Italian Sociological Review
ISSN 2239-8589
DOI: 10.13136/isr.v16i16S.973 2026, 16(16S), pp. 413-438

Post-Digital Data-Gathering and the Adaptive Epistemological Framework: Navigating the Human-Algorithm-Platform Nexus*

Suania Acampa[a], Giuseppe Michele Padricelli[b], Gabriella Punziano[c]

Abstract

In the post-digital condition, where the digital is no longer a separate domain but the pervasive environment of social life, this article addresses the methodological and epistemological challenges of studying human-algorithm-platform interactions. Digital platforms are understood as socio-technical systems that mediate user behavior through opaque algorithmic mechanisms, producing data as co-constructed artifacts rather than neutral traces. The paper proposes an adaptive epistemological framework that responds to digital data's hybrid and processual nature, emphasizing the need for flexible, plural, and reflexive research designs. The work conceptualizes data hybridization as a methodological paradigm capable of capturing active engagement and passive traces through a comparative analysis of digital and computational ethnography, web scraping, APIs, and data donation. The discussion culminates in a typological framework that systematizes data-gathering techniques according to user awareness and researcher intervention, offering practical and theoretical guidance for navigating an increasingly algorithmic and datafied research landscape.

Keywords: digital data-gathering, adaptive epistemology, digital and computational ethnograpy, data scraping, data donation.

[a] University of Naples Federico II, Naples, Italy.
[b] University of Naples Federico II, Naples, Italy.
[c] University of Naples Federico II, Naples, Italy.
* The work presented is the combined result of the reflections of the three authors. Specifically, Introduction, paragraphs 1, 2 and 4.2 should be attributed to Suania Acampa; paragraphs 3, 5 should be attributed to Gabriella Punziano; paragraphs 4, 4.1, 4.3 should be attributed to Giuseppe Michele Padricelli.

Corresponding author:
Giuseppe Michele Padricelli
E-mail: giuseppemichele.padricelli@unina.it

Received: 23 May 2025
Accepted: 15 July 2025
Published: 12 May 2026

1. Introduction: pervasive logics and processual data

Digital platforms have become indispensable elements of daily life, reshaping how individuals communicate, work, access information, and engage in the public sphere. Though these tools appear intuitive, they operate within complex socio-technical systems governed by opaque algorithmic logics that influence content visibility and modes of interaction.

Platforms are not neutral infrastructures; they are dynamic environments where computational models and technological affordances intertwined (Gillespie, 2014). Algorithms mediate access to information, prioritize content and define the structure of digital ecosystem (Pasquale, 2015; Noble, 2018). These opaque mechanisms raise concerns about user autonomy, informational diversity, and the broader implications of algorithmic governance (Zuboff, 2019).

Every data flow within these ecosystems is processed through predictive mechanisms that shape what is seen, when, and by whom (Kitchin, 2017). This dynamic are central to the broader logic of surveillance capitalism (Zuboff, 2019), in which user action (from a Google search to a Facebook like) are transformed behavioral data that fuels platform optimization strategies (Introna, 2016; Bishop, 2019) The algorithmic infrastructure aims to maximize engagement and profitability, often reinforcing selective visibility and reducing user agency (Beer, 2017a).

The influence of algorithms extends beyond content organization to information regulation, as platforms exercising an editorial power defining what is highlighted and what remains hidden without users being fully aware (Bucher, 2018; Noble, 2018). Through mechanisms digital platforms reinforcing existing power structures and shaping public discourse (Gillespie, 2014) with socio-political implications, influencing knowledge production, media representation, and collective decision-making processes (Vaidhyanathan, 2018).

In this hybrid digital environment, users interact not only with content or interfaces, but with what we call the *platform-algorithm unicum:* a socio-technical subject that dynamically engages with and reconfigures user behavior. Understanding this relationship requires us to explore *types* and *modalities* of interaction within these environments, the implications in terms of *algorithmic perception, algorithmic awareness, and the mediation of consumption practices*, ranging from informational to cultural and beyond.

This article explores these dimensions, analyzing algorithmic mediation's *practical and factual repercussions* and the *methodological challenges* that arise in studying these processes. A key issue in this regard is the *collection of processual data*, data that is not static but emerges through interaction, possessing an inherently hybrid and dynamic nature. The very act of generating data in these digital

environments cannot be seen as a neutral or independent event; rather, it is a process of *co-construction* between user activity, platform affordances, and algorithmic decision-making (Marres, 2017; Couldry & Mejias, 2019).

The methodological implications of this perspective extend beyond conventional data-gathering techniques. The evolving landscape of digital research necessitates new tools and methodologies capable of capturing these interactions' fluid, adaptive, and recursive nature while also *addressing the technical, ethical, and regulatory challenges* associated with accessing and analyzing such data (Dencik et al., 2019). As computational action becomes increasingly pervasive - amplified by the intervention of generative artificial intelligence and sophisticated algorithmic infrastructures - the need for critically informed, interdisciplinary approaches becomes ever more pressing. This entails addressing technical limitations and ethical concerns regarding data transparency, bias, and user autonomy (Zuboff, 2019; Mittelstadt et al., 2016).

This scenario is part of a condition that can be defined as *post-digital* (Cramer, 2014; Marres, 2017), in which digital no longer represents a separate sphere, but constitutes the ubiquitous environment in which social practices and research activities take place. In this context, the study of algorithmic governance cannot remain confined to a static analysis of pre-existing structures. Instead, it must be framed within an *adaptive epistemological framework* that acknowledges the continuous evolution of platform logic and the emergent strategies users develop in response (Cotter, 2019; Eubanks, 2018). This shift in perspective opens new avenues for *rethinking data collection and retrieval strategies* in an era increasingly shaped by computational agency and artificial intelligence, where the boundaries between human decision-making and algorithmic intervention become increasingly blurred.

2. The platform-algorithm as an interactive subject: implications for the individual-algorithm relationship and its analysis

The relationship between users, platforms, and algorithms constitutes a dynamic system in which the continuous interaction between technology and human agency gives rise to a process of data co-production. From this perspective, digital platforms can be understood as sociotechnical systems where technological and human components mutually influence each other, generating situated and contextual effects (Marres, 2017). However, this interaction is asymmetrical: while users navigate within constraints defined by affordances and algorithmic architectures, platforms possess advanced tools to shape behaviors in ways that remain invisible to their users (Gillespie, 2014; Eubanks, 2018). In doing so, platforms redefine the boundaries of the digital

experience and establish what is considered relevant within the informational ecosystem (Couldry & Mejias, 2019).

This convergence of platforms and algorithms creates a singular sociotechnical entity, what we call *platform-algorithm unicum*: a hybrid construct that operates as an interactive subject, both shaping and being shaped by user engagement. Unlike traditional media, algorithmic platforms dynamically structure visibility, and relevance through opaque mechanisms (Gillespie, 2014), producing emergent forms of agency, power, and perception that redefine digital subjectivity (Bucher, 2018).

This is particularly evident in recommendation systems on entertainment content platforms such as YouTube and Netflix, where algorithmic strategies influence what users consume (Seaver, 2019). Drawing from Foucault (1977), platforms function as instruments of power, capable of disciplining user behavior without the need for explicit obligations. Features like gamification, badges, and interaction metrics act as behavioral regulation tools (Bishop, 2019).

Recommendation mechanisms play a central role in shaping cultural consumption practices, influencing access to music, videos, and entertainment through algorithmic personalization (Airoldi, 2021) and contributing to the proliferation of a culture of algorithmic optimization (Napoli, 2019). The result is that a new algorithmic perception emerges from this particular dynamic of digital reality construction. The perception of algorithmic agency - how users interpret and internalize the presence and function of algorithmic decision-making - plays a crucial role in shaping interactions within digital environments (Eslami et al., 2016).

Users do not passively consume content but navigate algorithmically curated spaces with varying level of awareness and interpretive frameworks (Rader & Gray, 2015). Some users develop *folk theories* about how algorithms work attempting to optimize visibility or engagement (Cotter, 2019), while others remain unaware or indifferent to the algorithmic structuring of their experience (Couldry & Mejias, 2019). Algorithmic perception is thus not uniform but is shaped by social, cultural, and cognitive factors (Felaco, 2022). For instance, content creators often exhibit a heightened awareness of algorithmic mechanics, engaging in *platform literacy* practices to manipulate ranking and engagement metrics (Bishop, 2019). In contrast, casual users may demonstrate *algorithmic blindness*, failing to recognize the curated nature of their feed and assuming an organic or neutral information flow (Eubanks, 2018).

Awareness of algorithmic governance does not automatically translate into agency. While some users engage in *algorithmic resistance*—adjusting their behavior to counteract perceived biases or manipulations (Noble, 2018)—many remain caught in feedback loops of engagement optimization designed to

maximize time spent on the platform (Zuboff, 2019). The platform algorithm unicum thus generates a paradox of control (Zuboff, 2019): users are offered a sense of personalization, while their behavior is subtly directed through nudges, affordances, and reinforcement mechanism

This form of algorithmic governance fosters a condition of *algorithmic habituation*, whereby users unconsciously align their behaviors with platform incentives without being explicitly aware of the underlying computational logic (Gekker & Hind, 2019). Social media, for example, leverage engagement-driven metrics systems to prioritize emotional charged content reinforcing echo chambers and affective polarization (Tufekci, 2018). Similary, Netflix and Spotify exemplify this phenomenon, where algorithmic curation creates self-reinforcing consumption patterns that privilege high-engagement content over diversity and novelty (Lobato, 2019; Eriksson et al., 2019).

Algorithms curation also extends to news consumption, where feeds prioritize virality rather than informational rigor, encouraging sensationalism, polarization and fragmentation of public discourse (Cadwalladr & Graham-Harrison, 2018). The interplay between algorithmic sorting and user interaction thus establishes a new mode of public engagement, one driven less by editorial judgment and more by the opaque optimization processes of the *platform-algorithm unicum* (Diakopoulos, 2019).

The *platform-algorithm unicum* thus represents an interactive subject that shapes, directs, and co-produces digital experiences. Its influence extends beyond content organization to the structure of perception, agency, and consumption within digital ecosystems. A key challenge lies in fostering greater algorithmic literacy and transparency, ensuring that individuals understand how platforms function and how their behaviors are being modulated within these algorithmic environments. Moreover, the increasing opacity of algorithmic systems calls for renewed interdisciplinary inquiry to investigate their socio-technical impact. The pervasiveness of algorithmic systems and the opacity surrounding their operational mechanisms make it increasingly urgent to reflect on the relationship between users, platforms, and algorithms while simultaneously questioning the power structures that govern the digital public sphere. This scenario raises new methodological challenges: on the one hand, there is a need to *rethink approaches and tools* capable of investigating algorithmic mediation processes; on the other, it is crucial to *capture the dynamic nature of the relationship between users and platforms*, as well as the *process of data co-production* that emerges from this interaction (Gillespie, 2014; Beer, 2017a).

The growing algorithmic mediation of digital environments necessitates methodological frameworks that account for both the structural constraints imposed by platforms and the agency of users in negotiating these constraints (Seaver, 2017). This requires interdisciplinary approaches that *integrate*

computational analysis, qualitative inquiry, and critical theory to examine how algorithmic systems shape visibility, participation, and discourse in online spaces (Kitchin, 2017; Bucher, 2018), as well as the consumer's practices in the digital environment (Airoldi, 2021). Furthermore, the dynamic interplay between users and platforms underscores, as just recalled in the opening of this section, the need to consider platforms not merely as neutral intermediaries but as sociotechnical systems that actively structure information flows and user behavior (Couldry & Mejias, 2019).

3. The data dilemma: active contribution or passive trace?

The advent of digital platforms has transformed social research, offering innovative methods that extend far beyond the boundaries of traditional techniques such as surveys, interviews, observations, and content analysis (Padricelli, Punziano & Saracino, 2021). Researchers increasingly rely on digital methods - including web scraping, digital ethnography, text mining, data donation, and even walk-through approaches - to access rich, real-time data. This shift reflects how digital data capture the dynamic, multifaceted nature of social interactions in ways that conventional methods often cannot (Kitchin, 2014; Marres, 2017; Pizzul & Caliandro, 2025).

Today's social research operates within a *post-digital condition* (Cramer, 2014), in which the digital is deeply normalized, integrated, and pervasive element of everyday life and knowledge production. The post-digital paradigm challenge the online/offline binary, framing the digital as both an ontological and methodological condition of research (Marres, 2017; Ruppert, Law & Savage, 2013). Data are *socio-technical artefacts*, emerging from the intersection of human agency, algorithmic logics, and platform affordances (Beer, 2017b). This reframes attention from mere data-gathering techniques to critically reflecting on the infrastructures and *epistemic imaginaries* that shape what can be known (Rogers, 2013; Bucher, 2018) demanding a situated, reflexive, and critical approach to social research.

Digital data-gathering offer the ability to quickly process vast amounts of data and uncover patterns that may not be immediately apparent through traditional approaches. Web scraping and text mining enable large-scale analysis of public opinion, trends, and behavioral patterns (Boyd & Crawford, 2012). Similarly, digital ethnography provides an immersive lens through which digital communities' online practices and cultures can be examined, blending participant observation with digital trace analysis (Hine, 2015).

However, the adoption of these digital techniques is not without challenges. Key issues surrounding data access, privacy, and ethics have become

central to the discourse on digital social research. Access to data is often restricted by proprietary algorithms and terms of service, complicating the process of data collection (Tufekci, 2018). Moreover, the ethical implications of harvesting digital data (especially when it involves personal information) necessitate stringent privacy protections and informed consent protocols (Markham & Buchanan, 2012).

An additional layer of complexity concerns the inherent limitations of the research questions that digital data can address. While digital methods open new avenues for inquiry, they also risk reducing complex social phenomena to quantifiable metrics, potentially oversimplifying nuanced human experiences. Researchers must therefore be cautious in interpreting digital traces and consider the sociocultural contexts underpinning these data (Rogers, 2013). Ongoing debates question whether digital data can fully capture the depth and richness of human interaction or if it will forever remain a supplementary tool to more traditional, qualitative methods (Caliandro & Gandini, 2016).

Looking ahead, the future of social research in the digital age appears promising, with significant potential to refine methodological frameworks that integrate both traditional, digital, and generative approaches. There is a growing consensus that a hybrid methodology, which leverages the strengths of digital data-gathering while addressing its ethical and interpretative challenges, could lead to more comprehensive and informed social analyses. The continued development of digital tools, alongside evolving ethical standards and data governance frameworks, will likely expand the horizon of social research, enabling scholars to tackle increasingly complex questions about social behavior in a digital society (Marres, 2017; Kitchin, 2014).

Integrating these methods with traditional research techniques and supplementing them with new dynamic tools based on computation and generative artificial intelligence promises a more nuanced and robust understanding of contemporary social dynamics, paving the way for innovative and ethically sound investigations into the digital dimensions of social life.

This position reflects what we could call *digital data's duality*. It is both a product of users' active engagement and an artifact of passive digital traces, more or less impacted by the *platform-algorithm unicum* interaction. This complexity raises further questions about the extent to which these data accurately reflect individuals' intentions and behaviors.

The data-gathering landscape in traditional social sciences has historically relied on structured methodologies designed to capture human behavior, social interactions, and institutional dynamics. Classical research instruments include surveys, interviews, ethnographic observations, and experiments, each embedded within specific epistemological and methodological frameworks (Denzin & Lincoln, 2018). These methods are grounded in systematic data

acquisition, prioritizing representativeness, reliability, and validity (Babbie, 2020). Surveys, for instance, have been fundamental in generating structured and comparable responses, while ethnography has allowed for in-depth qualitative insights into lived experiences (Geertz, 1973). The methodological rigor of traditional approaches ensures that data-gathering is intentional and controlled, with researchers explicitly framing their inquiry through theoretical lenses and predefined research questions (Bryman, 2016). The logic underpinning these instruments is one of direct researcher intervention: data is understood as an artifact of designed measurement rather than an emergent byproduct of digital interactions.

The rise of digital environments, however, has brought about a paradigmatic shift in how data is conceptualized, collected and retrieved. With the proliferation of online interactions, social media platforms, and participatory digital spaces, traditional research instruments have been transposed into digital formats, facilitating new data extraction forms (Lazer et al., 2009). Surveys are not conducted online, ethnographic studies have adapted to digital fieldwork, and new techniques such as sentiment analysis, topic modeling, and computational ethnography have emerged (Marres, 2017).

This transformation aligns with the conceptualization of data as an active product of user engagement. Digital environments continuously record deliberate user interactions, such as social media posts, blog entries, and forum discussions, providing a novel form of behavioral data (Boyd & Crawford, 2012). These digital footprints differ from conventional survey responses in that they emerge from organic, self-directed interactions, raising both methodological opportunities and challenges. While digital engagement data offer rich behavioral insights, they also introduce issues of self-selection bias and contextual dependency (Ruths & Pfeffer, 2014).

Parallel to the active engagement model, an even more radical epistemological shift has occurred with the emergence of passive digital traces (Kitchin, 2014). Unlike self-reported survey responses or explicitly curated content, these traces (metadata) are often involuntarily generated as byproducts of digital interactions. The quantification of human behavior has thus expanded beyond direct user input to include implicit, often invisible, processes of data capture (Zuboff, 2019). From a methodological standpoint, this shift poses significant challenges regarding data representativeness, ethical concerns, and the interpretability of digital traces (Mittelstadt et al., 2016). While passive data can provide unprecedented granularity in tracking user behavior, it also raises critical issues about surveillance, consent, and algorithmic opacity (Eubanks, 2018). Researchers must therefore navigate the duality of increased data availability and the ethical complexities of using unobtrusively retrieved information.

Post-Digital Data-Gathering and the Adaptive Epistemological Framework: Navigating the Human-Algorithm-Platform Nexus

Suania Acampa, Giuseppe Michele Padricelli, Gabriella Punziano

Despite significant advances in data analysis through artificial intelligence (AI) and machine learning (ML) techniques, developing suitable data retrieval tools remains a critical challenge in studying the dynamic, adaptive, and processual nature of user interaction with the *platform-algorithm unicum*. Traditional social sciences methods have struggled to capture the complexity of these interactions as they unfold in real-time and are shaped by ever-evolving computational infrastructures (Kitchin, 2014; Boyd & Crawford, 2012). Unlike static data collection approaches, studying algorithmically mediated environments requires adaptive and responsive tools capable of tracing and interpreting interaction patterns within digital ecosystems (Gillespie, 2014).

One of the central obstacles is the *issue of data access*, which presents challenges across three primary dimensions: *normative*, *technical*, and *ethical*. Normatively, the challenges to data access include regulatory constraints (GDPR, CCPA); technical challenges regarding proprietary APIs, and the increasing of algorithmic opacity, called the "black box" problem (Binns, 2018; Pasquale, 2015) which complicate access to how content is filtered, ranked, and recommended (Ananny & Crawford, 2018). Moreover, the scale and speed of user-platform interactions necessitate computationally efficient and scalable retrieval methods, requirements often beyond the scope of conventional survey-based or experimental designs (Bruns, 2019a).

Ethically, the collection and analysis of digital traces raise concerns regarding user autonomy, informed consent, and the potential for surveillance-based research methodologies (Zuboff, 2019). Many users remain unaware of the extent to which their interactions generate data, leading to asymmetrical power dynamics between platforms, researchers, and participants (Couldry & Mejias, 2019). This necessitates the development of ethical frameworks that balance the need for empirical insight with respect for digital rights and participatory research ethics (Nissenbaum, 2010).

To address these challenges, a shift towards adaptive, dynamic, and processual data retrieval approaches is essential. One promising avenue is computational ethnography and real-time tracking systems, which leverage AI-driven methodologies to study user engagement with digital platforms as it unfolds (Marres, 2017). Additionally, participatory and crowdsourced data collection models, such as those based on citizen science principles, empower users to contribute data on their own terms while maintaining transparency and agency (Andrejevic, 2020). Finally, the recalled concept of data hybridization - which integrates passive digital traces with active user-generated inputs - offers a holistic perspective on platform-user interactions, allowing for the simultaneous study of behavioral patterns, algorithmic influences, and contextual variabilities (Felaco, 2022).

The interplay between active engagement data and passive digital traces has led to a reconceptualization of digital data as inherently hybrid. Rather than treating these data types as separate entities, scholars have increasingly recognized their interdependence within digital ecosystems (Couldry & Mejias, 2019). The hybrid nature of digital data blurs traditional methodological boundaries between quantitative and qualitative, self-reported and observational, deriving from intrusive and nonintrusive collection techniques, and explicit and implicit data (Latour, 2005).

The methodological evolution of data retrieval tools must align with the epistemological frameworks underpinning digital social research that could reflects the multiprocessability, complexity, and contextuality of digital data. The hybridization of data results in the development of an *adaptive epistemological framework,* a paradigm shift in the social sciences that is not only more efficient but also adaptive, capable of dynamically adjusting to new data inputs and offering deeper insights into complex social phenomena. Such an approach acknowledges algorithmic interactions' fluidity and situated nature, calling for methodological pluralism, interdisciplinary collaboration, and continuous refinement of data-gathering techniques (Ruppert, Law, & Savage, 2013). By embracing these principles, researchers can move beyond static snapshots of digital behavior toward a more nuanced, real-time understanding of the evolving interplay between users and the *platform-algorithm unicum*. This shift acknowledges that digital traces are not mere reflections of user behavior, but active constructs shaped by algorithmic mediation, platform affordances, and sociotechnical systems (Beer, 2017a) that emerge from the interaction between user and *platform-algorithm unicum*. As a result, researchers must develop adaptive, iterative, and cross-disciplinary techniques (Venturini & Latour, 2010) rooted in computation, artificial, and generative methods.

The move from traditional to digitally mediated research calls for rethinking how data is conceptualized, collected, retrieved and analyzed. Data hybridization offers a powerful lens for understanding online behavior, while maintaining ethical integrity and methodological rigor. Embracing an a*daptive epistemological framework* is essential for meeting the challenges of computational social science.

4. Craft and collect: adapt data-gathering techniques to evolving platforms

The exploration of intervention contexts, opportunities, and current challenges in social research unfolds along various trajectories, all anchored by a pivotal factor: the researcher's choices. These decisions shape the adaptive

epistemological frontiers and methodological approaches that define research practices in contemporary digital contexts.

It is therefore imperative to delineate the technical dimensions and scopes these trajectory may encompass, shedding light on the limitations and opportunities that can influence the effectiveness of each technique employed to generate knowledge. The configurations of the interplay between users and the *platform-algorithm unicum* are molded by the foundational premises for data access and data retrieval. The dualism between data crafting and data-gathering represents the initial crossroads that determines its direction.

In data crafting, the emphasis lies in the researcher's active construction of primary data, operationalized through intrusive or non-intrusive techniques, specifically designed for the study. Conversely, data collection and retrieval involves non-intrusive techniques to gather pre-existing data which the researcher uses secondarily.

Within this landscape, data-gathering techniques emerge as adaptive components of a broader epistemological reconfiguration. The growing complexity of digital ecosystems calls for methodological strategies that are sensitive to both the technological infrastructures and the socio-technical dynamics shaping data production (Marres, 2017; Kitchin, 2014). What follows is a comparative presentation of two distinct data collection approaches: one grounded in qualitative, presence-based, and more intrusive research engagement; the other rooted in quantitative, unobtrusive, and researcher-absent modalities. These dual trajectories aim to exemplify how, within an *adaptive epistemological framework*, data-gathering techniques must evolve also in relation to the shifting nature of methods, the fluid roles of actors—both human and algorithmic—and the redefinition of data itself as a socio-technical construct (Couldry & Mejias, 2019; Dencik et al., 2019). Such a perspective reframes digital data as the emergent outcome of layered and multi-influenced interactions, rather than as a neutral resource awaiting extraction.

This organization is essential to underscore how the interplay between the researcher and the observed context-subjects is challenged — technically, ethically, and regulatorily, depending on the circumstances in a digital scenario and inside the interplay among users and *platform-alghoritm unicum*. Following this systematization - and without claiming exhaustiveness - a scalable typological framework is proposed to assist researchers in identifying, based on situational and contextual conditions, the strengths and limitations of each technique.

4.1. Digital vs computational ethnography

When discussing the intrusiveness of research techniques, it is essential to begin with the premises of the ethnographic method, typically characterized as a direct observation approach that generally involves the researcher's participation and interaction in the daily lives of individuals over a certain period (Spradley, 1980; Gobo & Marciniak, 2016; Hammersley & Atkinson, 2007). Today, the ethnographic method has experienced a revival, enabled by new opportunities to study and observing online social environments.

Digital ethnography entails a thorough reconsideration of observational opportunities within communities, both technically and contextually. Technically, the first reconsideration aligns with a new interpretation of the concept of intrusiveness. Ethnographic observation is most of the time classified as "participant" due to the researcher's presence in the field. However, the premise of participation changes when operating in online observation fields. In this context, the researcher's presence blends seamlessly with the field and the observed users, leading to new forms of online participation, such as lurking — a constant and concealed presence of the researcher in the observation field. In its digital version the ethnographer is responsible not only for identifying the relational principles within communities but also for collecting what users produce and share in terms of meaning and knowledge. In line with Rogers' perspective (2013), the digital ethnographer operates within a post-demographic research vision. The selected unit of analysis is typically not the individual but rather the product of their actions. This perspective fits into a broader debate on the ethical characterizations. It is necessary to emphasize, however, that adopting a post-demographic vision is not a discretionary choice, but an obligatory passage stemming from the very ontology of the observation field, where information such as gender, age, origin, and social status are not always structural components of the platforms. In this sense, the post-demographic vision assumes a dual connotation: it constitutes both a limitation and an opportunity, as it opens up new methodological possibilities within the adaptive epistemologies underpinning the digital research.

However, the necessary recourse to post-demographic research pathways does not exempt the social researcher from ethical responsibility. The question of regulation and respect for personal data and identities — once these are transferred into digital spaces — goes far beyond the necessary adherence to general research best practices.

These practices are often negotiable, as they are frequently subject to exceptions arising from the subjectivity of research designs.

This is enough to sustain an ongoing debate around ethical principles. At the heart of this debate are two major areas of discussion: the first concerns the

potential post hoc identifiability of users who produce the content collected by digital ethnographers online; the second pertains to the long-standing issue of the boundary between what is public and what is private. The first area of discussion relates to the various sampling opportunities across digital spaces from which data are collected. The second area focuses on issues of intellectual property.

The identification of users involved in sampling strategies represents a weak point in research designs that concentrate on small-scale observational settings such as social media groups or comments on posts with limited user interaction — especially when addressing sensitive topics.

In these cases, despite the researcher's effort to anonymize the sources and obscure the identities involved in the presentation of results, a post hoc intervention by third parties does not always rule out the possibility of reverse-following the research design to trace back to the individuals involved in the study. This hypothetical scenario presents a dilemma that goes far beyond the dichotomy of what is permissible versus what is not. Instead, it foregrounds the need to establish an ethical threshold — a line that should not be crossed when full protection of participants' identities cannot be unequivocally ensured.

The second area of discussion emerges as a much more fluid and blurred boundary within digital environments.

While the platforms' terms of service (TOS) generally state that ownership of content rests with the platform itself, it must be acknowledged that users' digital productions are still potentially tied to individual actions and decisions. Therefore, when evaluating the boundary between public and private, the issue is not simply one of ownership; it is necessary to focus on the extent to which the researcher may be intruding into spheres external to their own.

As with the first area of discussion, here too it is necessary to draw a non-negotiable ethical boundary — one that cannot be shifted. What can and should shift, however, are the epistemological horizons of knowledge production. These can be expanded through the advancement of research techniques which, when applied within ethical boundaries, allow for deeper insights into the research subject. This is, ultimately, a methodological issue.

In light of this, digital methods open up new research possibilities previously unattainable through traditional ethnographic and survey research techniques. However, the single-method application of digital methods — depending on the research design — may leave crucial elements of meaning and context unexamined, which cannot be investigated solely through digital ethnography.

The transition from traditional to digital ethnography has not been immediate or linear.

The migration of ethnographic applications online has enabled a shift from data-crafting to data collection and retrieval, thereby allowing for the inclusion of new types of information to be transformed into data. Data flows, digital traces of online activity, and unintentional information — shaped by the algorithm's embeddedness within the platform — are subjected to analytical categories typically employed in the study of digital content. For instance, content analysis frameworks, traditionally applied to digital texts, are not being adapted to the study of structured digital content (e.g., frequency of posts on a given topic, access times, session durations).

This theoretical shift has given rise to a novel interpretative model known as *Computational Ethnography* (Peponakis et al., 2023; Abramson et al., 2017; Brooker, 2022). From the researcher's perspective, within a hybrid view of technical intervention, this model entails a processual approach to the *unicum* of data and acknowledges the potential overlap between the phases of data gathering and those of modeling and analysis. This framework enables the analysis of large datasets that can "reduce the cost of analysis and save time toward finding answers to anthropological questions which would, otherwise, be very difficult to address" (Peponakis et al., 2023, p. 7).

Examples include the use of ethnographic heatmaps (Abramson & Dohan, 2015), the implementation of Social network analysis (SNA) and data visualization outcomes (Abramson et al., 2017) and code-analytic procedures that require comparisons with an interpretively coded dataset, connecting behavior to linguistic and cognitive structures (Abramson et al., 2017; Gonzalez-Bailon & Paltoglou, 2015).

Specifically, the computational perspective in ethnography is closely tied to the concept of innovation, understood both as the repertoire of available technological tools and the researcher's capacity to employ them effectively. Within this framework, two broad modalities of data collection and processing can be identified. On the one hand, there are approaches that use technologies exogenous to the field of observation as facilitators of the ethnographic task (e.g., recording, note-taking tools). On the other hand, there are strategies that rely on the direct deduction of information — for instance data visualization procedures and semi-automated programs — whose linguistic logic helps illuminate relational processes. According to Peponakis et al. (2023), such tools enable the representation of facts and rules, supporting the ethnographer in constructing knowledge through syntax, semantics, and reasoning. These instruments help bridge the gap between qualitative and quantitative types of research, and between intrusive and nonintrusive data-gathering models, thus creating a more integrated research paradigm.

4.2. Data scraping and API's querying vs data donation

The evolution of non-intrusive digital data collection techniques mirrors the broader transformations of digital society - marked by the growing availability of data and the tension between openness and control - and the growing complexity of digital environments. This trajectory spans from early web scraping practices to structured access via APIs, and more recently, to user-centric approaches like data donation. While we acknowledge that, from a post-digital perspective, it is necessary to move beyond the dichotomy between digitalized and digitally native techniques (Rogers, 2013) in favor of an evolutionary vision (Amaturo & Aragona, 2019), it is clear that web scraping and API querying exhibit characteristics that reflect their relationship with the technical infrastructure and digital objects; characteristics that make them specific to digital environments.

Web scraping emerged as a way to automate data collection from websites, by passing the limits of manual retrieval. Initially based on simple scripts parsing HTML structures, the scraping evolved and refined into sophisticated tools to respond to the complexity of digital spaces, such as automated tools and headless browsers (for example Selenium), capable of simulating the real navigation of a user – even in complex and dynamic sites – without requiring visual resources, thus bypassing possible anti-scraping control systems activated by websites. Despite its utility, scraping is fragile - sensitive to minor changes in site structures - and operates in a legal and ethical grey area; even when the data is public, automatic extraction can violate the TOS. Particularly in the wake of GDPR (GDPR, Regulation EU 2016/679), which restricts the collection of personal data without consent.

With the advent of algorithmic platforms, the digital ecosystem grew more complex, articulated, interconnected, and simultaneously more closed. Initially, collecting data from social media through scraping techniques was possible, often with tools that directly accessed the source code of profile pages, posts, or comments. However, starting in 2014, many platforms have progressively "institutionalized" access to their data through APIs (Application Programming Interfaces), i.e., official interfaces that allow querying and obtaining data from platforms in a semi-structured way.

APIs have quickly become a privileged access point to explore the socio-technical structures and cultural processes generated by users' daily interaction with digital platforms, opening up new perspectives for analyzing complex phenomena with quantitative and computational approaches in a longitudinal way. They have fostered a small but significant paradigm shift: through intuitive tools it has been possible to collect detailed data on a large scale, making digital research accessible even to those who didn't possess advanced programming

skills (Venturini & Rogers, 2019). However, this approach also presents critical issues, including new sources of opacity, incompleteness, and information asymmetry. In addition to the general limitations related to data collected on social media (for more information, see Roger, 2018), the first limitation of data collection via APIs query concerns the control exercised by the platforms on what is made available. Some platforms may restrict the number of requests per time time frame, require authentication via revocable keys. A further problem is represented by the lack of transparency on the criteria with which the data is returned. Often, the researcher has no way of knowing exactly how the APIs operate, nor which algorithmic mechanisms or business logics influence the return of the data.

Nonetheless, social media data collected through APIs have enabled longitudinal investigations on users, content, and metadata and their explicit (user-user; user-content) and implicit (user-metadata; content-metadata) relationships (Veltri, 2021). The latter, in particular, offers insight into processes of content creation and dissemination and on the algorithmic logics underlying them, thus shifting the research focus from the user or content to what we have defined as a user-platform-algorithm unicum.

APIs are not only a technological change, but also a change in understanding the nature of the interaction between human and non-human actors in the data collection. The progressive institutionalization of access makes it clear that data production and collection can no longer be considered a neutral process but the result of a continuous and situated interaction between the user, platform affordances, algorithmic architecture, and economic and governance logics that regulate APIs.

Following a period of accessibility and flourishing creation of data extraction tools (for deeper information, visit the Digital Methods Initiative), accessibility has progressively decreased, especially following the Cambridge Analytica scandal in 2018. This event marked a turning point after which many technology companies (notably Meta and Twitter) have drastically closed free access to their APIs (reserving it to selected commercial partners) and reduced the possibilities for independent researchers and academics to conduct large-scale analyses.

These restrictions have had profound effects on research. Bruns (2019b) aptly describes this transformation as a real 'APIcalypse', denoting a structural crisis for social research, which must now continuously negotiate data accessibility amid the growing tension between openness and control.

In response to the numerous criticisms from scholars (Freelon, 2018), some platforms have tried to activate specific initiatives to facilitate access to data for academic purposes including Social Science One or Crowdtangle, in an effort to balance the needs of research with the protection of privacy. However,

these efforts have been ineffective, partial, and short-lived. The current solution proposed by Meta to open up to the world of research is the Virtual Data Enclave (VDE) by SOMAR, which introduces a new constraint: not only is access to data regulated and controlled, but also the type of analysis that can be carried out with those data. It is possible to access data exclusively within the virtual enclave, a protected, isolated, and controlled environment where researchers must conduct their analyses. Data cannot be downloaded or exported from the VDE, and the review board must approve any results before they are included in a research report and disseminated.

In light of these structural constraints on data access, researchers have begun to explore new strategies that reframe users not as passive data subjects but as active contributors to the research process. One of the most promising of these strategies is data donation.

Data donation represents an emerging and transformative paradigm in digital social research, evolving from a simple data-gathering technique to a methodological approach grounded in user self-determination, ethical responsibility, and the innovative use of digital trace data (Haim et al., 2023; Araujo et al., 2022). Through the voluntary sharing of personal data by users - via browser extensions, dedicated applications, or direct export of Data Download Packages required by the GDPR - it is possible to access granular information on online interactions: browsing histories, algorithmic preferences, suggested content, and consumption behaviors (Boeschoten et al., 2020). The data obtained is hybrid and processual, generated during the interaction between subject, platform, and algorithmic architecture: a concrete form of data co-production in the interplay between user and platform-algorithm unicum. From a technical point of view, data donation is mainly based on the download-upload model which allows collecting cross-platform data produced in real digital environments. As platform access continues to narrow, data donation emerges not only as a methodological and political response, capable of returning agency to users and access to researchers, offering a legitimate and transparent alternative to collection via scraping or APIs. However, the quality and interpretability of data depend on the researcher's ability to transparently structure the extraction, cleaning, categorization, and enrichment processes, while minimizing the risk of analytical biases (Boeschoten et al., 2020). The rise of native tools confirms the growing accessibility of self-determined data sources — accessible without the researcher's direct intervention (Ohme et al., 2021). The effectiveness of data donation also lies in its ability to reduce biases deriving from algorithmic/platform mediation, as seen in API-based collection , and from users' self-perception as in surveys (Haim et al., 2023). Ethically, data donation represents a transparent alternative to extractive practices, as it is based on informed consent and user agency. However, this approach must be

supported by transparent, responsible, and privacy-oriented protocols that are fully sustainable. Several studies underline the importance of designing communication messages and collection mechanisms that strengthen participants' sense of control, avoiding information ambiguities (Ohme & Araujo, 2022). The motivations that drive individuals to share their data are often prosocial, linked to the desire to contribute to the common good and the progress of public research (Skatova & Goulding, 2019). Yet, this presupposes an ethical and transparent management of the shared data. One key limitation concerns sample representativeness. However, while data donation may be subject to self-selection bias, recent studies have downplayed this concern, showing that the willingness to donate data does not vary significantly across demographic groups (Keusch et al., 2024). Other research highlights that factors such as clarity about data processing, the research purpose, and trust in the researcher have a greater impact on willingness to participate than age or social position (Kmetty et al., 2024). Furthermore, while data donation is formally non-intrusive, it is not immune to reactivity effects. Users, aware that their data will later be donated for research purposes, may alter their behavior during data production. This conscious self-monitoring can influence how users interact with platforms, potentially reducing the authenticity and spontaneity of their digital practices. Thus, while data donation enhances ethical standards and user empowerment, it may also introduce new forms of reactivity. From an operational point of view, technical obstacles still persist: not all platforms allow easy data export; formats can be heterogeneous, incomplete, or complex to validate; and the need for secure infrastructures for data management requires significant effort from research teams. To respond to these challenges, open-source frameworks have been proposed for the safe and compliant collection of donated digital data, capable of combining scientific standards, legal security, and user empowerment (Araujo et al., 2022). Data donation offers a valuable way to capture the interaction between users and algorithmic systems in native contexts, as it allows access to in-situ behavioral traces that reflect the interplay between platform affordances and user practices. At the same time, it marks an epistemological shift that, as Caliandro (2024) notes, redefines the user not merely as a data source but as an active methodological resource. This transition from an extractive model to one of reflective and responsible collaboration represents a key turning point for digital social research in the post-API era.

4.3. Systematizing the process

The dualism features of digital ethnography and computational ethnography, as well of data donation and web scraping technics, clearly

illustrate the opportunities for integrating these techniques within a dynamic epistemological perspective that is also adaptable to the evolving nature of platforms and the knowledge objectives centered on users' movements across them.

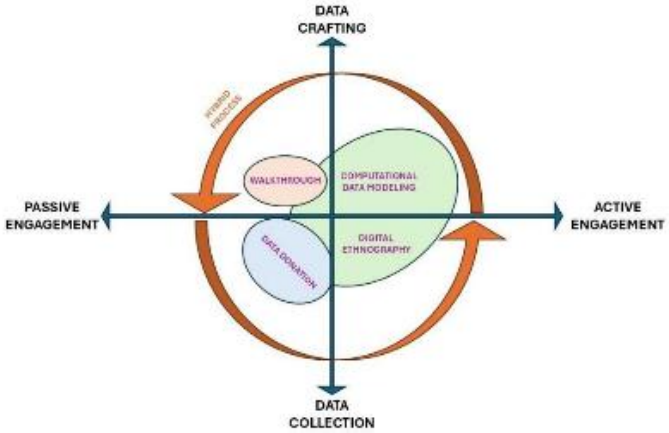
In this context, a typological framework is proposed, which highlights on the one hand, the multi-processability, complexity, and contextualization of data, and on the other, the degree of user awareness regarding their digital trajectories.

The framework presented in the diagram reflects the interplay between two dualities discussed in this paper: first, the level of user engagement and awareness in generating data that contribute to algorithmic processing; second, the researcher's role — whether in translating raw interactions into structured data through their operational definitions, or collecting ready-to-use data for further analysis.

As illustrated in Figure 1, the model follows a process of hybridization that has no clear starting or ending point on the scheme but rather unfolds in a circular direction. The techniques discussed in the previous paragraph are positioned within this circular schema to represent their methodological orientation. In line with the methodological pluralism which, as mentioned, allows the epistemological perspective to adapt to the specific case, Figure 1 shows not only the placement of each individual technique within its respective quadrant, but also highlights potential overlaps among the techniques themselves. These overlaps align with the idea of integrated and compensatory use of the techniques, insofar as each can help to overcome the limitations of the others (e.g., data access, ethical issues, etc.).

In a research design that adopts sequential or nested mixed-digital approaches, the use of multiple investigative techniques not only expands the interpretive horizon but becomes strategically necessary when studying rapidly evolving phenomena, such as the complex interactions between users, platforms and algorithms. This methodological structure, in this sense, provides a way to adapt to the evolving nature of the research object.

Figure 1 - a systematization proposal



The structure of the framework, ready-to-use and adaptable according to the research objectives, is not limited to the techniques described in this paper, though they are closely interwoven in the brief example presented.

For example, a study focused on human-platform-algorithm interactions, such as understanding recommendation mechanisms on cultural distribution platforms, would align closely with this framework. Consider a case study examining a music streaming platform: the researcher may begin with analyzing the platform's structure through walkthrough operations, a technique that explores how digital platforms guide, channel, and shape user interactions through visual, functional, and symbolic affordances. Following this initial exploration, and upon discovering that the playlist recommendation system is based on users' listening histories and on the sharing of songs outside the original platform, the researcher may encounter limitations related to accessing the streaming data of a sampled user set needed for empirical investigation.

Faced with this limitation, the researcher could pivot to collecting passive digital traces or implement a data donation strategy, wherein users voluntarily provide their interaction data. Moreover, if additional contextualization is required, in an integrational research vision the researcher should shift toward active engagement and gather user-generated content on the topic from across the web and, based on that material, develop listening profile models.

5. Conclusions

Human actions are increasingly entangled with opaque algorithmic processes that shape everyday behaviors through mechanisms of

recommendation, personalization, and influence (Zuboff, 2019). These dynamics are especially visible in domains such as cultural consumption and news curation, where platform algorithms modulate access to information in real time. The result is a divergence between online and offline behaviors that underscores the socio-technical complexity of digital environments. From a methodological standpoint, this evolving landscape necessitates hybrid approaches that allow researchers to observe algorithmic personalization in situ, trace real-time interactions, and reconstruct the continuum of user-platform-algorithm relations.

The techniques explored in this paper reflect this need for methodological complementarity. They are not limited to digital tools but instead integrate traditional research strategies to account for the entangled nature of human and algorithmic agency (Seaver, 2017). This shift also raises critical questions about data privacy, identity negotiation, and the evolving contours of knowledge production in an AI-mediated world. As Sebastian et al. (2025) argue, artificial intelligence and natural language processing technologies introduce new cognitive and methodological frontiers that challenge the epistemic foundations of social inquiry.

Rather than adopting either deterministic or Luddite positions, the current epistemological imperative is to develop adaptive frameworks that engage critically with the technosocial dynamics at play. The digital datum, understood here as the core unit of human-algorithm mediation, reveals the need to rethink both data construction and research design. These elements are increasingly co-constructed within the user-platform-algorithm nexus, requiring that methods evolve alongside the contexts they study. Just as early digital research necessitated the adaptation of traditional methods, today's approaches must remain agile and responsive to technological change.

A case in point is ethnographic methodology in digital environments. The shift from data creation to data modeling, typical of computational ethnography, illustrates how analysis is now embedded within the collection process itself. This development occurs within an epistemological regime that is inherently adaptive, driven by the continuous interaction of technological advancement and cultural transformation. It also demands a reconsideration of the researcher's positionality, which now encompasses strategic choices about tools, techniques, and the interpretive frameworks applied to increasingly complex datasets.

For example, data donation mechanisms and walkthrough methods exemplify how contemporary strategies attempt to decode the situatedness of user experiences within platforms. Yet these innovations do not erase methodological challenges—they merely reframe them. This underscores the need to systematize debate around epistemological adaptation by designing new

tools and frameworks that address the limitations and potentials of digitally integrated research. As Housley et al. (2014) suggest in his concept of *digital remastering*, the task is not simply to revise existing techniques but to reconfigure them entirely in response to the new epistemic conditions.

While algorithmic opacity remains, methodological innovation helps decode the mechanisms through which meaning is constructed in digital settings. Researchers must now work within dual tensions: the evolving, often unpredictable nature of digital infrastructures, and the disciplinary conventions that define valid social science inquiry. The rise of generative AI demands that epistemology not only adapt but transform—shifting from method-driven answers to a new emphasis on asking the most generative, context-sensitive questions possible.

References

Abramson C.M. & Dohan, D. (2015) Beyond text: Using arrays to represent and analyze ethnographic data. *Sociological Methodology,* 45, 272–319.

Abramson, C. M., Joslyn, J., Rendle, K. A., Garrett, S. B. & Dohan, D. (2017). The promises of computational ethnography: Improving transparency, replicability, and validity for realist approaches to ethnographic analysis. *Ethnography, 19*(2), 254–284. https://doi.org/10.1177/1466138117725340.

Airoldi, M. (2021). *Machine habitus: Toward a sociology of algorithms*. Hoboken, NJ: John Wiley & Sons.

Amaturo, E., & Aragona, B. (2019). Per un'epistemologia del digitale: note sull'uso di big data e computazione nella ricerca sociale. *Quaderni di sociologia*, *81*(81-LXIII), 71-90.

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.

Andrejevic, M. (2020). Automated Media. London: Routledge.

Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Welde, B., de Vreese, C. & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, *4*(2), 372-387.

Babbie, E. (2020). *The Practice of Social Research*. Boston: Cengage Learning.

Beer, D. (2017a). *The data gaze: Capitalism, power, and perception*. London: Sage.

Beer, D. (2017b). The social power of algorithms. In *Information, Communication & Society, 20*(1), 1–13.

Binns, R. (2018). *Fairness in Machine Learning: Lessons from Political Philosophy*. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, 149-159.

Bishop, S. (2019). *Managing visibility on YouTube through algorithmic gossip. New Media & Society, 21*(11–12), 2589–2606.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

Bryman, A. (2016). *Social Research Methods*. Oxford: Oxford University Press.

Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). Digital trace data collection through data donation. https://doi.org/10.48550/arXiv.2011.09851

Brooker, P. (2022). Computational ethnography: A view from sociology. *Big Data & Society*, *9*(1), 1-6. https://doi.org/10.1177/20539517211069892

Bruns, A. (2019a). *Are Filter Bubbles Real?*. Cambridge: Polity Press.

Bruns, A. (2019b). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society,* 22(11), 1544-1566.

Bucher, T. (2018). *If... Then: Algorithmic power and politics*. Oxford: Oxford University Press.

Cadwalladr, C., & Graham-Harrison, E. (2018). *The Cambridge Analytica files: How Facebook's data scandal unfolded. The Guardian.*: https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm

Caliandro, A. (2024). Follow the user: Taking advantage of Internet users as methodological resources. *Convergence*, 13548565241307569.

Caliandro, A., & Gandini, A. (2016). *Qualitative research in digital environments: A research toolkit*. London: Routledge.

Cotter, K. (2019). *Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. New Media & Society, 21*(4), 895-913.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford: Stanford University Press.

Cramer, F. (2014). What is 'Post-digital'? In A. Berry & M. Dieter (Eds.), *Postdigital aesthetics: Art, computation and design* (pp. 12–26). Palgrave Macmillan. https://doi.org/10.1057/9781137437204_2

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). *Data Justice and the Right to Political Imagination. Data Justice: A Critical Inquiry into Datafication*, 1-20.

Denzin, N. K., & Lincoln, Y. S. (2018). The SAGE Handbook of Qualitative Research. London: SAGE Publications.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media.* Cambridge: Harvard University Press.

Eriksson, M., Fleischer, R., Johansson, A., & Snickars, P. (2019). *Spotify teardown: Inside the black box of streaming music.* Cambridge: MIT Press.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2016). *First I "like" it, then I hide it: Folk theories of social feeds. CHI '16 Proceedings*, 2371-2382.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* New York: St. Martin's Press.

Felaco, C. (2022). *Algorithmic awareness and user behavior in digital platforms. Journal of Digital Culture, 15*(3), 120-137.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. New York: Pantheon Books.

Freelon, D. (2018). Computational research in the post-API age. *Political Communication,* 35(4), 665-668.

Gekker, A., & Hind, S. (2019). *Gaming the metrics: Misconduct and manipulation in video game culture. Games and Culture, 14*(6), 1-21.

Geertz, C. (1973). *The Interpretation of Cultures*. New York: Basic Books.

Gillespie, T. (2014). *The relevance of algorithms*. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). Cambridge: MIT Press.

Gobo, G., Marciniak, L. (2016). What is Ethnography. In D. Silverman (Ed.), *Qualitative Research* (pp. 103-119). London: Sage.

Gonzalez-Bailon, S. & Paltoglou, G. (2015). Signals of public opinion in online communica tion: A comparison of methods and data sources. *The Annals of the American Academy of Political and Social Science* 659, 95–107.

Haim, M., Leiner, D., & Hase, V. (2023). Integrating data donations in online surveys. *Medien & Kommunikationswissenschaft*, (1-2), 130-137.

Hammersley, M., Atkinson, P. (2007). *Ethnography: Principles in Practice*. London: Routledge.

Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Bloomsbury Academic.

Housley, W., Procter, R., Edwards, A., (2014). Big and broad social data and the sociological imagination: A collaborative response. *Big Data & Society,* 1(2), 1–15. https://doi.org/10.1177/2053951714545135.

Introna, L. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology & Human Values, 41*(1), 17–49.

Keusch, F., Pankowska, P. K., Cernat, A., & Bach, R. L. (2024). Do you have two minutes to talk about your data? Willingness to participate and nonparticipation bias in Facebook data donation. *Field Methods*, *36*(4), 279-293.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage.

Kitchin, R. (2017). *Thinking critically about and researching algorithms*. *Information, Communication & Society, 20*(1), 14–29.

Kmetty, Z., Stefkovics, Á., Számely, J., Deng, D., Kellner, A., Pauló, E., & Koltai, J. (2024). Determinants of willingness to donate data from social media platforms. *Information, Communication & Society*, 1-26.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King., G., Macy, M., Roy, D. & Van Alstyne, M. (2009). Computational Social Science. *Science*, 323(5915), 721-723.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Lobato, R. (2019). *Netflix nations: The geography of digital distribution*. New York: NYU Press.

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (version 2.0)*. Association of Internet Researchers.

Marres, N. (2017). Digital sociology: The reinvention of social research. Polity.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big data & society*, *3*(2), doi: 2053951716679679.

Napoli, P. M. (2019). *Social media and the public interest: Media regulation in the disinformation age*. New York: Columbia University Press.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford University Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.

Ohme, J., & Araujo, T. (2022). Digital data donations: A quest for best practices. *Patterns*, *3*(4).

Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, *9*(2), 293-313.

Padricelli, G. M., Punziano, G., and Saracino, B. (2021). Virtual methods and digital methods: examples of netnography and digital ethnography for a comparison between methods for analyzing the digital scenario in tourism studies. *Athens Journal of Social Science*, 8, 259–278. doi:10.30958/ajss.8-4-4

Peponakis, M., Kapidakis, S., Doerr, M., Tountasaki, E. (2023). From Calculations to Rea soning: History, Trends and the Potential of Computational Ethnography and Computational Social Anthropology. *Social Science Computer Review,* 1-20. doi: 10.1177/08944393231167692.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.

Pizzul, D., Caliandro, A. (2025). The digital data double standard: economic value extraction versus scientific knowledge limitation and the opportunities of data donation. *The Lab's Quarterly*, forthcoming.

Rader, E., & Gray, R. (2015). *Understanding user beliefs about algorithmic curation in the Facebook News Feed. CHI'15 Proceedings*, 173-182.

Rogers, R. (2013). Digital Methods. Cambridge: MIT Press.

Ruppert, E., Law, J., & Savage, M. (2013). *Reassembling Social Science Methods: The Challenge of Digital Devices. Theory, Culture & Society, 30*(4), 22-46.

Ruths, D., & Pfeffer, J. (2014). Social Media for Large Studies of Behavior. Science, 346(6213), 1063-1064.

Seaver, N. (2017). *Algorithms as culture: Some tactics for the ethnography of algorithmic systems. Big data and society, 4*(2). https://doi.org/10.1177/2053951717738104

Seaver, N. (2019). *Captivating algorithms: Recommender systems as traps. Journal of Material Culture, 24*(4), 421-436.

Sebastian, R., Kottekkadan, N.N., Thomas, T.K., Niyas KK, M. (2025). Generative AI tools (ChatGPT) in social science research. *Journal of Information, Communication and Ethics in Society*, 23(2), 284-290. https://doi.org/10.1108/JICES-10-2024-0145

Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. *PloS one*, *14*(11), e0224240.

Spradley, J.P. (1980). *Participant Observation*. New York: Holt, Rinehart and Winston.

Tufekci, Z. (2018). *Twitter and tear gas: The power and fragility of networked protest*. New Haven: Yale University Press.

Vaidhyanathan, S. (2018). *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford: Oxford University Press.

Veltri, G. (2021). *La ricerca sociale digitale*. Ostiglia: Mondadori.

Venturini, T., & Latour, B. (2010). The Social Fabric: Digital Traces and Quali-quantitative Methods. Paris: Proceedings of Future En Seine.

Venturini, T. & Rogers, R. (2019). API-based research or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. *Digital Journalism*, 7(4), 532-540.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.