hold their creators accountable (Kemper & Kolkman, 2019). Recognizing algorithmic patterns and their effects does not come naturally but depends on targeted educational interventions paired with transparent communication about data practices. Cultivating such awareness and understanding is not per se dependent on numeric-/mathematical literacy.

Where current calls for transparency are about the technical dimension of algorithms behind data-driven interfaces, we argue that demanding transparent communication about the black box of algorithms is more urgent. As Bucher (2018) shows, tech companies frame their algorithms as *strategic unknowns*. For Big Tech, 'cultivating ignorance is often more advantageous, both institutionally and personally, than cultivating knowledge' (McGoey, 2012, p. 555), as it allows to excuse themselves for the detection of issues that were allegedly undetectable beforehand. Strategic unknowns provide companies with ways for not taking responsibility through layers of induced ignorance. When scandals emerge, companies can immediately jump the bandwagon and apologize for the programmed unknowns within their products and services (e.g., when Alphabet's AI labelled black people as gorillas (Nieva, 2015)). The black box metaphor has become a communication strategy that is pulled by Big Tech to avoid taking responsibility for mistakes and for legitimizing (un)intended abuses of power.

Instead of arguing for more transparency in algorithmic code, users should call for more transparency in communication about algorithms, i.e., what their goals are, how they work conceptually, and what data they use. However, it is important to consider the limitations of transparency since making processes within algorithmic designs visible does not per se lead to fairness, trust, and net-wins for users (Ananny & Crawford, 2018). A broader view is needed that critically evaluates the context in which data-driven AI is deployed, including all relevant actors (companies, regulators, users).

Big Tech companies should no longer portray their own algorithms as unknowns and black boxes but rather take responsibility and accountability for the algorithmic decisions that they set in motion. The same black box metaphor extends itself to data ownership. One example is how in 2019 Alphabet excused itself for unknown microphones in its Nest alarm systems that were not even present in the patent blueprints of the device. Although Google publicly apologized for the microphones and claimed they were not meant to be secret, the communication on their data practices still invokes the idea that the company itself is uncovering the unknowns at the same time as its users. These kinds of scandals show how power is a medium of social change that is expressed by users. There is a clear desire for more agency and transparency by users of Alphabet's services, while there is also the need to use these services. For users, breaking the black box is essential to express their power to eventually generate transformative change in how these companies operate. To counter excuses based on its own black boxes, it is essential to study how Big Tech frames algorithms and tech specs as unknowns, for whom, and for what purpose.

## Data communication and critical data literacy

### *Recognizing data practices and discovering affordances*

Interfaces are the time and space in which the assemblage of inputs comes together as an output for users (Galloway, 2012). The interface is a medium and primary lens for the data gaze. It frames the user in specific contexts and converts behaviour into data points. The sensors that enable this gaze are not limited to the functionalities that are visible to the user but are partly hidden to capture valuable exhaust data (Kitchin, 2014).

To unpack interfaces, it is important to centre on the elements that invite users to engage with them in specific ways. These *affordances* steer the interaction that users have with the technology through e.g., the use of buttons, ways of scrolling through an interface or using a search engine (Hartson, 2003). Affordances can be examined in ways of communication, as they invite users to do things, but the underlying data practices and potential forms of surveillance behind them are not always visible or self-explanatory. For example, Facebook built the News Feed and its interaction possibilities as an important affordance to consume content. Through constant interactions with interfaces, users grow into specific patterns for engaging with these interfaces in what become conventionalized ways.

### *Critical data literacy and data emancipation*

To describe this learning process through conventional interactions that users have with Big Tech interfaces, we use the term *literacy*. Conceptually, literacy concerns ‘the relationship of learners to the world, mediated by the transforming practice of the world taking place in the very general milieu in which learners travel’ (Freire & Macedo, 1987: viii). Users become literate about the ways in which platforms require interactions from them and what data they need to operate (Crary, 2014). Whenever users interact with Big Tech interfaces, they become literate to the interactions that they have with these, which could be a smartphone, a laptop screen, or a voice-controlled device such as Alexa. When operating systems such as Android and iOS dominate the market, Alphabet and Apple can monopolize the ‘grammar and vocabulary’ needed to become proficient in using their devices. The predominant position of Big Tech is indirectly responsible for shaping users’ literacy.

There are possible alternatives to dominant modes of inducing literacy in technology use. Most pressingly, the idea of literacy needs to expand to abilities in understanding data practices and critically scrutinizing communication strategies of tech companies. Users would benefit from being literate in how data and algorithms are considered black-boxed, while simultaneously comprehending the technography of how algorithmic outputs are formed by different connections that are made in the assemblage of algorithmic decision-making represented within the interface. This kind of literacy is what we put forward as *critical data literacy*.

In recent years, data literacy became an important research concept in the intersection of critical data studies, media literacy research, and media studies. We can generally differentiate between technical-practical definitions (Wolff et al., 2016) and critical-conceptual ones (Gray et al., 2018). These should not be seen as separate boxes, as practical skills and conceptual understandings connect. More technical definitions of data literacy

focus on (basic) knowledge about statistics and numeric skills (Gould, 2017). For example, being able to read data visualizations is considered an important skill for informed citizens (Kennedy et al., 2020). In a broader sense, ‘technical’ data literacy also relates to practical internet skills, and computer skills (Frank et al., 2016). Critical-conceptual proposals move beyond technical skills. They prioritize conceptual knowledge and skills in critical thinking about datafication and automation (cf. Carmi et al., 2020; Pangrazio & Selwyn, 2018), especially regarding the impact on individual privacy, fairness, equality, inclusion, safety, and access to technological benefits and freedom (Park, 2013). Both general types of data literacy aim for individual empowerment.

Building on Gray et al. (2018), we argue that a focus of data literacy on numerical and statistical knowledge is important but too limited to deal with the societal and political implications of datafication. It may be unrealistic to expect large proportions of society to acquire hands-on skills in statistics. What is required is an infrastructural way of thinking: to examine the extent to which platforms are influencing data practices and users’ perception of them (Gray et al., 2018). We propose that critical data literacy determines how aware users are of datafication processes and how much they understand about their impacts. It transcends practical skills relevant for using data and expands into a conceptual understanding of data as tools of control, media of communication, and powerful framing devices in narratives that have effects on users individually and collectively. Critical data literacy then becomes a prerequisite for individuals to effectively engage in critical debates about data justice (Dencik et al., 2019) and data politics (Jørgensen, 2021). Other related forms of literacy are coding literacy (Vee, 2017) and algorithmic literacy (Silva et al., 2022). Coding literacy challenges us to view programming as a use of language with all the social, cultural, and political implications that this entails. Algorithmic literacy focuses on laypeople’s conceptual understanding of algorithms commonly subsumed under AI. Both forms of literacy have strong links to our proposal of critical data literacy and can be viewed as specializations within its framework.

Critical data literacy by itself neither leads to a more cautious stance towards data practices per se nor automatically increases the likelihood of an individual user to proactively take data protection measures. People may know a lot about data practices and their potentially negative impacts, but they may not care, simply resign, or willingly accept risks for having access to the value that they see in digital technology (Solove, 2021). Yet critical data literacy is the fundament for making better-informed choices about the use of data-driven technology and developing a political stance on datafication. This is the starting point for exploring forms of political action and resistance, provided critical data literacy is combined with education about emancipatory strategies (Pangrazio & Selwyn, 2019). Critical data literate individuals can better protect their personal data by understanding privacy settings, may consider data practices as a factor in choosing between digital services, possibly boycott and protest tech companies, and contribute to the politicization of datafication by demanding regulation from political decision-makers. Each of these actions need concrete strategies to move from critical data literacy as a mere foundation for awareness to a catalyser for change (Pangrazio & Sefton-Green, 2020).

Building data literacy at several fronts through (public) education, critical news reporting, and transparent communication by ethical data-driven organizations could lead to the formation of data civics that are capable of challenging current power

imbalances politically (Andrejevic, 2020). While strong arguments have been made for formally including data literacy into educational policies (Knaus, 2020), it is for the European context largely unclear whether, where and how exactly the issue is integrated in school- and university curricula. There does not seem to be a concerted effort in the form of European and/or national education policies. Considering the transnational scope of the challenge and how tech companies try to exploit national legal frameworks to their benefit, taking a European-transnational approach seems necessary for the European context. Within the European Union (EU), collaboration and alignment in regulatory responses to big tech have led to some successes in the past (e.g., the GDPR or Digital Services Act). In respect to education, the EU has infrastructures for the exchange of knowledge and educational policies that could facilitate the formal introduction of critical data literacy across member-states.

Current forms of using resistant power as a transformative capacity for social change are some political institutions that attempt to make citizens more critical data literate. Examples are new EU legislations such as The Data Act which ‘will ensure fairness in the digital environment, stimulate a competitive data market, open opportunities for data-driven innovation and make data more accessible for all’ (European Commission, 2022). Such legislation should provide platform users more access to data, allow them to switch easily between cloud data-processing services, and safeguard against unwanted data transfers. While this can appear abstract for average EU citizens, such measures could become more tangible through representation via local aldermen and councilors. An example is the representative for Digitalization, Citizen Services Participation, and EU Affairs in Frankfurt, Germany. Also, European cities such as Barcelona and Frankfurt are investing in new digital structures of institutionalized citizen participation to enable citizens to engage in the future of their city through different digital formats. Critical data literacy is taken more seriously among some European politicians but still requires broader macro-legislations before these resistant forms of power can be classified as power that is not dominated by Big Tech. The growth in critical data literacy amongst platform users should therefore not only generate more awareness of datafication practices, but above all social change through data ownership. As digital republics like Estonia have shown, investing in both digitalization and critical data literacy gives citizens more trust in black-boxed data practices and their governance (Tsap et al., 2020). Just as Estonian citizens feel more empowered in their personal control of privacy and data (Priisaul & Ottis, 2017), others worldwide would be able to politically participate more in their governance if they become as critical data literate as the average Estonian.

Governments are not the only actors that can and should lead critical data literacy initiatives. Governmental organizations themselves make use data and AI in ways similar to tech companies (and often use technology services from the private tech sector). This can entail the intended or unintended negative effects of current data communication mentioned above. State institutions need to be held accountable for their data practices. A third critical perspective is needed through journalism and organizations in the civil society (e.g., Netzpolitik.org in Germany or Privacy International in the UK). Second, building critical data literacy should not be confined to primary, secondary, and higher education but needs to be possible for diverse demographic groups in society. The limited scope of public education needs to be supplemented by civil society organizations and NGOs (e.g., School of Data Germany).

Communication practices of data-driven organizations can have an impact on users' understanding of datafication and automation. Clear communication about how data is being collected, stored, shared, processed can create trustworthiness and confidence while it may address potential risks and how organizations deal with them. Some tech companies such as Apple began to acknowledge this. However, it is important to critically assess Apple's policies against the background of fierce competition between Big Tech companies. Ethical data practices may become additional selling points without addressing the core issues of subjugation and subjectification. Critical monitoring of data practices in the private and public domains by e.g., independent journalism and non-governmental organizations are indispensable for a transparent information ecosystem about data practices.

## Conclusion

This conceptual paper explored the role of communication around data practices of Big Tech companies. Through using theories and concepts from different fields, we provided an interdisciplinary understanding of datafication that demonstrates the value of critical data literacy. We argue that communication should be examined from three perspectives that matter in the appropriation of data and subjectivity of users. First, communication about the controlling practices of Big Tech companies is molding users into subjects who automatically opt-in for surveillance when entering a platform. Simultaneously, these companies colonize users' routines and appropriate their data. Through this data colonization, they frame users into data identities as compilations of metrics. Second, by communicating about their algorithms as black boxes, Big Tech companies deploy strategic unknowns that legitimize their abuses of power regarding their data practices. By enforcing the concept of the black box, companies obscure communication about their data practices that users then often subliminally accept. Third, the logic in which communication is mediated through the interfaces of Big Tech platforms is normalizing the subjectification by these companies through their control over the discourse.

Although the controlling practices of Big Tech companies seem inevitable, we have outlined ways that may empower users to take control over their data and subjectification. We suggest that critical data literacy, as a way of becoming aware of datafication practices and the molding of user perceptions, offers ways of questioning controlling practices. Big Tech companies may not have much interest in critical data literacy, as sustaining the strategic unknowns of their technologies allows them to remain in a position of power and deniability. By making strategic unknowns known, users would be able to call out Big Tech companies on their subjugating practices. As Big Tech companies might not proactively encourage users to develop this literacy, critical data literacy should be promoted and initiated by governments. A first step to demonstrate the value of critical data literacy by governments would be appointing local, national, international ministers and ambassadors with a focus on datafication practices, data transparency, and educating people in their critical data literacy. There is need for a concerted effort to integrate critical literacy systematically in educational programs of schools and universities. It is from a moral and ethical obligation that governments and users should challenge data practices and foster empowerment.



One might ask to what extent tech companies can act ethically and remain economically viable. Search-engine platforms such as Brave, DuckDuckGo, and Swisscows show that it is possible to be an ethical company while sustaining growth. These companies provide de-personalized search results for their users and block trackers, which means that users are using data without simultaneously being datafied. When looking at the Big Five, it seems like a longshot that these companies would adapt their business model to the logic of these smaller platforms. However, these examples illustrate alternatives that users could imagine outside the scope and logic of Big Tech data practices.

The question remains what communication research could practically contribute. We argue that communication research is vital in analysing the power of Big Tech companies and should explore ways of addressing data ownership with or without the intervention of Big Tech companies. As this conceptual paper argues, a communication perspective offers an interdisciplinary intervention by combining different theoretical lenses on the communication relationships between users and Big Tech. Focusing on black boxes and the PR by these tech giants obscures the real issue with user communication. Communication research should investigate the normalizing communication on data practices by these companies to give users control over their media ecologies and their data.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

### Notes on contributors

*Dr. Dennis Nguyen* is Assistant Professor for Digital Literacy and Digital Methods at Utrecht University in the Netherlands. He holds a PhD in Media, Culture & Society from the University of Hull (UK). His main research interests are critical data studies, public discourses on datafication, digital culture, and empirical methods for media research.

*Bjorn Beijnon* is a PhD candidate at the Amsterdam School for Cultural Analysis, University of Amsterdam, and a lecturer at the Institute for Communication, HU University of Applied Sciences Utrecht. His interest lies in digital subjectification by big tech companies through attentional capturing. He has published articles on numerous topics, varying from the visibility of consciousness to the appliance of smartwatches. Through cultural analyses, Deleuzian theory and media ethnographies, he currently studies how digital platforms subjectify their users in contemporary surveillance cultures, with a focus on online conspiracy theories on Facebook and techniques of resistance by DuckDuckGo users.

### References

- Amoore, L. (2020). *Cloud ethics*. Duke University Press.
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>



- Andrejevic, M. (2020). Data civics: A response to the “ethical turn”. *Television & New Media*, 21(6), 562–567. <https://doi.org/10.1177/1527476420919693>
- Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Interview Policy Review*, 9(2), 1–22. <https://doi.org/10.14763/2020.2.1481>
- Cheney-Lippold, J. (2011). A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28(6), 164–181. <https://doi.org/10.1177/0263276411424420>
- Couldry, N., & Mejias, U. A. (2020). *The costs of connection: How data are colonizing human life and appropriating it for capitalism*. Oxford University Press.
- Crary, J. (2014). *24/7*. Verso.
- Deleuze, G. (1992). Postscript on the Societies of Control. *October* 59, 3–7.
- Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society*, 22(7), 873–881. <https://doi.org/10.1080/1369118X.2019.1606268>
- Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, 5(2), 1–10. <https://doi.org/10.1177/2053951718784083>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- European Commission. (2022, February 23). Data act: Commission proposes measures for a fair and innovative data economy. *European Commission Press corner*. [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_22\\_1113](https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113)
- Foucault, M. (1977). *Discipline & punish: The birth of the prison* (A. Sheridan, Trans). Vintage Books. (Original work published 1975).
- Foucault, M. (1980). *Power/knowledge* (C. Gordon, Ed). Pantheon.
- Foucault, M. (2003). *Society must be defended: Lectures at the College de France, 1975–76*. Allen Lane The Penguin Press.
- Frank, M., Walker, J., Attard, J., & Tygel, A. (2016). Data literacy – What is it and how can we make it happen? *The Journal of Community Informatics*, 12(3), 4–8. <https://doi.org/10.15353/joci.v12i3.3274>
- Freire, P., & Macedo, D. (1987). *Literacy. Reading the word and the world*. Bergin and Garvey Publishers.
- Galloway, A. (2012). *The interface effect*. Polity.
- Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22–25. <https://doi.org/10.52041/serj.v16i1.209>
- Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 1–13. <https://doi.org/10.1177/2053951718786316>
- Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, 22(5), 315–338. <https://doi.org/10.1080/01449290310001592587>
- Horne, C., & Przepiorka, W. (2021). Technology use and norm change in online privacy: Experimental evidence from vignette studies. *Information, Communication & Society*, 24(9), 1212–1228. <https://doi.org/10.1080/1369118X.2019.1684542>
- Jia, L., & Ruan, L. (2020). Going global: Comparing Chinese mobile applications’ data and user privacy governance at home and abroad. *Internet Policy Review*, 9(3), 1–22. <https://doi.org/10.14763/2020.3.1502>
- Jørgensen, R. F. (2021). Data and rights in the digital welfare state: The case of Denmark. *Information, Communication & Society*, 26(1), 123–138. <https://doi.org/10.1080/1369118X.2021.1934069>
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081–2096. <https://doi.org/10.1080/1369118X.2018.1477967>

- Kennedy, H., Weber, W., & Engebretsen, M. (2020). Data visualization and transparency in the news. In M. Engebretsen, & H. Kennedy (Eds.), *Data visualization in society* (pp. 169–188). Amsterdam University Press.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1–12. <https://doi.org/10.1177/2053951714528481>
- Knaus, T. (2020). Technology criticism and data literacy: The case for an augmented understanding of media literacy. *Journal of Media Literacy Education*, 12(3), 6–16. <https://doi.org/10.23860/JMLE-2020-12-3-2>
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford University Press.
- Lee, A. J., & Cook, P. S. (2020). The myth of the “data-driven” society: Exploring the interactions of data interfaces, circulations, and abstractions. *Sociology Compass*, 14(12), 1. <https://doi.org/10.1111/soc4.12691>
- Lyon, D. (2017). Surveillance culture: Engagement, exposure, and ethics in digital modernity. *International Journal of Communication*, 11(2017), 824–842.
- Marwick, A. E. (2015). Instafame: Luxury selfies in the attention economy. *Public Culture*, 27(1), 137–160. <https://doi.org/10.1215/08992363-2798379>
- McGoey, L. (2012). The logic of strategic ignorance. *The British Journal of Sociology*, 63(3), 533–576. <https://doi.org/10.1111/j.1468-4446.2012.01424.x>
- Nieva, R. (2015, July 1). Google apologizes for algorithm mistakenly calling black people ‘gorillas’. *Cnet*. <https://cnet.com/tech/services-and-software/google-apologizes-for-algorithm-mistakenly-calling-black-people-gorillas/>
- Oeldorf-Hirsch, A., & Obar, J. A. (2019). Overwhelming, important, irrelevant: Terms of service and privacy policy reading among older adults. *Proceedings of the 10th International Conference on Social Media and Society* (pp. 166–173).
- Pangrazio, L., & Sefton-Green, J. (2020). The social utility of ‘data literacy’. *Learning, Media and Technology*, 45(2), 208–220. <https://doi.org/10.1080/17439884.2020.1707223>
- Pangrazio, L., & Selwyn, N. (2018). “It’s not like it’s life or death or whatever”: Young people’s understandings of social media data. *Social Media + Society*, 4(3), 1–9. <https://doi.org/10.1177/2056305118787808>
- Pangrazio, L., & Selwyn, N. (2019). ‘Personal data literacies’: A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, 21(2), 419–437. <https://doi.org/10.1177/1461444818799523>
- Park, Y. J. (2013). Digital literacy and privacy behavior online. *Communication Research*, 40(2), 215–236. <https://doi.org/10.1177/0093650211418338>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Priisaul, J., & Ottis, R. (2017). Personal control of privacy and data: Estonian experience. *Health and Technology*, 7(4), 441–451. <https://doi.org/10.1007/s12553-017-0195-1>
- Saïd, E. W. (1979). *Orientalism*. Vintage Books.
- Schie, G., van Smit, A., & López Coombs, N. (2020). Racing through the Dutch governmental data assemblage: A postcolonial data studies approach. *Global Perspectives*, 1(1), 12779. <https://doi.org/10.1525/gp.2020.12779>
- Silva, D. E., Chen, C., & Zhu, Y. (2022). Facets of algorithmic literacy: Information, experience, and individual factors predict attitudes toward algorithmic systems. *New Media and Society*. <https://doi.org/10.1177/14614448221098042>
- Solove, D. J. (2021). The myth of the privacy paradox. *89 George Washington Law Review*, 89(1), 1–42.
- Tsap, V., Lips, S., & Draheim, D. (2020). EID public acceptance in Estonia: Towards understanding the citizen. *The 21st Annual International Conference on Digital Government Research* (pp. 340–341). Association for Computing Machinery, New York, NY.
- Vee, A. (2017). *Coding literacy. How computer programming is changing writing*. MIT Press.

- Williams, J. (2018). *Step out of our light: Freedom and resistance in the attention economy*. Cambridge University Press.
- Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3), 9–26. <https://doi.org/10.15353/joci.v12i3.3275>
- Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society*, 43(5), 807–824. <https://doi.org/10.1177/0163443720972314>