

“What is your digital identity?” Unpacking users’ understandings of an evolving concept in datafied societies

Media, Culture & Society

2025, Vol. 47(2) 336–353

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01634437241282240

journals.sagepub.com/home/mcs**Jussara Rowland** 

Instituto de Ciências Sociais da Universidade de Lisboa, Portugal

João Esteves

Portuguese Institute of International Relations, NOVA University of Lisbon, Portugal

Abstract

Digital identity has become a central concept in understanding how people’s online presence is shaped and made sense of. Although extensively studied, the prevailing focus has been on how online identities are shaped by digital platforms or how users curate and perform these identities. How users perceive and assess this concept, however, has received comparatively less attention. In this article, we take a qualitative, user-centric approach to the meaning of digital identity, drawing on insights from 17 online focus group discussions involving 86 participants in Portugal. We identify three distinctive understandings – digital identification, self-presentation, and the datafied self – each corresponding to specific facets of users’ online experiences. Our findings underscore the multidimensionality of digital identity, highlighting its dynamic nature and potential for ongoing reinterpretation. This work contributes to the existing literature by highlighting how the concept’s ever-expanding complexity relates to people’s sensemaking practices about identity, agency, and digital platforms in datafied societies.

Keywords

datafication, digital identification, digital identity, self-presentation, social media

Corresponding author:

Jussara Rowland, Instituto de Ciências Sociais da Universidade de Lisboa, Avenida Professor Anibal de Bettencourt, 9, Lisbon 1600-189, Portugal.

Email: jussara.rowland@ics.ulisboa.pt

Introduction

In the digital age, where screens and algorithms mediate our interactions, the concept of identity has undergone profound transformations. Identity, a complex construct with significant implications on personal, political, and legal levels (Krotoski and Hammersley, 2015), has its origins in the Latin word “idem,” signifying “the same” (Turkle, 1995). Paradoxically, in contemporary times, it has come to encompass a wide range of definitions and applications. As Florian Coulmas (2019: 2) aptly notes: “Sameness and difference, this is what identity is all about. It should be simple, but it isn’t. For ‘identity’ means different things to different people and in different scientific disciplines”.

The emergence of the internet and the subsequent introduction of a “mass self-communication” paradigm, as proposed by Castells (2007), brought new dimensions to our understanding and conceptualization of the term and development of the notion of digital identity. Just as the overarching concept of identity, digital identity is subject to varied interpretations across different academic fields and discourses. For instance, in social media studies, digital identity has often been examined through the lens of online interactions and identity construction in virtual spaces (Marwick, 2013; Papacharissi, 2010). This perspective considers not only the performative nature of digital identities such as self-presentation and identity expression (Buss et al., 2022; Kavakci and Kraeplin, 2017; Liu, 2024) but also explores how platform designs guide and constrain users’ actions and interactions (Dyer and Abidin, 2022). In critical data studies, the focus on digital identity tends to highlight the social and ethical implications of data collection, surveillance, and control (Zuboff, 2019), alongside issues regarding power dynamics and exclusion mechanisms caused by algorithmic decision-making and automated systems (Burrell and Fourcade, 2021; Noble, 2018). Information Systems literature, meanwhile, focuses on aspects such as identity verification, authentication, and authorization, while also considering concerns about trust and privacy (Masiero, 2023; Sullivan, 2018).

These perspectives, while not mutually exclusive, illustrate how differently this concept can be conceptualized, explored, or enacted in contemporary societies. As posited by Søe and Mai (2022), the varying interpretations of digital and data identity can be seen as metaphors, each signifying a distinct viewpoint on the interplay between identity and digital information. Although analyses of digital identity have sometimes faced criticism for their emphasis on the online domain – which risks reinforcing the problematic dichotomy between digital and real-life identities (Cover, 2016; Davis, 2014; Schüll, 2019) – these perspectives are instrumental in unpacking the multifaceted socio-technical dynamics of how digital technologies interact with personal identity.

While existing studies provide valuable insights into the various conceptualizations of digital identity, they sometimes fail to explore the full range of meanings that can be assigned to the concept. Nevertheless, these meanings are important, particularly as digital identity becomes a term users are increasingly likely to encounter in their everyday lives, in contexts such as their social media activity, digital footprint management, or engagement with national or de-facto digital identity schemes (Feher, 2019; Sullivan, 2018; Szulc, 2019). Therefore, understanding how individuals perceive and use this concept in their day-to-day activities becomes a vital part of grasping its evolving role in contemporary society.

This article aims to advance our understanding of the concept of digital identity by adopting a user-centric approach. Relying on 17 online focus group discussions conducted in Portugal, which included 86 participants as part of the “We, The Internet” citizen dialogue, this paper interweaves insights from digital identity and datafication studies. It also contextualizes the term “digital identity” by tracing its evolution over the past 25 years. The goal is to illuminate how users conceptualize digital identity, emphasizing its complex and multifaceted nature and its numerous possibilities for varied (re) interpretations and conceptualizations.

Conceptualizing digital identity

To fully grasp the multiple meanings attributed to digital identity today, it is important not only to understand its interconnections with various fields of application but also to acknowledge its deep-rooted ties to the history of the internet and how it was conceptualized in response to the emergence of novel media technologies and practices (Han, 2019). In the early days of the Internet, the emphasis revolved around anonymity, signifying the notion that online individuals could establish an identity entirely distinct from their real-life persona (Han, 2019; Marwick, 2013). Sherry Turkle was particularly instrumental in popularizing the idea of multiple and flexible online identities. In her 1995 book “Life on the Screen,” she highlighted how the relative anonymity of people’s online experiences opened possibilities for a variety of role-play games and the adoption of decentered identities. Turkle argued that this phenomenon reinforced an existing cultural trend, encouraging the perception of identity as diverse and flexible (Turkle, 1999).

The idea of detaching oneself from the physical body played a significant role in this perspective: “as others cannot see who we really are, we are free to claim to be whoever we want to be” (Zhao, 2005: 388). This “disembodied communication” was viewed, for a moment, as an opportunity that could potentially eliminate discrimination based on race, sex, gender, sexuality, or class (Marwick, 2013). However, this notion was gradually challenged by users’ actual experiences with online services (Han, 2019). Scholars also began to emphasize how traditional categories used for identification and discrimination, such as gender, age, or race, could take on various forms in the online domain. This highlighted how systemic inequities, namely racism and sexism, could manifest online in ways that went beyond mere personal appearance (Han, 2019; Marwick, 2013; Nakamura, 2013).

In the late 2000s, with the rising popularity of social media platforms such as Facebook, the understanding of digital identity underwent significant change, moving closer to the idea of a “networked self” as described by Papacharissi (2010). Boyd (2010) highlighted the importance of networked publics in the shaping of digital identities, emphasizing the crucial role of online profiles in identity performance on the internet. According to her, online profiles provided users on social networks with a space to “write themselves into being,” allowing them to express different aspects of their identity in digital spaces. These digital identities were constructed based on four specific affordances of networked publics: persistence, replicability, scalability, and searchability. This was because the personal data they are built upon was indefinitely available, easily duplicated, highly visible, and easily searchable. Additionally, these digital identities were

also increasingly being subjected to “context collapse” which means they are performed to multiple audiences in a single digital context (Marwick and boyd, 2011).

Concurrently, the 2000s also marked the era where digital identity started to be discussed through the lenses of information privacy and identification (Camp, 2004). The central concern revolved around the absence of robust self-authenticating mechanisms on the internet, akin to those employed offline. Pioneering authors like Lessig (1999) advocated for the establishment of an identity-based regulation relying on discrete and verifiable attributes such as name, sex, address, education, driver’s license number, and social security number, among others. These permanent or long-lived temporal attributes could be assigned to an entity that would allow identity proofing, authentication, and authorization – fundamental elements to the foundation and development of future national and international digital identity programs (Sullivan, 2018).

By the 2010s, scholars began voicing concerns over the increasing desire to fixate identities into single entities. They observed how commercial, governmental, and cultural forces discouraged the idea of embracing multiple and performative experiences of identity online by actively defining online personas that differ from offline identities as “risk factors” (Lareki et al., 2023; van Zoonen, 2013) and by promoting the creation of singular, anchored identities (Szulc, 2019; van Dijck, 2013). van Dijck (2013) specifically critiqued social media platforms for championing the notion of a singular, transparent self, or identity. She argued that digital platforms’ interfaces do more than provide spaces for self-expression, they actively shape our public identities to enhance long-term traceability of social behavior.

Echoing a similar perspective, Szulc (2019) discussed how platformization has translated datafication logics into the creation of unique profiles. Acknowledging that the push to the constitution of individual profiles is not unique to digital culture, the author contends that social media platforms “enable and incentivize particular identity performance and construction, both online and offline since the two are mutually constitutive” (Szulc, 2019: 14). The author concludes that, for commercial reasons, platforms incentivize their users into making abundant but anchored selves: identities that are capacious, complex, and volatile but also singular and coherent. However, research also highlights the online curation work that users, especially those from marginalized groups, perform across multiple platforms to escape this push, enabling them to express multiple, or emerging identities (Buss et al., 2022).

With the intensification of commercial-driven personal data tracking and extraction, attention also started to be given to the impact of algorithmic processes on identity construction. Cheney-Lippold (2011; 2017), for example, explored how algorithms collect, classify, and make sense of our data, ultimately defining what can be termed as our “datafied selves” – consumer profiles that aggregate a user’s visible and invisible online identity information, establishing dynamic correlations between different types of available data. According to the author, a new algorithmic identity emerges from this constant and dynamic interaction between the data we produce through our ever-changing online behavior and the various State and corporate entities that collect and interpret it. The identity of a specific user is hence defined by the clustering of descriptive information collected about them. However, the meaning and scope of this data remain unknown to the user and, as a result, beyond their control (Cheney-Lippold, 2017).

The lack of clarity in the algorithmic collection and processing of users' information leads to what Kant (2020) describes as a state of "epistemic uncertainty" among users. An uncertainty that makes it exceedingly difficult for users to discern and assert their agency regarding their own data and identity (Kant, 2020; Milan, 2018). As a result, research has shown that individuals have an increasingly disoriented perception of their role in this process and a muddled understanding of what is their digital identity (Garbaşevschi, 2021).

Research on datafication processes also highlights the impactful and recursive nature of these processes, underscoring the importance of considering users' perspectives within these dynamics (Bucher, 2017). In these studies, users' understandings, imaginaries, theories, and emotional connections with technology are viewed as integral elements in constructing and defining online identities (Bishop and Kant, 2023; Bucher, 2017). These perspectives shed light on the productive entanglements between users, data, and algorithms in the construction of our digital identities. They also underscore how differing interpretations of algorithmic processes can lead to different choices and actions in daily life (Das, 2023; Siles, 2023).

Kant's analysis, for instance, emphasizes how "doing" identity online is inescapably tied to the ways users feel and negotiate algorithmic personalization, and is a reflection of how they learn "'to cope with', trusting, distrusting, being legitimized, or being disciplined by algorithmic personalization systems" (Kant, 2020: 87). Siles et al. (2023), on the other hand, bring forth the idea of "in-betweenness" in the realm of digital agency. This notion reveals how users are shown to simultaneously resist, conform to, challenge, and obey algorithms in their everyday lives, thereby asserting their identity and autonomy through different acts of resistance in the context of datafication.

By examining the diverse socio-technical dimensions of the relationship between digital technologies and identity, these views emphasize not only the growing importance of digital identity in our understanding of selfhood and subjectivity in the digital age (Cover, 2016) but also the impact these interactions have on our understanding of how identity is expressed online. Understanding how users interpret these dynamics and connect them to the concept of digital identity provides additional insights into the complex interplay between technology and identity.

Our research aims to shed light on the concept of digital identity from a user-centric viewpoint. We focus not on users' technical abilities or digital knowledge but on the range of understandings they have of their digital identity. In doing so, we aim to gain insights about how the concept shapes people's sense-making practices concerning the relationship between their selves and the internet in increasingly datafied societies.

Methods

This paper uses data collected during the online citizen dialogue "We, The Internet" conducted in Portugal in October 2020, which was organized in the context of a global consultation in more than 80 countries promoted by Missions Publiques, a French NGO, at the request of the High-Level Panel on Digital Cooperation convened by the UN Secretary-General. This consultation had a general goal regarding a discussion about the future of the Internet, using citizens' perceptions and understandings as insights to shape Internet governance. The general script covered five main sections: internet perceptions,

Table 1. Demographics ($N=86$).

Demographics	Count (n)
Gender	
Female	58
Male	28
Age (years)	
18–24	21
25–34	16
35–44	15
45–54	12
55–64	11
65–74	7
75+	3
Education	
Advanced	53
Intermediate	30
Basic	3

One participant did not share their age.

digital data, digital public sphere, artificial intelligence, and internet governance. This paper focuses on discussions related to digital data, and more specifically, on semi-structured discussions about digital identity.

In Portugal, the recruitment of participants predominantly occurred in the month leading up to the consultation. Although the sample was one of convenience (Stewart and Shamdasani, 1990), efforts were made to ensure diversity. Recruitment strategies included promoting the consultation on the social media platforms of participating institutions and through the personal accounts of the researchers. Additionally, direct outreach was conducted with various civil society associations and intermediaries. As an incentive for their participation, individuals were offered a 50€ gift card.

Prior to the event, all participants received informative materials about the event and topics being discussed. They consented and agreed with the protocol for the collection and dissemination of data within the context of this research project. Before the event, the organizing team arranged an optional online meeting for Q&A to address any doubts about the Zoom platform and the planned dynamics. This was a way to ensure that participants who were less familiar with online events could get comfortable with the online dynamics of the event beforehand.

Participants were asked to fill out a form indicating their gender, age, level of education, and occupation. The final sample consisted of 86 citizens spanning various age groups (ranging from 19 to 77 years), gender (28 male, 58 female), and educational background (the majority holding a graduate degree) (See Table 1). The sample was also diverse in terms of place of residence and professional situations. No information was collected regarding their ethnic self-identification, but the sample was predominantly composed of individuals of white ethnicity.

The event was conducted on Zoom and took place on two occasions in October 2020. It spanned a total of 5 hours, divided into two parts, each lasting 2 hours and 30 minutes. Participants initially gathered for an opening assembly but were then assigned to separate breakout rooms. In total, we had 17 focus groups, with each group consisting of between four and seven participants, following recommendations from the literature on this method (Barbour, 2007). Each group was led by a facilitator who guided the discussions. These facilitators were experienced researchers present to stimulate quieter participants, ensure that a few participants did not dominate the discussion, and create a safer environment in case of disagreements. Each of the five sections of the general script was discussed separately, with every participant engaging in all five discussions. Before each section, an organizing team member provided a brief introduction, highlighting important aspects of the topic. During the consultation, facilitators proposed several exercises to gather information reflecting the participants' opinions and perceptions. These discussions were recorded and later transcribed.

The semi-structured discussion on digital identity which is the focus of this article was included in the first part of the event with each of the 17 focus groups spending approximately 20 minutes on this topic (17 m × 20 m = 5 hours 40 minutes). In this discussion participants were intentionally prompted with a broad question: "What is your digital identity?". Facilitators encouraged active participation and the exchange of opinions, fostering engaging discussions in which different viewpoints on digital identity and related aspects emerged.

Our analysis, guided by the principles of focus group methods (Lunt and Livingstone, 1996), primarily aimed to uncover socially expressed and contested opinions and discourses on digital identity, rather than individual viewpoints. This approach prioritized the exploration of collective understandings rather than individual attitudes. We followed a constructivist grounded approach to analyze this data, applying codes to identify key themes and emerging issues (Charmaz, 2006). Using MaxQDA software, we searched for common patterns and dimensions commonly mentioned by participants during discussions. This initial phase resulted in the creation of a set of descriptive codes covering aspects such as privacy, public image, risks, authentication, multiplicity, and control. Initial codes were then compared, re-coded, and organized to clarify their relevance. This process led us to reorganize them into three main analytic codes, which appeared to exhibit sufficient conceptual autonomy and scope to capture the diverse viewpoints shared by participants (Charmaz, 2017). The codes were then further developed by the authors into three categories related to different understandings of digital identity through the detailed (re)examination of the data, clarifying their relations, and reflecting on their connections to the relevant literature.

In the following section, we detail our findings using relevant quotes from participants to illustrate the core themes addressed. These quotes are accompanied by the participant's gender and age group to provide context, with fictitious names assigned to enhance readability. The authors have translated and mildly edited these quotes for better clarity. The section begins with an exchange between participants, showcasing the variability in users' views on digital identity. Subsequently, we examine the three distinct categories identified in our study which we named: *Digital Identification*, *Self-Presentation*, and the *Datafied Self*.

Results

The discussion around the topic of digital identity elicited varied responses among participants. While some had not considered it deeply, others showed a clearer understanding. This diversity in conceptualization became evident during a notable exchange between two participants, Carlos and Maria, who held contrasting views:

- Carlos: The concept of digital identity is a bit difficult. It seems to me to be the set of our personal data we use when we register on any platform. . . (. . .) because it is not exactly the data generated from our searches, as someone said before. . . that is data, it is not identity. [male_75+]
- Maria: In my opinion, identity is more than that. Our identity is not only the data that we put there, but also our interests, our sources of interests, reflected in what we search, because when I introduce myself to someone I don't talk about my tax card number or my identity card number. The way I identify myself to someone has to do with my interests, what I'm interested in. So, I think identity is more than that. [female_35-44]

For Carlos, digital identity is something very specific – it is the identification data created automatically when we register as users on a platform and that is used to access our accounts. On the other hand, Maria views our digital identity as intertwined with the digital traces we leave behind, which combined reveal a larger story about who we are. This example highlights the significant differences in how users make sense of the concept of digital identity but also underscores how identity and data have become intertwined in digital societies.

Digital identity as digital identification

The first understanding of digital identity we identified relates closely to the idea of digital identification. Identity, in this case, is associated with the data we use to access a service or platform and as something that identifies us online as unique individuals.

It is the data that allows me or any other person to access a certain information platform, whatever it may be. (. . .) To use an email I must have an identity, to access a bank the identity is the same or different, each one must define that situation (. . .) After all, that is my digital identity. [Bruno_male_65-74]

A username, an email, a password, and an official signature were some of the examples mentioned by the participants. Digital identity is linked to a very specific type of information, namely the information used to identify and authenticate ourselves online. In this sense, it aligns more closely with the legal definition of digital identity (Lessig, 1999; Sullivan, 2018). Ana, for example, associated digital identity with the Mobile Digital Key, a method of authentication created by the Portuguese State to access to public e-services:

Last week I asked for some prescriptions at my health center. And the doctor texted me to tell me to install My SNS [National Health Service] app. I opened it and it asked me for my digital

identity, but I don't have one. To create my digital identity, I have to go to a "citizen store". . . it has a series of procedures. (. . .) So I didn't even download the app. I went to the doctor to get the prescription in paper form and that was it. [Ana_female_65-74]

The idea that this identity is created not by choice, but by necessity, was also mentioned by other participants. Duarte thinks that digital identity is the data used to sign in for something, which is mostly dependent on the requirements of the websites and applications we use.

For example, on a website I will register in a certain way. On that site I just want to buy something, I will only put the necessary data. If it is something more official, I will have to put another type of data. So, I think that digital identity is a little bit what we write and that it depends on the sites and applications we use. [Duarte_male_25-34]

This idea was widely shared and most agreed that it is almost impossible not to have a digital identity in a world that increasingly creates incentives and obligations to use online services. As the digitalization of different spheres of our lives grows, the number of digital identities also increases (Masiero, 2023). Several participants mention that we have not one, but several identities associated with different apps and contexts.

We often have several digital identities. When I register on a website or in a company or regarding. . . the COVID application or whatever, I get that digital identity, but then I seem to have other digital identities. When I access through my login at my workplace, I'm obviously related right away to my work network, in a health institution the same thing. I think that will be the digital identity. [Élio_male_45-54]

For certain participants, having multiple identities is perceived as presenting various risks, such as having to manage too many passwords and the apprehension of losing or having sensitive information stolen. For others, however, it is seen as something positive. As Francisco explained, having several identities online means that our personal information and data are more protected, since one is less exposed to manipulations and hacks.

There is a discussion about whether there should be only one access or not. I am in favor of having several accesses. I am not in favor of a single number, one identity only, for everything. I think there should be several identities because citizens will be better protected if they are manipulated or attacked anywhere. [Francisco_male_65-74]

The understanding of digital identity as digital identification clearly associates it with personal data that functions online as an identifier of the offline self and a set of unique identifiers that we use to authenticate ourselves online.

Digital identity as self-presentation

The second understanding of digital identity is closer to what the literature defines as self-presentation. In this case, digital identity is seen as something imminently constructed online, something performed to an audience (boyd, 2010).

Our digital identity is, for example, the image that we want other people to perceive of us. We sell our image as a product in relationships, or social networks. (. . .) For example, LinkedIn, where our work information is. [Carolina_female_25–34]

Since it is a projected image, this identity does not have to correspond necessarily to our offline self. As some participants explained, “it can be distorted” [Rosa_female_55_64], “manipulated” [Inês_female_n/a], an “escape from reality” [Marta_female_25–34]. Dalila, in turn, considered that the way we present ourselves online reflects only partially who we think we are, resulting in an online identity that differs from the image we have about ourselves.

I understand that we may have a digital identity that may be distinct from what we consider to be our true identity as a whole, apart from the digital world. And an image can be created about us that diverges from what we consider to be us. The way we present ourselves in certain social networks will always be a slice and never the totality of what we consider ourselves to be. [Dalila_female_55–64]

The common thread associated with this understanding is that we have more control over our online identity than we have over our offline identity. As one participant explained, by choosing to post certain things and presenting information in a certain way, we construct an idealized image that mostly projects what we want others to think of us. Additionally, this identity can be plural since we tend to share different things on different platforms.

I find it a bit difficult to give a definition, but I think it’s basically our presence on the internet, that is how we are on different platforms. How are we on Facebook, how are we on LinkedIn, which is a more professional application. . . . [Hilário_male_55–64]

On the other hand, several participants were keen to explain that their online identity had a clear correspondence with their offline self, assuming a closer notion of the anchored self (Szulc, 2019). One participant stated that she is the same person on all social media. Cristina [female_45–54] said that we should not pretend to be someone else because our digital identity is still who we are. This is the result of a conscious effort as well as concerns about how people tend to present themselves online. Estela shared a similar opinion:

I had never thought about digital identity. But, anyway, my digital identity is my real identity. I don’t invent anything at all about myself. I don’t create a virtual identity that doesn’t coincide. My identity is my identity. [Estela_female_45–54]

Here, digital identity corresponds to the offline self as a specific form of self-presentation, fostered by the desire to control one’s image and privacy. It is anchored in the idea that, ultimately, we must be responsible for how we present ourselves online and manage our online identity. This theme emerged several times during the discussions, particularly concerning users’ control over the privacy levels of their digital identity. One participant highlighted that people have the choice to make their accounts private, restricting who can have access to what they post. Another referred to the possibility of having a layered identity, with different levels of privacy defined by the settings we choose to use.

When identity is thought of in terms of self-presentation, the need to manage this public identity comes to the front stage. Participants explained that “things on the internet” are permanent and searchable (boyd, 2010), so, when we choose to post something, we are necessarily exposing ourselves to others. This can be tricky, especially because of context collapse (Marwick and Boyd, 2011). Several participants talked about the risk of employers, or recruiting companies, searching online for information about individuals and ending up with an image that does not correspond to reality. Irina also raised the question of how we can lose control over our online image in the long term, and how what we share can be differently interpreted and used later against us:

Anything we publish on the Internet is there forever, whether we like it or not. Something that we see a lot now is people who manage to get things that people published many years ago and (. . .) turning it against them and ruining many careers (. . .) sometimes people publish things without thinking about the consequences they may have in the future. [Irina_female_25–34]

Finally, a few participants also mentioned the way we are encouraged to share certain data and to project a certain image online poses risks not only to our privacy but also to our well-being. They consider this is mainly due to the risks of disconnection between online and offline selves that can have terrible consequences for our identity construction and mental health.

I think that has a lot to do with the image that we have of ourselves and the image that we want to put across. And that can lead to a dichotomy of identity issues and health issues. More often we see that digital identity is not entirely trustworthy and sometimes it can be related to mental health issues. [Carolina_female_25–34]

Digital identity as datafied self

Digital identity is also seen as something not created by us but by digital platforms or algorithms, based on data often extracted and collected without our permission or knowledge (Cheney-Lippold, 2017). For these participants, digital identity is composed of everything we do online, every search, every place we go, and anything we share online.

All our information, all our digital history. The websites we visited, the information, the content we shared. [Manuel_male_45–54]

Identity is clearly associated with data, but all kinds of data: “the collection of all our data,” a “database of everything that exists about us,” “the cross between everything I’ve been doing on the internet,” participants explained. As Nuno put it, digital identity is something built from all the data we voluntarily and involuntarily leave online.

I think it is a set of all the data that is collected . . . and what profile is built about me from this data. From what I search on the internet, you can find out what else I need. He is a student, he is so many years old, his consumption pattern is such, that is, it is something that is built about me from the data that I, although involuntarily, not unconsciously, hand over when I access a website. [Nuno_male_18–24]

This collection of data is made of everything we actively do and share online, but not only. Some participants highlighted that data is collected even when we are not actively on the internet, especially through our smartphones. They gave the example of platforms being able to trace all our movements through the location of our mobile or capturing sound and information without our consent. Laura, for example, mentioned the possibility of data-capturing conversations we have on the phone.

It's all our activities, it's our searches, it's even what we don't put directly on the internet, when we are talking, like in the case of the colleague who was on the phone [Laura_female_25–34]

According to some participants, all this data is captured, registered in databases by companies, and associated with an online profile, a “persona on the internet,” said Vasco [male_18–24] which we involuntarily create. Madalena [female_75+] referred to the existence of multiple identities, “one as a wardrobe consumer, another as a Facebook consumer, and another as an Instagram consumer.” Paulo mentioned the same idea, indicating that for digital companies it is still difficult to combine all the collected data, something that he considered to be a possibility in the future.

The truth is that we have several identities and depending on the platform we use to work or to talk, to publish, to do things, to interact. . . we have several identities as a whole and the danger, is the crossing of all these areas, crossing everything that I have been doing on the internet in the last ten years, fifteen years, twenty years. (. . .) Fortunately, it is still extremely difficult to combine all this data, but we are increasingly moving towards this situation. [Paulo_male_55–64]

Other participants said these profiles are created for commercial purposes by the digital platforms we use as Google or Facebook. Some were more abstract without naming a specific company or platform: it's “the machine,” “the algorithms,” “a kind of artificial intelligence,” the “system,” “the algorithm trained for that.” All these processes result in a digital identity we no longer recognize. Teresa [female_45–54] referred to it as “an identity built as a consumer,” Xavier [male_35–44] as “a product to be commercialized,” and Olga as something robotic:

Because it's more and more from the owners of social media. Less and less is about our identity. More and more it's really an algorithm or a kind of artificial intelligence with our data, which is no longer our essence. It can already be a little bit robotic. [Olga_female_18–24]

However, most of these processes tend to be not visible to users. Like in Bucher's (2017) analysis, participants become aware of the existence of a datafied self only when confronted with personalized content on the web. Diana [female_18–24] explained that companies tend to bombard her with information about things she searched for, recognizing “that's part of my profile, I'm a person who supposedly likes tennis.” Patricia [female_65–74] gave a similar example of searching for a hotel in a specific location and then receiving similar suggestions. She said that when that happens, she knows “the algorithm is working”.

Furthermore, this “amalgamation” (a term used by one of the participants) is often perceived as “epistemically unknowable” (Kant, 2020). Its relationship to the offline self is extremely ambiguous and the clues given by the algorithm through personalization are considered confusing. Some participants mentioned that they often do not identify with what is suggested by the platform and do not understand how their data is being interpreted. For these participants, identity is built by the algorithm but often has little to do with them. It is an imperfect, selective, and commodified product (Garbaşevschi, 2021), assuming things about people as consumers, frequently misinterpreting what we do online and making wrong assumptions about likes and needs (Bishop and Kant, 2023).

In contrast, others argued that our digital identity can also be a less performed representation of who we are. Tomás [male_25–34] explained that because social media companies collect information about things we may not even be aware of “they can get a clearer idea of the person” – clearer than the image we have of ourselves or that others have of us. Similarly, Maria noted that Google knows more about us than we do ourselves because they can cross-reference information on everything we do:

We think we are one thing and then we are another thing to those who are running the internet. And they know more about us than we and those around us do. For example, when we have a tummy ache, we’re not going to tell everyone, obviously, but we might google - how to avoid diarrhea, right? We don’t realize that this is going to be registered in our digital identity because for us it’s a simple search, but whoever is managing from behind has broader access to everything we do, on Facebook, on Twitter, and this is all crossed. There are algorithms that can cross-reference all this, and perhaps they have a very different view of what we have of our own digital identity. [Maria_female_35–44]

Because of platforms’ data collection practices, our digital identities are seen as something we do not completely control. For Beatriz, for example, it’s the interconnectedness between our personal accounts, or identifiers, and the information we share online that shapes our (more or less) public digital identities. These, in turn, are seen as inescapably entangled with our datafied selves.

I believe that our digital identity is not entirely under our control. I think it’s not just the data we share online, because the data we provide is essentially being captured by companies with a commercial intent (. . .). We provide some initial inputs with our data, clicks, and choices, but then we are also influenced because the system and algorithm are trained for that. There is also another aspect, related to the services we have at the State level, in terms of having an identification number (. . .), but these don’t represent such a large portion of our digital identity. In other words, we partially contribute to the construction of our digital identity, but there’s another layer built on top of it through the stimuli that come from the digital world to us. [Beatriz_female_35–44]

During the discussions, participants also highlighted several risks associated with their digital identity and the rise of datafication processes. Increased personalization is seen as problematic because it tends to keep us inside our bubbles and condition what we do and think. One participant referred to this as being a growing problem since we are increasingly heading for data crossing between platforms, long-term accumulation of

information, and personalized conditioning associated with our online identities (Cheney-Lippold, 2017).

The feeling and fear of being constantly watched at any time also transpired (Zuboff, 2019). Bernardo [male_45–54] expressed apprehension regarding the utilization of personal data by private companies, particularly the ability to track our location at any given time. Other participants lamented the lack of regulation, transparency, and the escalation of these surveillance practices. Notably, Rui [male_65–74] emphasized the belief that privacy is a human right, expressing concerns about deregulated and potentially dangerous data collection practices, such as the collection of biometric information without consent.

Discussion

The historical conceptualizations of digital identity previously discussed offer a useful backdrop for the three identified understandings of the concept. *Digital identification* was mostly conceived as something that identifies us as unique users online, largely required to access services or platforms, similar to the identification systems discussed by Lessig (1999) and Sullivan (2018) and more broadly in the Information Systems literature.

The understanding of digital identity as *self-presentation*, on the other hand, aligns more closely with the perspectives that have focused on online interactions and identity construction in virtual spaces (Marwick, 2013; Papacharissi, 2010). *Self-presentation* was mainly associated with what we choose to do on these platforms and the data we share publicly, which are seen as searchable and permanent (Boyd, 2010). For some, this digital identity can have a dissimilar presentation from the offline self. For others, it should be singular and coherent, aligning more closely to the notion of the anchored self (Szulc, 2019).

Finally, the *datafied self* represents a more recent conceptualization of digital identity. Predominantly seen as an inevitable consequence of our internet use and interaction, it refers to an unknowable data entity constructed by digital platforms and fed by the many traces we leave behind. In this view, identity is seen mostly as a construct of algorithmic processes and data collection, as discussed by Cheney-Lippold (2017) and Kant (2020).

These three understandings of digital identity are not always mutually exclusive, often intertwining in our participants' discussions. For instance, when referring to our public profiles participants noted that our posts can be trackable through metadata, reflecting an implicit acknowledgment of overlapping. In other cases, the complexity of digital identity was explicitly recognized, with participants recognizing its dual or even triple nature. They perceived digital identity as multiple, not solely due to different online accounts or the varied images projected to different audiences (Buss et al., 2022), nor just because of personal data being subjected to algorithmic (re)interpretations (Kant, 2020). Rather, they saw it as inherently encompassing various aspects of their online presence.

This analysis further underscores the pervasive nature of digital platforms in our daily lives, rendering the online-offline dichotomy increasingly restrictive (Cover, 2016). Nonetheless, it also highlights how different meanings attributed to digital

identity mediate people's perceptions of this relationship (Søe and Mai, 2022). In the case of *digital identification*, virtual identity is mostly seen as an identifier of the offline self. In the case of *self-presentation*, the self is seen as performed and without necessary correspondence with the offline self. Finally, in the *datafied self*, the relation to the offline self is extremely ambiguous, being imagined both as a fraction of what we are and/or as a manifestation of a truer us.

Moreover, these understandings also relate to different perceptions users have of their agency and identity construction as well as the sociotechnical environments shaping it (Milan, 2018; Siles, 2023). When discussing identity in terms of *self-presentation*, agency was mostly ascribed to the user, who was seen as the one responsible for shaping their online persona. On the other hand, *digital identification* was mostly viewed as an unavoidable consequence of being a citizen in an increasingly digitalized world. Explicit concerns about the lack of user agency were particularly emphasized when considering the *datafied self*, where issues like data extraction, consent, and privacy acquired salience.

Ultimately, our analysis reveals that digital identity is, foremost, a complex and multi-layered concept, one that is becoming increasingly harder for users to define. This is so because it is understood as something that we cannot completely know and control (Kant, 2020). It is not a singular entity; rather it has layers built into it, resulting not only from what we do but also from the socio-technical infrastructures we rely on. This underscores the intricate dynamics at play when discussing our digital identities, and emphasizes the complex interactions among users, identities, personal data and digital platforms in datafied societies.

Conclusion

In this article, we draw upon research in social media and datafication processes to explore the significance of digital identity in datafied societies and users' understandings of this concept. Our research extends the extant exploration of digital identity by centering on users' viewpoints regarding the concept and its underlying justifications. The analysis of 17 online focus groups revealed three understandings of digital identity: *digital identification*, *self-presentation*, and the *datafied self*, each encompassing specific considerations about users' online presence. Last, we showed these understandings are not always mutually exclusive, as they sometimes appear intertwined. It became evident that, for some participants, identification systems, online self-presentation, and data collected by commercial platforms are seen as inherently interconnected, converging to create layers in what we might define as our digital identity.

These findings underscore insights from previous research that have shown how making sense of the notion of digital identity has grown increasingly complex for users (Garbajevschi, 2021). Various factors, such as the widespread commercial-driven personal data tracking (Cheney-Lippold, 2017; Szulc, 2019), expanded algorithmic processing and personalization (Bucher, 2017; Kant, 2020), the proliferation of digital identity schemes (Sullivan, 2018), and platform-mediated surveillance (Masiero, 2023), among others, contribute to this continuous reconceptualization of the term. As the digital entanglement of

our lives intensifies and evolves, perceptions of the internet expand, inevitably influencing and reshaping users' understanding of what their digital identity entails.

However, our study has also highlighted that these dynamics do not necessarily produce similar viewpoints. While some participants acknowledged the layered nature of their digital identity, others held specific understandings of the term, often tied to specific facets of their online presence. By integrating historical conceptualizations with empirical findings, our analysis underscores how the concept's ever-expanding complexity and multidimensionality provide opportunities for its continuous reinterpretation. This allows for the emergence, coexistence, and even competition of diverse conceptualizations of digital identity among users. Additionally, the evolving nature of the concept and the coexistence of different understandings have significant implications for how user agency is conceptualized, influencing perceptions of control, responsibility, and autonomy in the construction and management of digital identities.

While our study offers valuable insights, it is important to acknowledge its limitations, particularly regarding the composition of our sample and the context of data collection. As Davis (2014) notes, meanings of identity are inherently context-dependent, and our discussions on digital identity occurred within a broader event focused on data governance. This contextual backdrop may have influenced participants, especially in their reflections related to the datafied self.

Future studies could examine users' understanding of their digital identity, by considering individual variables such as socio-demographic backgrounds, patterns of internet use, and varying levels of digital literacy. Additionally, it would be important to focus on the perspectives of minority groups, whose online experiences and identity expressions are often uniquely shaped by distinct cultural, social, and political dynamics. Furthermore, it would be valuable for researchers to explore how users' understandings of digital identity vary across different national, political, and socio-cultural contexts.

Finally, beyond identifying who holds specific views on digital identity, it is also crucial to understand the broader implications of its multifaceted nature. Such exploration could be extended to how digital identity meanings are mobilized in everyday scenarios, including their application by governments and commercial platforms in user interactions, and how this, ultimately, influences perceptions about the role that digital identities play in datafied societies.

Acknowledgements

We would like to thank our coordinator, Ana Delicado, whose guidance and committed efforts made the consultation and this article possible, as well as Ana Jorge for her encouragement and valuable suggestions. We also appreciate Missions Publiques for organizing the international citizen dialogue, "We, the Internet," and all the facilitators who guided the Portuguese group discussions. A special thanks goes to the 86 participants who generously shared their perspectives on the topic. Finally, we thank the reviewers for their kind and insightful comments on the paper.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The online citizen dialogue 'We, The Internet' was financially supported by Missions Publiques and the Institute of Social Science (University of Lisbon).

ORCID iD

Jussara Rowland  <https://orcid.org/0000-0003-2639-3714>

References

- Barbour R (2007) *Doing Focus Groups*. London: Sage Publications.
- Bishop S and Kant T (2023) Algorithmic autobiographies and fictions: A digital method. *The Sociological Review* 71(5): 1012–1036.
- Boyd D (2010) Social network sites and networked publics: Affordances, dynamics and implications. In Papacharissi Z (ed.) *A Networked Self: Identity, Community, and Culture on Social Network Sites*. New York: Routledge, pp.39–58.
- Bucher T (2017) The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information Communication and Society* 20(1): 30–44.
- Burrell J and Fourcade M (2021) The society of algorithms. *Annual Review of Sociology* 47: 213–237.
- Buss J, Le H and Haimson OL (2022) Transgender identity management across social media platforms. *Media, Culture & Society* 44(1): 22–38.
- Camp JL (2004) Digital identity. *IEEE Technology and Society Magazine* 23(3): 34–41.
- Castells M (2007) Communication, power and counter-power in the network society. *International Journal of Communication* 1(1): 238–266.
- Charmaz K (2006) *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London: Sage Publications.
- Charmaz K (2017) Special invited paper: Continuities, contradictions, and critical inquiry in grounded theory. *International Journal of Qualitative Methods* 16(1): 509–524.
- Cheney-Lippold J (2011) A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society* 28(6): 164–181.
- Cheney-Lippold J (2017) *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.
- Coulmas F (2019) *Identity: A Very Short Introduction*. Oxford: Oxford University Press.
- Cover R (2016) *Digital Identities: Creating and Communicating the Online Self*. London: Academic Press.
- Das R (2023) Parents' understandings of social media algorithms in children's lives in England: Misunderstandings, parked understandings, transactional understandings and proactive understandings amidst datafication. *Journal of Children and Media* 17(14): 1–17.
- Davis JL (2014) Triangulating the self: Identity processes in a connected era. *Symbolic Interaction* 37(4): 500–523.
- Dyer HT and Abidin C (2022) Understanding identity and platform cultures. In: Housley W, Edwards A, Beneito-Montagut R, et al. (eds) *The SAGE Handbook of Digital Society*. London: Sage, pp.170–187.
- Feher K (2019) Digital identity and the online self: Footprint strategies – An exploratory and comparative research study. *Journal of Information Science* 47(2): 192–205.
- Garbaşevschi D (2021) *Infoselves: The Value of Online Identity*. New Jersey: John Wiley & Sons.
- Han S (2019) Virtual identities: From decentered to distributed selves. In: Elliott A (ed.) *Routledge Handbook of Identity Studies*. New York/London: Routledge, pp.219–236.
- Kant T (2020) *Making it Personal: Algorithmic Personalization, Identity, and Everyday Life*. Oxford: Oxford University Press.
- Kavakci E and Kraepflin CR (2017) Religious beings in fashionable bodies: The online identity construction of hijabi social media personalities. *Media, Culture & Society* 39(6): 850–868
- Krotoski A and Hammersley B (2015) Identity and agency. In: *The International Encyclopedia of Digital Communication and Society*. Hoboken, NJ: John Wiley & Sons, pp.1–9.

- Lareki A, Altuna J and Martínez-de-Morentin J-I (2023) Fake digital identity and cyberbullying. *Media, Culture & Society* 45(2): 338–353.
- Lessig L (1999) *Code: And Other Laws of Cyberspace*. New York: Basic Books.
- Liu Y (2024) “I want to be a bridge”? The digital identity positioning of transnational bloggers on Chinese social media platforms: Between ethnic differences and cultural affinities. *Media, Culture & Society* 46(1): 130–147.
- Lunt P and Livingstone S (1996) Rethinking the focus group in media and communications research. *Journal of Communication* 46(2): 79–98.
- Marwick A (2013) Online identity. In: Hartley J, Burgess J and Bruns A (eds) *A Companion to New Media Dynamics*. Malden, MA: Blackwell Publishing Ltd, pp.355–364.
- Marwick A and Boyd D (2011) I Tweet honestly, I Tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13(1): 114–133.
- Masiero S (2023) Digital identity as platform-mediated surveillance. *Big Data & Society* 10(1). <https://doi.org/10.1177/20539517221135176>
- Milan S (2018) Political agency, digital traces, and bottom-up data practices. *International Journal of Communication* (12): 507–525.
- Nakamura L (2013) *Cybertypes: Race, Ethnicity, and Identity on the Internet*. New York: Routledge.
- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Papacharissi Z (ed.) (2010) *A Networked Self: Identity, Community, and Culture on Social Network Sites*. New York: Routledge.
- Schüll ND (2019) Self in the loop: Bits, patterns, and pathways in the quantified self. In: Papacharissi Z (ed.) *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*. New York: Routledge, pp.25–38.
- Siles I (2023) *Living With Algorithms: Agency and User Culture in Costa Rica*. Cambridge, MA: MIT Press.
- Siles I, Gómez-Cruz E and Ricaurte P (2023) Toward a popular theory of algorithms. *Popular Communication*, 21(1), 57–70
- Søe SO and Mai JE (2022) Data identity: Privacy and the construction of self. *Synthese* 200(6): 492.
- Stewart DW and Shamdasani PN (1990) *Focus Groups: Theory and Practice*. Applied Social Research Methods Series. Newbury Park, CA: Sage Publications.
- Sullivan C (2018) Digital identity – From emergent legal concept to new reality. *Computer Law & Security Review* 34(4): 723–731.
- Szulc L (2019) Profiles, identities, data: Making abundant and anchored selves in a platform society. *Communication Theory* 29(3): 169–188.
- Turkle S (1995) *Life on the Screen: Identity in the Age of the Internet*. New York: Touchstone.
- Turkle S (1999) Cyberspace and identity. *Contemporary Sociology* 28(6): 643–648.
- van Dijck J (2013) ‘You have one identity’: Performing the self on Facebook and LinkedIn’. *Media, Culture & Society* 35(2): 199–215.
- van Zoonen L (2013) From identity to identification: Fixating the fragmented self. *Media, Culture & Society* 35(1): 44–51.
- Zhao S (2005) The digital self: Through the looking glass of telecopresent others. *Symbolic Interaction* 28(3): 387–405.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.