

The algorithmic public opinion: a literature review

Alessandro Gandini¹, Alessandro Gerosa¹, Beatrice Gobbo²,
Silvia Keeling¹, Luisa Leonini¹, Lorenzo Mosca¹, Marco Orofino¹,
Urbano Reviglio¹, Sergio Splendore¹

¹ University of Milano

² Politecnico di Milano, Density Design Lab

In the context of a society where digital technologies have come to pervasively intermediate most social, cultural and economic processes, algorithms represent an important socio-technical component of ‘the social’. Algorithms of various kinds - recommendation algorithms, reputation algorithms, and many more - regulate and organize how users access content online. Some recommend what movies to watch, what items to purchase, who to trust when using an online service, and are increasingly employed by public and private actors for a variety of purposes, including work, crime prevention, credit risk assessment. Some other algorithms moderate the content posted online and decide about which types of materials are deemed appropriate to circulate. Algorithms are the key infrastructures sustaining the business models of social media platforms, which are based on the prediction of user behaviour and its influence for purposes of monetisation of targeted advertising - a model that has been labelled as ‘surveillance capitalism’ (Zuboff, 2019). However, their inner workings continue to remain somewhat obscure. We still know very little about how algorithmic outputs are generated, and the social, cultural and economic implications that derive from their implementation and proliferation.

The opaque nature of algorithmic workings in social media and platforms of all sorts has been commonly described through the ‘black box’ metaphor (Fig. 1), which has come to quintessentially represent the main way to describe the unknowability of algorithmic elaborations: we know the input and the output of an algorithm, but we are largely unaware of what happens in between (Pasquale, 2015).



Figure 1: The Black Box metaphor

Yet, while certainly useful at a metaphorical level, the ‘black box’ metaphor epistemologically constrains social researchers in their capacity to account for the complex set of factors that concur to the determination of algorithmic elaborations (Bucher, 2018). In particular, it tells only part of

a story where the technical blends with the cultural and the social in original ways, and where human action remains central. While often provided with their own agency, algorithms are ultimately created by humans within enterprises with specific economic logics underpinning their creation. The investigation of how algorithms work therefore cannot be limited to the ‘box’ that hides the ‘code’. It must, instead, extend onto the critical observation of the structures and infrastructures that surround algorithmic elaborations, as well as onto the cultures that these intervene within, and the economic goals these are programmed to contribute to.

Unsurprisingly, algorithms as conceived here also hold an important role in the formation of public opinion. The way people access news and informational content, thus forming their political and social views, is now largely mediated by social media platforms. These represent new types of gatekeepers that contribute in decisive ways to shape individual and collective access to information. As the infamous Cambridge Analytica scandal helped reveal, the recommendation algorithms that constitute the backbone of social media spaces play a key and largely unaccounted role in filtering personalised news content to users (Pariser, 2011). Yet, their actual impact in the ways in which individuals conceive and make sense of informational content remains difficult to accurately assess. In particular, it is difficult to account for how much users know about algorithms and their role in this context. On the one hand, knowing how algorithms work and understanding the way their outcomes are produced is imperative if we want to understand what informational sources users are and are not able to access. At the same time, knowing more about the ways in which users approach and act in relation to algorithmically-circulating or algorithmically-created content is essential in order to sustain a call to make algorithmic processes on social media more accountable to the general public. The **ALGOCOUNT**¹ research project, within which this literature review originates, is primarily concerned with this task.

The scope of this literature review is to map the existing academic knowledge about the relationship between algorithms and the formation of public opinion. In the pages that follow, we discuss what we consider to be the key theoretical and empirical research on algorithms and public opinion to date, bringing together the various scholarly strands that inform this debate, from critical media studies to political communication, digital policy and journalism studies, and present the state-of-the-art of this lively research field. The overarching goal of this review is to provide the baseline for the conceptualization of what we call an ‘algorithmic public opinion’: with this term we intend the algorithmically-driven process by which a certain issue becomes a salient matter of public opinion, and the central role algorithms play as the gatekeeping infrastructure through which individuals access information, produce their opinions and consolidate their social and political views. From the the moment it comes into being, to when it reaches the wider public, information is prioritised, filtered and hidden (Pariser, 2011; Bozdag, 2013) across a thick mixture of elements that come together in the algorithmic infrastructure of social media and digital platforms, and that involve user behaviour as well as third mediating parties - known as data brokers. From the interaction among all of these elements, we contend, an algorithmic public opinion

1 www.algocount.org

emerges – one that is unavoidably affected by the biased nature of technology (Friedman Kahn, Borning & Huldgtren, 2006) and its affordances, that concur to the extreme personalisation of online activity and the salience of issues of disinformation and misinformation, partial or partisan information environments and situations of psychological and political polarisation (Settle, 2018; Dylko, Dolgov, Hoffman, Eckhart, Molina & Aaziz, 2018).

The emergence of an ‘algorithmic public opinion’, we maintain, bears huge social and cultural implications that require researchers to ramp up their efforts in expanding the existing understanding of algorithmic processes and the cultural conceptions surrounding them, without stopping at the ‘unknowability’ of black-boxed code. Beyond the little public knowledge about algorithms such as Facebook’s News Feed, or YouTube’s Related Videos, or the Google Search algorithm itself, in order to understand their relevance in the processes of public opinion creation there is a necessity to know these as social and cultural objects first. The perceptions, opinions and understandings of algorithmic interventions in day-to-day information consumption and content filtering from the side of users matter as much as knowing about the code and mathematical formulations of these algorithms. The fallout resulting from the already-mentioned Cambridge Analytica scandal, and the publication of a variety of documentaries that presented an exposè of the internal workings of social media platforms in relation to data management, content moderation and ethics, has given new space to discuss about reducing the opacity of algorithmic recommendation systems and enhancing their transparency. Through this research, we pursue the overarching goal of seeking to improve our understanding of how algorithms intervene in the social fabric of Western democracies.

Our chief focus is on recommendation algorithms in the social media sphere. We contend public opinion formation processes in the digital society have become ‘natively’ algorithmic. Recommendation algorithms constitute the often invisible, but largely inescapable infrastructure through which all informational content circulates in the saturated information environment of the 21st century. As algorithmic forms of organization, circulation and access to informational content innervate the formation of public opinion at all levels, from the media industry to individual everyday lives, the ways in which different kinds of information are cognitively processed by individual users in their social lives change accordingly. Our ambition with the Algocount project is to contribute to an improvement in the social and cultural understanding of these processes, and to question their increasing relevance.

In this review we focus on 5 main areas of inquiry. In the first section, we present the debate on algorithms in the context of digital media research, which has coalesced in a dedicated strand of scholarship known as ‘critical algorithm studies’. This body of work has focused on the understanding of algorithms as technological, social and cultural objects, from the standpoint of a critical observation of the implications that the presence of algorithms engenders and fosters in a variety of contexts. This is the main strand within which the **ALGOCOUNT** project aims at making a

contribution. Subsequently, we reflect on the specific entanglement of algorithms and public opinion, articulating their ‘complicated relationship’ as it emerges from existing research across different disciplines. In the third section, we turn our attention to the newsroom and the emergent role of algorithms in the context of the journalistic profession. We show that the ways in which old and new, human and technical gatekeepers interface at the level of journalistic work is a key aspect to observe in the relationship between algorithms and public opinion formation at large.

In the fourth section, we look at the methodological challenges that the critical study of algorithms and public opinion entails. We discuss how existing research has approached the study of algorithms and highlight the potential ways forward in this endeavour, paying close attention to public opinion issues. Relatedly, in the fifth section we turn our attention to how to visualize algorithmic interventions, seeking to understand how emergent data visualization practices have helped, and can further help, knowing about algorithms from a social and cultural perspective. Finally, in the Conclusion we discuss limitations and boundaries of this review, and reflect on the broader issue of algorithmic awareness and accountability that the *algotcount* project comprehensively aspires to address.

Understanding algorithms. Key concepts

Algorithms are a recent, but all the more significant element of concern in the study of society, culture, and the economy. Research on algorithms as conceived in this literature review exists at the intersection of a variety of scholarly disciplines, including computer science, digital media studies, cultural and digital sociology, science and technology studies, and social policy. In this context, a new strand of research, known as ‘critical algorithm studies’, has arisen, which is purported to critically observe the emerging relevance of algorithms in a variety of domains and applications, with a peculiar focus on controversies and inequalities (**Lomborg & Kapsch, 2020**). This is cognate to another emergent stream of research, that of ‘critical data studies’, with which it partially overlaps (**Kitchin & Lauriault, 2014**).

Overall, the problematization of algorithms as an object of research has so far prioritized the aim to a) inquire what algorithms do; b) investigate the potential forms of bias, surveillance, inequality and societal disruption the diffusion of algorithms and related forms of technological innovation bring; c) question the opacity of algorithmic elaborations and the lack of transparency and public accountability surrounding them. This has developed across three main lines of inquiry. A first one conceives of algorithms as technical objects. From this perspective, algorithms are primarily understood as infrastructures and objects of mediation in the tradition of media research. As a result, ‘algorithms’ broadly intended are taken here as the digital infrastructures underpinning and enabling the

workings of the most important social media platforms, such as Facebook, YouTube, or Instagram. In this context, algorithms have been described as ‘engines of order’ (Rieder, 2020) whose peculiar capacity consists in the work of categorizing, filtering, sorting, ordering and hierarchizing content. Conceived as such, algorithms “play an increasingly central role in selecting what information is more relevant to us” (Gillespie, 2014, p.168). These, Gillespie continues, “need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations”. Upon this basis we can say that algorithms, comprehensively intended, are:

- **Patterns of inclusion**, which define what makes and what does not make part of an index;
- **Cycles of anticipation**, i.e., tools that help predict individual behaviour;
- **Relevance evaluators**, as they discern what is relevant from what is not relevant on the basis of a given query;
- **Promises of objectivity**, which derive from their mathematical nature and henceforth undisputedly assumed to be neutral and fair;
- **Entangled with practice**, as users’ practices constantly reshape the working of algorithms;
- **Producers of calculated publics**, insofar as they give a digital public a perception of its own existence and articulation.

(based on and elaborated from Gillespie, 2014).

A rich body of literature has developed over the years around this conceptual baseline. In particular, many scholars have raised a critique about the unknowability of how the aforementioned processes actually take place. This has been popularised through the already-mentioned metaphor of the ‘black box’ (Pasquale, 2015), which grasps the opacity and lack of transparency about algorithmic elaborations and the secrecy that digital media companies impose around them. Nonetheless, it has also been noted that knowing how an algorithm works is often impossible at a practical level for many researchers, for four main reasons: a) intentional secrecy by the companies creating and owning them about the computational codes employed, treated as key economic information; b) technical illiteracy by the researchers who study them, who are often trained in disciplines that historically require little (if no) computer science expertise; c) unintelligibility, because of the evolution of the very same algorithmic output as a result of machine learning processes; d) size, which makes it impossible to infer the workings of an algorithm due to the lack of machinery power (Burrell, 2016; Christin, 2020b).

Despite (and to some degree as a result of) the unknowability of algorithms as technical objects, other research has made a significant effort in trying to counter algorithmic opacity, most commonly using a combination of issue-based digital methods and reverse engineering (cfr. SEC-

TION 4). Research from this perspective has attempted at questioning the technical functioning of algorithms in contexts such as, among others, music (Airoldi, Beraldo, Gandini, 2016; Airoldi, 2021) and music streaming services (Eriksson et al., 2017), digital influencer economies (Bishop, 2018) and data-driven advertising models (Pybus, 2019), just to name a few. The predominant aim of this emergent stream of research is the deconstruction of the apparent neutrality and objectivity of machine learning processes (Airoldi, 2021; Mackenzie, 2019), unveiling the hidden role played by human agency and knowledge (Bechmann & Bowker, 2019). Simply put, machine learning processes consist of ‘computers that learn from experience’, meaning algorithms that are given a dataset of historic events upon which to apprehend and, in turn, use to identify patterns in new data (Airoldi, 2021). The outputs of machine learning algorithms depend on properties and inputs that are often unassumed but can have relevant consequences on the robustness of the output (Doshi-Velez & Perlis, 2019). Furthermore, the nature and scopes of machine learning algorithms themselves bring developers to concentrate their attention and efforts only on the optimization of predictive performances (Enni & Herrie, 2021), favoring an unreflexive and uncritical implementation. This myopia obscures many equally relevant aspects, such as the reproduction of existing societal biases and discriminations (Veale & Binns, 2017). Other researchers, following the path of considering machine learning highly dependent on human agency, are starting to analyze the ‘cultural life of machine learning’ (Roberge & Castelle, 2021) and to theorize machine learning algorithms as cultural objects (Airoldi, 2021) with their own morality (Jaton, 2021).

The study of algorithms inevitably intertwines with research on big data (Ruppert et al., 2013; Lupton, 2014), digital platforms (Poell et al., 2017), affordances (Marres, 2017) and metrics (Beer, 2017; Gerlitz & Helmond, 2013) and the roles these individually and simultaneously hold in the intermediation of social, economic and cultural processes. These entities may be considered, respectively, the raw matter of input (data), the milieu (platforms), the enablers (affordances) and the byproduct (metrics) of technical algorithmic intermediation. This contiguity concurs to devise a second line of inquiry, which understands algorithms as social objects. From this perspective, algorithms coordinate the establishing of social engagements among users and more generally enable social activity within digital environments (Marres, 2017). Conceived as such, algorithms are to be taken as non-human social actors who, together with human social actors, inhabit social media as social environments whereby different forms of sociality exist and unfold (Caliandro & Gandini, 2016). A substantial amount of research exists on the critical study of recommendation systems and forms of personalisation (Airoldi et al., 2016, Milano, Taddeo & Floridi, 2020; Helberger, 2019; Greene & Shmueli; 2019, just to name a few), on the role of bots and automated content production in social media spaces (Kovic et al., 2018), and on the forms of exploitation of the ‘free labour’ of digital media users in the creation of the content that circulates throughout them (Scholz, 2012; Andrejevic, 2013; Postigo, 2017). A particularly important contribution in this debate is given by the notion of ‘algorithmic identity’, coined by Cheney-Lippold (2011). This underlines how algorithms contribute to the determination

of identity categorizations of populations according to their Internet history (cfr. also **Lomborg & Kapsch, 2020**). In this same context, a rich set of contributions focuses specifically on algorithms as byproducts of tech corporations, their cultures and economic-driven decision-making processes, questioning the externalities these produce and the extent to which these affect public decision-making processes. From this point of view algorithms are conceived as everyday forms of artificial intelligence that embed ‘opinions in code’ (**O’Neil, 2016**) and may contribute to the reproduction of social inequality. Pivotal research has specifically focused on inequality and bias surrounding facial recognition systems (**Eubanks, 2018**), forms of algorithmic surveillance (**Amoore, 2020**) and racism (**Noble, 2018**). Relatedly, a rich stream of research has highlighted the issues surrounding the role algorithms play in online content moderation, and the complicated interplay between automated and human moderation work (**Roberts, 2019; Gillespie, 2017; Gorwa et al., 2020**).

A third line of inquiry focuses on algorithms as cultural objects. Key contributions in this debate are concerned with the implications that originate from the intervention of algorithms in the context of cultural production. Research in this context has highlighted the emergence of an ‘algorithmic culture’ (**Striphos, 2016**) and of related processes of ‘platformization’ of cultural consumption (**Duffy et al., 2019; Poell et al., 2021**). From this perspective, algorithms are deemed to re-mediate the organization, circulation and access to cultural content; as a result, they constitute key components of cultural inquiry and analysis. The study of algorithms as cultural objects also involves the epistemological and methodological terms of inquiry; the unknowability of algorithms as technical objects does not have to extend to the cultural context surrounding their interventions, which can - and must - be researched (**Seaver, 2017; Christin, 2020b; Bonini & Gandini, 2020**). Particularly, it has been questioned how, when and to what extent digital media users perceive, understand and reflect back on the algorithmic intervention in their everyday practices of use of digital service. Pivotal in this context is the work of Taina Bucher (2017, 2018), who has theorized the existence of certain ‘algorithmic imaginaries’ that users of a given platform hold and develop in relation to their user experience of certain (recommendation) algorithm - e.g. the Facebook News Feed. Bucher’s research on algorithmic imaginaries represents a foundational inspiration for the present work.

Algorithms and public opinion. A complicated relationship

Beyond the general observation of algorithms as technical, social and cultural objects is the specific inquiry on the role of algorithms in the political arena. This must be intended in two complementary nuances: a first one concerns the all-out relevant intervention of algorithms in political debate; a second one sees algorithms as a broader problem of the ‘polis’, in its Greek sense, thus intending an emergent issue concerning civic society at large.

It is building on this understanding that we critically observe the relationship between algorithms and public opinion. The extent to which algorithms intervene in the formation of public opinion remains a grey area, as it is often made object of (appropriate) criticism as a form of technological determinism (Moeller & Helberger, 2018). Furthermore, as said, the degree of knowledge and understanding about the workings of algorithms by the general public also remains largely limited. Yet, in line with the aforementioned research on algorithms as social and cultural objects, this should not discourage researchers from questioning the cultures surrounding algorithmic intervention in public opinion formation.

We problematize the relationship between recommendation algorithms and information circulation through the notion of the ‘algorithmic public opinion’. This has the objective of promoting the understanding of the role algorithms play not just within society, but also by society. Since the emergence of social media, the role of digital technologies of communication in the processes of public opinion formation has been a key concern for scholars and experts. Social media have been shown to play a significant role in political contention, from fostering ‘connective’ action of political publics (Bennett & Segerberg, 2012) to facilitating ‘citizen journalism’ practices (Robinson et al., 2009). This occurs in the context of an ecosystem of communication driven by ‘affective’ logics whereby social media, albeit not ‘producing’ uprisings in a deterministic way, arguably facilitate new forms of engagement and participation (Papacharissi, 2015). Throughout the years, this has coupled with increased concerns about how a digitally-mediated public sphere might ultimately foster opinion polarisation and sow division, instead of expanding the debate among different viewpoints (Tucker, Guess, Barberá, Vaccari, Siegel, Sanovich, Stukal & Nyhan, 2018; Sunstein, 2017). These concerns arise from the fact that today’s unprecedented availability of media sources is able to offer citizens a wide array of heterogeneous information - which is, however, tailored to specific interests and appeals to delimited audiences. This detailed targeting of narrower and ‘niche’ publics enables a customised consumption of news that may lead to audience fragmentation as opposed to building a common ground for cooperation.

The fragmentation of the public agenda, together with the physiological state of arousal that some social networks seem to provoke (Mauri, Cipresso, Balgera, Villamira, Riva, 2011), can result in the psychological polarisation of Internet users, who would then strengthen their identification with their in-group and exacerbate the differences with the out-group (Settle, 2018). These discourses have encouraged researchers to look for connections between the usage of social media networks, the affirmation of radical right-wing parties (Bennett & Livingston, 2021) and internet subcultures (Marwick & Lewis, 2017). The rise of populist parties has also been associated with social media in light of an ‘elective affinity’ between social media and the populist message, with social media being able to convey ‘the voice of the people’ vis-a-vis mainstream media, perceived as representatives of the elites (Gerbaudo, 2018).

Within this context, the ways in which algorithmic infrastructures contribute to channeling, facilitating or hindering such processes of public

opinion formation, as well as the users' awareness with regards to these technological processes, has been rendered an urgent matter following the Cambridge Analytica scandal, which has put under question the outcome of the 2016 Brexit referendum and the election of Donald Trump as President of the United States in the same year. In both cases, personal Facebook data have been deceptively obtained by a private corporation – Cambridge Analytica – which has used these to construct 'psychographic profiles' of users who were then targeted on the same social media with personalised political advertising designed to leverage on emotional triggers and individual personality traits (Cadwalladr & Graham-Harrison, 2018). The success of this operation was ultimately granted by the workings of the algorithmic elaborations of Facebook News Feed that are known to, at a basic level, offer users content based on their existing preferences and prior behaviour on the platforms.

Yet, many users remain largely unaware of the specific ways in which this algorithmic dynamic works, as this has long remained away from public scrutiny.

Among others, Swart (2021) notes that young users who access news via social media showcase a lack of knowledge about how social media algorithms operate. This, Swart suggests, together with their opacity and limitations in technological vocabularies, hinders the capacity of young people to articulate their "experiences" of algorithmic encounters.

The formation of an 'algorithmic public opinion' - an 'everyday' public opinion that is necessarily shaped by the algorithmic infrastructure of social media and digital platforms - is a particularly important matter of concern if we consider that social media and digital resources now count as the main means to access informational and news content by the general public, especially for young demographics. In the US, for instance, the Pew Research Centre reports that, in 2018, 43% of users accessed social media as their main news source – a mere 7% points less than television – while 67% declared to consume news content on social media 'often'. The same report also reveals that a growing percentage of users access news content using a mobile device (85%), showing in other words the increasing relevance of social media platforms and their algorithms as 'gatekeepers' in the access to information (Smith, 2018). It is reasonable to expect this data can only increase as the pervasiveness of social media platforms in information circulation processes further extends.

In particular, algorithmic infrastructures intervene significantly in this picture as they coordinate the access to news content by users. To describe this process, Eli Pariser (2011) has coined another popular metaphor, that of the 'filter bubble'. This term grasps how the circulation of personalised content facilitated by (recommendation) algorithms on social media platforms prevents access to diversified sources, therefore leading to the formation of informational 'bubbles' within which each user is exposed to essentially the same type of content, suffering the absence of different viewpoints. In this context, opinions come into being in a peculiar environment without the influence of disagreement. This poses both a theoretical and an empirical threat to the ideals of democracy, according to which individuals have the right to be fully informed,

acknowledge and contest any form of political oppression, be heard when part of a minority, and negotiate their interests and values in a moment of deliberation (Bozdag & Van Den Hoven, 2015).

This is further complicated by the proliferation of deceptive digital content, especially of a political nature, commonly known through the term 'fake news'. More appropriately, this kind of content should be considered, broadly speaking, as disinformation. Disinformation is intended as fabricated or altered informational content that is perceived as true by users, and debated or shared accordingly. This is intentionally contentious content, designed to cause public harm or political profit by discrediting opponents and influencing voters (Disinformation, 2018, p. 3). Research has demonstrated the relevance of 'fake news' in the formation of public opinion around the 2016 US presidential election (e.g. Allcott & Gentzkow, 2016). In fact, the issue of 'fake news' was so contested on that occasion, that the very term became a mediatic war to win the public's complicity - for example, it was used against American right-wing politics, and then by American right-wing politicians as a critique to mainstream media (Farkas & Schou, 2018). This epistemological battle resulted in the overflow of extremist, conspiratorial and other forms of junk news across social media platforms such as Twitter, where this type of deceptive information ended up representing the larger portion of all political content being shared (Howard, Bolsover, Kollanyi, Bradshaw & Neudert, 2017). Other research instead has pointed to an excessive emphasis on the relevance of fake news; for instance, in the context of the debate concerning the Covid-19 pandemic in Italy, a study by Caliandro, Anselmi and Sturiale (2020) has shown that only a fraction of content circulating online can be considered as fake news, and warned that its importance is somewhat exaggerated.

The circulation of fake news combines with the role played by bots in political campaigning. Bots are algorithms that automatically share content on a large scale, posing as 'real' human users on social media. The use of bots in political campaigning is argued to pose a threat to the formation of a free and unbiased public opinion as a result of the amount of disinformation they are able to convey. Looking again at the case of the 2016 US election, Bessi and Ferrara (2016) show that algorithmic bots accounted for one-fifth of the entire online debate around the election, and warn about the risks this poses for the integrity of electoral processes. Other research suggests that disinformation is amplified when backed by automation and promoted in a preprogrammed way, but it is actually disseminated by only a minority of accounts. In a similar study to that of Bessi and Ferrara (2016), Howard and Kollanyi (2016) notice that although bots had generated a good portion of Twitter traffic about Brexit during the UK referendum on EU membership (almost $\frac{1}{3}$), they physically amounted to less than 1 percent of the sampled accounts. Likewise, Caliandro and colleagues (2020) explore the processes of circulation of fake news within the Italian Twittersphere during the Covid-19 emergency, and conclude that disinformation seems to be circulating only in very specific periods and within closed communities - yet, the World Health Organization speaks of Covid-19 as an 'infodemic', i.e. an epidemic of misinformation. In this light, bots represent a threat to the social order in-

so far as they constitute a means to both disseminate disinformation and expand its reach, but also as a way to alter online users' perception of how polarized their fellow companions are on social media networks.

The immediate perception of extreme and untrue information circulating more widely because of digital media, and in a more personalised manner because of algorithmic logics, ends up creating the idea of radicalised publics that operate via different information logics and that escape the conventions of democracy and political communication (**Bennett & Livingstone, 2021**). This kind of concern has led Zeynep Tufekci (2018) to warn, in an op-ed for the New York Times, that algorithmic personalization of news content poses a serious threat to the social order, in that it creates the condition for what she describes as actual forms of political radicalization engendered by the consumption of increasingly extreme content, offered automatically for users to consume. More recent research has seemingly disproved this warning, showing with empirical data that radicalisation dynamics are not necessarily an outcome of social media content circulation (**Markmann & Grimme, 2021; Hosseinmardi et al., 2021; Ribeiro et al., 2020**) but rather a more nuanced process. Research also suggests that political polarisation stems from other variables that can precede online activity - e.g., partisanship, which is acquired at a very young age, as well as the historical background of the country one resides in (**Kreiss, 2021**). This casts doubt on the causal effect of the Internet in provoking extreme attitudes and widespread disinformation (**Iyengar, Lelkes, Levendusky, Malhorta, Westwood, 2019; Kreiss, 2021**). Yet, it also reinforces the necessity to open up the debate around the level of knowledge that online users have about algorithmic systems, and the role this awareness might play in mitigating - or, on the contrary, in emphasising - one's radicalisation or gullibility (see **Caplan and boyd, 2016**).

Algorithms have also been found to produce unwarranted perceptions about the importance of some news articles over others, with the ultimate risk of promoting political, racial, class and other discrimination (**Eslami et al., 2015**). As the role of search algorithms as informational gatekeepers becomes clearer, focusing attention on the users' awareness of these mechanisms and their sense of how these guide their online experience gains further relevance. However, while the literature on the risks and dangers of the affirmation of an algorithmic public sphere is growing, much less has been written about the perception of the role of algorithms in the formation of public opinion by the general public. Graefe et al. (2018), for instance, assessed the capacity of users to distinguish if a news content has been written by a human or an algorithm, concluding that the quality of algorithmically-generated news today is comparable to that written by a human, thus leading to concerns in terms of perception. Yet, more research is certainly required on this front.

In general, despite its ubiquitousness it seems that a majority of users still lack a clear understanding of how the Internet operates. In the aforementioned Pew Research Center US-based survey, that investigated people's perceptions of how the Facebook News Feed worked, it is revealed that more than half of the sampled respondents had little information and perceived little control, or had never tried influencing the content

they were fed (Smith, 2018). In a similar research, though administered to a much smaller sample, Eslami and colleagues (2015) also found that only a small portion of users knew their Facebook Home Feed was filtered. The same insights emerge in Gran and colleagues' research on users' algorithmic awareness (2020), which turns out to be very low, especially among the more demographically mature segments of their sample. Other research shows instead that users do have some (and in cases, an accurate) perception about content being excluded from their Facebook News Feed, although the reasons they provide for why that is or for how it happens are mostly confused (Rader & Gray, 2015). In this regard, the aforementioned notion of 'imaginaries' (Bucher 2017, 2018) proves to be an essential heuristic to the aim of grasping the understandings and expectations of users in relation to the functioning of algorithmic infrastructures and online content recommendation and personalisation processes. As Messing and Westwood (2014) suggest, opinion polarisation resulting from the consumption of online news content is only one (key) concern in a broader issue that entails a fundamental change in the context in which information access occurs - which many still find difficult to make sense of.

When Algorithms enter the Newsroom

This set of issues, however, cannot be comprehensively addressed without taking into consideration the changing role of the journalistic profession in the social media era. In the contemporary debate on journalism it is often discussed what journalists do with data, but it is commonly overlooked what data have done – and do – to journalism and journalistic practice. In this section we present a reflection on the incorporation of data into everyday practices of contemporary journalistic production, and its relationship with algorithmic elaborations. In so doing, we devise a typology aimed at illustrating how data analytics, metrics and algorithmic elaborations influence various forms of journalistic everyday work.

The seminal studies in the field on the relationship between data and journalism are those by MacGregor (2007) and Dick (2011). MacGregor examines the way online journalists react to new ways of knowing about their audiences. This research analyzed emergent narratives on journalists' use of tracking data from website servers, suggesting that tracking data are changing the way 'news values' are implemented. Dick's (2011) work instead represents an exploratory study into what he defined as 'an emerging culture' in UK online newsrooms: the practice of Search Engine Optimisation (SEO), of which he assesses the impact on news production. Dick concludes that SEO policy is found, in some cases, to inform editorial policy; nevertheless, there is still considerable resistance to the principle of SEO driving decision-making.

By presenting a more holistic approach, Anderson (2011) highlights how early newsroom sociology emphasized the submersion of audience-centered news judgments in favor of professional codes, while the increasing amount of research on the relationship between journalism

and data documents a process by which ‘deciding what’s news’ is increasingly influenced by quantitative audience measurement techniques (Anderson, 2011, p.551). As Carlson (2018) suggests, it is important to point out that journalism has always tried to measure its audience: it did through sales data as well as through audience data. In our contemporary media ecology, however, this measurement has taken different and more pervasive characteristics, returning real-time and personalized data about news consumption by the public. Additionally, it is worth mentioning that this relevance of metrics and data is also due to the fact that the journalistic context is impinged by a serious economic crisis and is in a constant search to find new business models for making journalism sustainable. Data are thus often seen as a primary avenue towards those models. This has brought about the notions of ‘calculated’ (Gillespie, 2014) and ‘algorithmic’ publics (Christin, 2020a), intended respectively as publics that do not exist until they are calculated and, relatedly, the granular representation of audiences in metrics, which have affirmed as important heuristic in the cross-disciplinary research on datafication and journalism (Moller-Hartley et al., 2021).

Beyond the continuously increasing buzz about data-journalism (see Coddington, 2015; Splendore, 2016), journalists and newsrooms essentially deal with three key aspects: algorithms, analytics and metrics. Algorithms in this context are configured as automated distribution selection and control mechanisms of information (Boccia-Artieri & Martinelli, 2018). Journalists see the output particularly in the positioning of news in their timelines or looking at results provided by Google. This combines with metrics and analytics; Zamith (2018), by discussing the proliferation of audience analytics and metrics in journalism, distinguishes between audience analytics (systems that capture information) and audience metrics (quantified measures output by those systems). This comprehensive datafication process applied to journalism manifests itself in four forms:

1. Journalistic outcomes are increasingly a byproduct of algorithms in terms of positioning in the search engines and in the timelines of social platforms.

This aspect is the best known and most debated one (Dick, 2011; Tandon, 2014). It has to do with widespread practices of search engine optimization (SEO) and the precautions that every editorial team takes to facilitate the circulation of articles through search engines and social platforms.

2. The metrics of engagement with journalistic content are available to everyone, including readers: how much an article is shared, how many likes it receives, what is its circulation.

This can have different implications: it can push the reporter to deepen an issue that seems to interest her readers in terms of sharing, but it may also consolidate an interpretative framework that captures the favor of the public in terms of shares and likes, or push journalists to adapt their writing style depending on the criticisms (or compliments) received in the comments (see Wright et al., 2019).

3. Journalistic metrics are the product of a larger analytics processing system that monitors individual and aggregated behavior (i.e. Google Analytics, Charbeat, Newsbeat, Parse.ly).

Following MacGregor (2007), many other studies have posed significant attention on the relevance of analytics and metrics in journalistic work. Lee and colleagues (2014), through longitudinal secondary data analysis of three U.S. online newspapers, found that (a) audience clicks affect subsequent news placement; (b) such influence intensifies during the course of the day; (c) there is no overall lagged effect of news placement on audience clicks (i.e. audiences are more influential on journalists behaviour than editorial choices on audience); (d) the lagged effect of audience clicks on news placement is stronger than the inverse. These results are revealing of a media ecology that is paradoxically less complex than commonly thought, whereby audiences rarely go to newspaper homepages and have more fragmented trajectories of news consumption. As argued by Diakopoulos (2015), the new media ecology has led to an increase in directly-related metrics and the growth of digital platforms in which the act of consuming news generates a different transmission of data. Those data are managed by different tools newsrooms may use (Omniture, google Analytics, Facebook Insights, and Twitter Analytics). Many organizations supplement them with their own home-grown analytics tools (see also **Cherubini & Nielsen, 2016**).

4. Journalistic metrics as a product of behavioral processing of networked users (such as NewsWhip, Crowdtangle, Ezyinsights).

Finally, there are softwares - not infrequently the same ones we have mentioned above - who monitor what is happening on the Internet: they keep track of social media signals, monitor tweets, shares and comments. The reception of the tools presented in this typology varied from country to country, from one editorial staff to another, and one journalist to another; at the same time, what is happening where they have already been introduced is their continuous normalization (**Lasorsa et al., 2012**). As Zamith (2018) contends, these four levels have an influence across the ABCDE of news production: attitudes, behaviors, content, discourse, and ethics. This brings to the conclusion that algorithmic elaborations have had a consequential impact on the ways in which the gatekeeping profession par excellence - journalism - has adapted to the digital era. Importantly, these issues reflect in the ways in which users encounter and consume journalistic content as it gets re-mediated by social media and digital platforms, and in turn significantly concur to the development of an algorithmic public opinion as here conceived.

How can we Study the Algorithmic Public Opinion?

The study of algorithms represents a significant challenge also from a

methodological perspective. As said, this review starts from the assumption of the relative unknowability of algorithms as objects of research. This is an obstacle at a technical, social and cultural level (Burrell, 2016); however, this has not stopped researchers from engaging in multiple empirical investigations of algorithms and their functioning, experimenting with new methods and approaches. In the majority of cases, these exploit the relational nature of algorithms (Ash, Kitchin & Leszczynski, 2018; Lomborg & Kapsch, 2020); prevented from accessing the black box, the inquiry focuses onto the relationships between algorithms and individuals. This approach leverages on the notion that algorithms form a complex socio-technical infrastructure embroidered with human agency (Dogruel, Facciorusso & Stark, 2020) both at the input level (programmers) and at the output level (users). In this sense, users are not conceived of as passive actors: rather, they possess an agency that can influence the algorithmic functioning (Wang, 2020) and might bring them to develop strategies to circumvent the algorithmic intermediation (Gerrard, 2018).

A first established method of algorithmic inquiry is rooted in the tradition of ‘digital methods’ research (Rogers, 2013), which has affirmed as the main pathway to researching social media and digital platforms over the last decade. Digital methods exploit the native technical features of digital platforms, seeking to ‘follow the medium’; typically starting from a given issue or controversy (ibidem), digital methods use the infrastructural features of social media as instruments for data collection and, in some cases, analysis. Digital methods also allow cross-platform analysis, obtaining valuable insights on the reception of platforms through the direct study of another platform as medium, something that can result to be particularly useful in case of de-platforming of users’ categories (Rogers, 2020). Furthermore, digital methods allow us to consider that platforms often show forms of ‘tiered governance’ (Caplan & Gillespie, 2020), differently modulating the algorithmic functioning for different users’ categories. Overall, they represent an important avenue for social research as far as they allow to study the interplay of social relations and digital affordances (Marres, 2012). They enable to expand the existing knowledge about how certain algorithms work but have also been subject to various challenges, including first and foremost the limited ‘openness’ of platforms to scrutiny. It is frequent, in fact, that digital platforms arbitrarily modify the API keys or their terms and conditions in order to reduce, if not prevent altogether, the collection of data by researchers (Caliandro & Gandini, 2016). Thus, the feasibility of the digital methods’ approach depends upon the platforms’ decisions to restrict access to the API (Bruns, 2019) and the researchers’ ability to overcome the restrictions (Bainotti et al., 2020).

Complementary to the digital methods approach is the one of ‘reverse engineering’, that seeks to look at “what data is fed into an algorithm and what output is produced” (Kitchin, 2017, p.24). Reverse engineering may or may not make use of digital methods techniques; in general, it represents a viable option yet an equally limited one, as it only relies on the input and output to assess and speculate on something – the actual algorithmic working – that remains away from a more comprehensive scrutiny, safely in the hands of commercially-driven private companies that treat their codes as ‘trade secrets’. The reverse engineering approach

comprises of a vast spectrum of different methods – also from different disciplines - such as accountability reporting (Diakopoulos, 2015) and the ‘walkthrough’ method (Light, Burgess & Duguay, 2018). Accountability reporting methods stand out from the others because these are derived from journalistic investigations seeking to reverse engineer the input-output structure of an algorithm (Diakopoulos, 2015). These share the same goal and approach but with a different audience and tone, due to the journalistic approach and the non-academic venues. The ‘walkthrough’ method instead is specifically designed for the study of apps, but the methodology can be applied to digital interfaces more in general. Drawing upon Actor Network Theory (Latour, 2005) and the concept of digital affordances (Bucher & Helmond, 2018), the ‘walkthrough’ method constitutes a systematic and autoethnographic exploration of the app by the researcher, inspecting how the socio-technical infrastructure of the app influence the relationship between user and the algorithm.

Reverse engineering operates at the core of the concern about what users think algorithms do. A variety of literature suggests that users make their own experimentations in guessing the workings of algorithms, by putting in practice rudimental forms of reverse engineering on the basis of their own online activity. The corpus of research focused on these aspects largely relies on traditional and well-established qualitative methodologies like ethnographies (Bishop, 2019; Cruz & Harindranath, 2020), semi-structured interviews (Dogruel et al., 2020; Lomborg & Kapsch, 2020; Woodruff et al., 2018), focus groups (Siles, Segura-Castillo, Solís, & Sancho, 2020) and, more rarely, on surveys (Gran, Booth & Bucher, 2020; Just & Latzer, 2017), experiments (Lee, 2018) or mixed-methods approaches (Shin & Park, 2019). In some cases, a digital methods or ‘walkthrough’ approach was also combined with more traditional qualitative methods (Caplan & Gillespie, 2020; Duguay, Burgess & Suzor, 2020). These approaches generally follow a complementary logic, using digital or walkthrough methods to analyse the platform or media produced on a platform and combining these with semi-structured interviews to gain knowledge on the users’ perspectives. For example, Caplan and Gillespie (2020) combined the visual analysis of 90 Youtube videos collected with digital methods tools with semi-structured interviews to 6 Youtube video creators and a discursive analysis of youtube partners program policies. Similarly, Duguay et al. (2020) combined a walkthrough analysis of the Tinder app with 20 semi-structured interviews to female, queer Tinder users.

Recently, beyond qualitative methods combined with reverse engineering we have witnessed a revival of ‘pure’ ethnographic approaches and strategies to study algorithms (Christin, 2020b; Lange, Lenglet & Seyfert, 2019; Seaver, 2017). These advance ethnography as a particularly suitable method for the study of algorithms as it allows to overcome their peculiar opacity. Seaver (2017) proposes a set of tactics to study algorithmic enactments, such as ‘scavenging’ data eclectically from a multiplicity of dispersed sites, to reconceptualise access to the field as a texture whose study is insightful in itself, including interviews as a substantive part of the fieldwork. Christin (2020b) instead focuses on algorithmic enrollments, developing three meso-level strategies: algorithmic refraction,

comparison and triangulation. Algorithmic refraction means to study algorithms in their interactions with the surrounding social networks and institutional structures, considering them as a prism rather than a black box. Algorithmic comparison suggests a cross-ethnographic analysis, studying similar algorithmic typologies in different sectors or different platforms. Algorithmic triangulation refers to the enrollment of algorithmic systems by employing the three pivotal ethnographic features of saturation, reflexivity and disengagement. Similarly, Lange et al. (2019) in an ethnographic study of financial algorithms employ Michel Serres' concepts of quasi-object and quasi-subject to account the dynamic nature of the human-nonhuman relationship between actors.

Finally, two other methods stand out for their originality and deserve specific mention. These are the “mechanological perspective” by Rieder (2020) and the “post-phenomenological method” by Ash et al. (2018). The mechanological perspective by Rieder, grounded in his experience as digital methods researcher, focuses on the software as object of study to overcome the black box obstacle, to be analysed through specific strategies of inquiry. In ‘Engines of Order’ (Rieder, 2020) he makes the example of the Google search rank and the Facebook News Feed. How Google concretely orders search's results is impossible to know, but it is possible to gain critical insights on its functioning by reconstructing the history, content, and substance of PageRank. Similarly, how Facebook filters the content seen by users on their News Feed is unknowable, but the researcher can study the origins and the functioning of the machine learning algorithms that also govern Facebook's News Feed filtering, to shed light on the black box. The ‘post-phenomenological method’ by Ash et al. (2018) is instead grounded in a cultural geography approach. It posits that digital interfaces can be analysed in their appearance as objects through the analytical concepts of unit, vibration, and tone, drawn upon sound and sonic geographic studies. In this method, the overall sensorial experience of the individuals becomes the basis to unpack the digital interface and make the algorithm's functioning visible.

All the exposed approaches share the same intuition, that is to exploit the relationship between users and the platform to reflexively gain knowledge on the algorithms otherwise hidden by the black box. However, the multiplicity of strategies they embrace lead to relevant differences in their focus. Algorithmic research based on digital methods typically analyses users as digital avatars on a platform. In other words, they observe accounts' behaviour in the context of a specific platform and its relationship with the algorithms regulating its functioning. In doing so, digital methods allow to focus more thoroughly on algorithmic power, and consider the ability of algorithms to shape the user experience or behaviour on the platform or a certain development of a debate or of a community. Instead, reverse engineering approaches – albeit comprising an array of different methodologies - generally analyse the relationship between the user and the platform in terms of inputs and outputs. Thus, their focus becomes the usage of a platform by the users (and in some way, the usage of a user by the platform), looking at the reciprocal and bidirectional influences. Lastly, ethnographic and survey approaches appear instead to be particularly employed to understand users' subjectivities. As such,

they focus on how individuals perceive and conceptualise the algorithms governing their digital experiences.

The emerging academic literature on algorithms is also developing a specific methodological vocabulary. As occurs with every topic that is rapidly gaining wide relevance and attention, the debate on algorithms is characterised by the flourishing of a multitude of terms, not always easily distinguishable one from the other, to identify specific aspects. In this brief review, only the ones most tightly connected with algorithmic public opinion are included. A first relevant strand of analytical conceptualisations refers to the ways in which users mentally build algorithmic representations and reconstruct algorithmic functioning mechanisms. As mentioned, informed by the notion of socio-technical imaginaries by Jasanoff and Kim (2015) and the actor-network theory (Latour, 2005), Bucher (2017) coined the idea of ‘algorithmic imaginaries’.

The algorithmic imaginary perspective is based on focusing on the feelings, moods and sensations that users express towards algorithms, and to maintain a reflexive approach that considers the imaginary as an actor that influences the development of algorithms too. Other relevant conceptualisations connected to the same research interrogative are ‘algorithmic folk theories’ (Siles et al., 2020) and ‘algorithmic gossip’ (Bishop, 2019). These two concepts share with that of algorithmic imaginary the focus on perception and users’ agency towards algorithms, but from different viewpoints, that enable to analytically focus on other aspects. Siles et al. envision algorithmic folk theories as ‘ways to enact data assemblages, that is, to bring into being a particular “data ontology”’ (Siles et al., 2020, p. 3). The interest on the enactment of data assemblages brings them to pay particular attention to the determinants that bring some users to embrace a specific folk theory. The concept of algorithmic gossip by Bishop (2019), instead, allows to expand the analysis of algorithmic engagements by underscoring that the perception of algorithm involves not only the dual user – interface relationship, but is a more complex social process in which users cooperate exchanging opinions and perceptions in order to build detailed reconstructions of the algorithm and its functioning.

Connected to this last conceptualisation is the one of ‘algorithmic sociality’ by Wang (2020), that allows to introduce a second research strain which analyses how algorithms intermediate the relationships between users and influence their action on the platform itself. The concept of algorithmic sociality grasps how users play with the algorithm and attempt to shape its functioning, transforming it into a ritual tool in order to achieve desired sociality. Similarly, the analytical concepts of algorithmic governance (Just & Latzer, 2017) and algorithmic management (Lee, 2018) underscore that algorithms organize social activity and shape the users’ perception of reality itself, focusing specifically on the processes of individualization, commercialization and the decrease of transparency (among others) these involve. Existing literature in this stream conceptualises the platform governance enacted by algorithms as a ‘patchwork’ (Duguay et al., 2020) that is highly centred around formal measures and mechanisms but lacking adequate self-assessment of the technocultures

generated by the platforms (cfr. Musiani et al., 2016).

A third relevant strain of analytical concepts is the one dealing not on the perception but on the assessment and evaluation of algorithms. In this context, the concepts of ‘algorithmic awareness’ (Gran et al., 2020) and ‘algorithmic fairness’ (Woodruff et al., 2018) are pivotal. Algorithmic awareness pays attention towards how much individuals are informed about the functioning of the platforms they use and the mechanisms beneath them. The employment of the algorithmic awareness concept could prove particularly useful in combination with previous analytical concepts concentrating on the perception and the instrumental use of algorithms by the users, to analyse whether a higher or lower level of algorithmic awareness brings to qualitative or quantitative variations in the algorithmic folk theories/imaginaries depicted by users and their algorithmic sociality. Algorithmic fairness instead calls into question the ethical and political concerns of users towards algorithms and platform management. Highly dependent on the awareness of algorithms’ existence and functioning behind the platforms, that will presumably grow in the following years, the inclusion of algorithmic fairness assessments by users is likely to become a relevant topic in future research.

Visualising algorithms: a critical perspective

Cognate to the methodological challenge of studying algorithms is the way in which algorithms as ‘black-boxed’ objects can be visualised and brought to public view. The visualization of algorithmic interventions remains a gap in the existing research and a fascinating dimension of inquiry that brings together digital media research (particularly the digital humanities) and more traditionally-creative disciplines such as communication design.

Until a few years ago, the use of data visualizations to explain the functioning of algorithms was limited to the field of visual analytics and interpretable artificial intelligence with the original goal of providing expert users (for instance data scientists and model developers) tools for training, comparing, debugging and improving different models. **TensorflowBoard** (Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, 2017), *Manifold* (Zhang, Wang, Molino, Li & Ebert, 2018) and *LSTMVis* (Strobel, Gehrmann, Pfister & Rush., 2018) are good examples of applications for conducting research and for explaining machine learning systems; all of them are addressed to model developers and model builders and exploit visualisation techniques for representing data, features and their relationships. Indeed, data visualisation and visual analytics are considered good means (Offert, 2017; El-assady, Jentner, Kehlbeck & Schlegel, 2019) for explaining internal states of machine learning models.

Yet, if we look at the didactic function of explaining artificial intelli-

gence techniques, it emerges how the most important aspect to visualise is the causal relation between data input and output. Data visualisation and interactive interfaces are used for teaching naïve users how machine learning models react to different data inputs: images (Yosinsky et al., 2017) poses and sounds (GoogleAI, 2017) taken from webcams. For this purpose, new formats for presenting research have been developed, such as *Distill*² that allows users to directly handle interactive parametric visualisations (Watterberg et al., 2017; Olah et al., 2018). Also *ML4A*³ (Machine Learning for Artists) is a noticeable guide for artists who want to experiment with machine learning: demos and diagrams are the preferred tools to communicate notions and experiments.

However, there are currently no standard methods and few success stories of visually communicating algorithmic decisions to the ‘affected’ general audience (Correll, 2019)

From the early 2000s to the present day the visual language of explanations has evolved and articulated, thanks both to technological improvements and the spreading of the topic across disciplines and domains. Indeed, promoting algorithmic transparency, awareness and accountability at all levels has become an important challenge in terms of governance. Moreover, given that the transparency, accountability and explainability of AI towards the lay public is considered problematic (Ziewitz, 2016), it is important to take into account that the level of details, sophistication and content proposed to users must be calibrated according to their ability and attitude to understand.

The complexity of the desired explanation changes depending on the end user and the purpose of the application. Thus, the message can be communicated in different ways, ranging from a transliteration of the process, presenting individual (Ribeiro et al., 2018) or holistic explanations (Offert, 2017), to metaphorical narratives for reducing the complexity of the vocabulary (Jentner, Sevastjanova, Stoffel, Keim, Bernard, El-Assady, 2018; El Assady et al., 2019). In the information design and computational art field it is not uncommon to find projects that try to explain the mechanisms of artificial intelligence algorithms and, as mentioned before, the ways in which the message is built and communicated are various: the explanation is never unto itself but is always included and inferable from the demonstration of how the algorithm produces a certain output. The further away we move from computer science and expert users, the more difficult it is to find explanations about the algorithmic model itself. It is more frequent that the algorithmic model is told through the effect it has, for instance, on a platform or on a specific dataset (Silva et al., 2020)

In the next section, a collection of projects coming from the design and art field are reviewed according to how the message is delivered to the final user, depending on:

- **Dissection tools:** the aim of these projects is to open and dissect the “black box” by visualising the content: hardware, data, lines of code. The overall view is that of a complex but usable set of content.
- **Friction activators:** these types of projects push the user to slow

³ <https://ml4a.github.io/>

² <https://distill.pub/>

down and to think. The absence of extensive descriptions leaves users free to reflect and find connections between AI and reality.

- **Collective gamified experiments:** In these cases the functioning of the algorithm(s) is discernible from the game mechanisms and is partially unveiled at the end using diagrams and illustrations.
- **Exposure devices:** usually “platform centered”, these types of communication artifacts showcase and visualise contents generated by artificial intelligence as they appear. Users can filter, cluster and group them.
- **Awareness triggers:** these can also be understood as campaigns of social activism. Video documentaries and animated graphics tell personal stories of users and explain limitations and overuses of the technology.
- **Annotated Visual Experiments:** the shape of these artifacts is that of a kind of logbook during the application of one or more algorithmic models.

Dissection Tools

Object Centered, Model Centered

Attempts that have been made to visually expose the operation of some algorithmic models can be found in the work done by the ShareLab (*Anatomy of an AI System*³) and by the *Anatomy of Intelligence* artistic research initiative⁵. While Share Lab focuses on specific conversational objects or social platforms (i.e., Facebook), the Anatomy of Intelligence project is focused on the prediction process of Machine Learning.

What these two approaches have in common is the intention of showing the transparency of different algorithmic technologies to the general public, being research projects suitable for museum exhibitions. “The notion of intelligence, especially when human-made but nonhuman, may seem distant and abstract”: Kate Crawford and Vladan Joler, with *Anatomy of an AI System*, propose huge data visualisations and diagrams showing the vast planetary network that underpins the “birth, life, and death” of a single Amazon Echo unit — this project has also been selected for the Broken Nature Exhibition at the Triennale Museum, Milano in 2019.

The main characteristic of these static diagrams and visualisations is that they are conceived as dissection demonstrator devices that show, through a holistic approach, the functioning of the analyzed technology. When looking at these huge diagrams, the viewer is almost overwhelmed by the amount of information. Chicau and Reu, by in-situ dissecting Machine Learning algorithms through real time performances are developing an online repository of terminology and techniques for a critical examination of the “anatomy of learning and prediction processes”. The main characteristic of their demonstration is the real-time dissection where the set of choreography, sounds, images and code plays

7 It is also true that the governance of online content on platforms is a far less multistakeholder than the typical Internet Governance (IG) of internet protocols and standards, with far fewer formalised institutions and fora. To be clear, multistakeholderism is no panacea and civil society has been often marginalised in IG, eventually merely serving to legitimise the process for other, more powerful actors.

4 <https://anatomyof.ai/>

5 <https://anatomiesofintelligence.github.io/about.html>

the main role.

Friction activators

Model centered

Another kind of approach, that is much closer to art, is that based on frictional experiments by Memo Akten on neural networks. These videos show how an artificial neural network makes predictions on live webcam input, and tries to make sense of what it sees, in context of what has been seen and learned before. For instance, a neural network that has been trained only with astronomical pictures will recognise parts of the human body such as starry skies, galaxies and nebulae. Looking at the recordings in parallel the lay user will have to make a cognitive effort to understand how the algorithm could work. Another project of the same author is True Colors⁶.

Collective gamified experiments

Tasks centered

Recently, the Google Creative Lab collected a set of projects of people doing experiments with AI: some of these can be included within the explanations projects. One of the most groundbreaking ones is the Quick-Draw project, that both helps train an algorithm with doodles sketched by users and gives explanations to them according to the relation between data input and output. Semantris, another game designed by Google AI is a set of word association games powered by machine-learned, natural language understanding technology. Each time users enter a clue, the AI looks at all the words in play and chooses the ones it thinks are most related. These types of experiments show how the training phase of the database is easily understood by the end user and as the staging of errors made during the game contributes to explanation. However, the ethical aspect of this research should also be considered. Users should be informed that their data (drawings, chosen words) will become part of a training database.

Exposure devices

Platform centered

Another way that brings non-expert users closer to understanding the functioning of complex algorithms that dominate specific platforms is to “put a spotlight on users’ tracking, profiling, on the data market and on the influence of algorithms” within specific social media platforms. The tracking.exposed project analyses evidence of algorithm personalisation by enabling social media users in scraping their personalised feeds to funnel perception of reliance and doubt of results of these technologies among the general public. For instance, the reproduction of genderedness in shopping spaces by Amazon has been explored by

⁶ <https://www.memo.tv/works/true-colors/>

exposing and visualising the search results of different user profiles. All the records of each profile have been collected using the Amazon Tracking Exposed extension and then displayed by merely visualising the showcase of products that Amazon’s algorithm(s) suggested to each user profile (see: <https://wiki.digitalmethods.net/Dmi/WinterSchool2020trackinggenderbiasamazon>). The use of visualisations that merely expose how and in which order different types of contents are presented to the users has also been experimented with data coming from Facebook. Indeed, the deferred staging of users’ experience is a good tool for bringing users closer to understanding the functioning of complex algorithms.

Awareness Triggers

User-centered

The mission of Algorithmic Justice League (AJL) is “to raise public awareness about the impacts of AI and galvanize researchers, policymakers, and industry practitioners to mitigate AI bias and harms”. They use storytelling techniques and video editing to make short films and documentaries that highlight the stories of people who have been impacted by harmful technology. Joy Buolamwini, the project leader, presents how facial recognition systems are not partial: they embed priorities, preferences and prejudices of those who create datasets and models. By presenting a series of use cases, she demonstrates how facial recognition can be used by the government and the private sector in a variety of ways. Moreover, in the AI, Ain’t I A Woman project, Buolamwini performs a spoken word piece that highlights the ways in which artificial intelligence can misinterpret the images of iconic black women — Oprah, Serena Williams, Michelle Obama, Sojourner Truth, Ida B. Wells, and Shirley Chisholm — using poetry as a medium.

Annotated Visual Experiments

Technology centered (multiple models, different types of data)

Finally, the Tactical Tech Group “investigates how digital technologies impact society and individual autonomy, using our findings to create practical solutions for citizens and civil society actors”. In *The Glass Room — Misinformation* edition, an itinerant exhibition welcoming various kinds of projects, the DeepFakeLab has been included. Specifically, a group of students presented the hands-on knowledge gained from their experiments in order to bring deep fakes closer to lay users who may run into this technology. By decomposing the process of creating a deep fake video and showing how different variables impact the output quality, they intended to have their audience develop an understanding of the technology’s real-world capabilities. To make the message clear, they used different types of visual languages — input/output videos, diagrams and graphs explaining technical details and an explanatory text.

Conclusion

This review has presented a selection of essential research on algorithms from a theoretical, methodological and empirical perspective. Inevitably, this account is not exhaustive; many more works could have been included in this document, having contributed in various ways to the critical study of algorithmic elaborations, algorithmic accountability and awareness. Yet, our attempt was to selectively highlight what we believe is key cross-disciplinary research on algorithms and public opinion, anchored in the perspective of digital sociology - within which this research is epistemologically located.

The study of algorithms, and particularly of recommendation algorithms, remains a lively and dynamic field of research, where a variety of scholarly areas concur to the advancement of knowledge from several standpoints. The process of re-mediation of the 'social' operated by digital technologies, of which algorithms are a key component, is still unfolding, and there is an imperative to continue to enhance accountability and transparency around algorithms - which we believe is vital for the health of public debate and society at large.

Acknowledgments

This review represents the work of all authors and their contribution at various stages, from preparation to writing. For purposes of authorship recognition, Introduction, Conclusion and section 1 are the result of collaborative work by Gandini, Leonini, Mosca and Orofino; section 2 is the result of collaborative work by Gandini, Keeling and Reviglio; the main author of section 3 is Splendore; the main authors of Chapter 4 are Gandini and Gerosa; the main author of section 5 is Gobbo.

Bibliography

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. <https://doi.org/10.1109/TIP.2003.819861>.
- Airoidi, M. (2021). *Machine Habitus: Toward a Sociology of Algorithms*. John Wiley & Sons.
- Airoidi, M., Beraldo, D. and Gandini, A. (2016) Follow the algorithm: An exploratory investigation of music on YouTube, *Poetics* 57 (2016): 1-13.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-36.
- Amoore, L. (2020). *Cloud ethics*. Durham: Duke University Press.
- Anderson, C. W. (2011). *Deliberative, Agonistic, and Algorithmic Audi-*

ences: Journalism's Vision of its Public in an Age of Audience Transparency. *International Journal of Communication*, 5(0), 19.

Andrejevic, M. (2013). Exploitation in the data mine. In Fuchs, C., Boersma, K., Albrechtslund, A., & Sandoval, M. (Eds.), *Internet and surveillance: The challenges of Web 2.0 and social media*, London: Routledge. pp. 91-108.

Ash, J., Anderson, B., Gordon, R., & Langley, P. (2018). Unit, vibration, tone: A post-phenomenological method for researching digital interfaces. *Cultural Geographies*, 25(1), 165–181.

Bainotti, L., Caliandro, A., & Gandini, A. (2020). From archive cultures to ephemeral content, and back: Studying Instagram Stories with digital methods. *New Media & Society*, DOI: 1461444820960071.

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, 6(1), 2053951718819569. <https://doi.org/10.1177/2053951718819569>

Beer, D. (2017). The social power of algorithms. *Information, Communication and Society*, 1-13.

Bennett, W. L., & Livingston, S. (2021). A brief history of the disinformation age: Information wars and the decline of institutional authority. In W.L. Bennett & S. Livingston (Eds.), *The disinformation age*, Cambridge University Press, pp. 3-40.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5), 739-768.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday*, 21(11-7).

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

Bishop, S. (2018). Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. *Convergence*, 24(1), 69-84.

Boccia Artieri, G., & Marinelli, A. (2018). Introduzione: piattaforme, algoritmi, formati. Come sta evolvendo l'informazione online. *Problemi dell'informazione*, 43(3), 349-368.

Bonini, T., & Gandini, A. (2020). The Field as a Black Box: Ethnographic Research in the Age of Platforms. *Social Media+ Society*, 6(4), 2056305120984477.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15(3), 209-227

Bozdag, E., & van den Hoven, J. (2015). Breaking the filter bubble: De-

mocracy and design. *Ethics and Information Technology*, 17(4), 249–265.

Bruns A. (2019) 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.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30–44.

Bucher, T., & Helmond, A. (2017). The affordances of social media platforms. In J. Burgess,, A.E. Marwick, T. Poell (Eds.). (2017). *The SAGE handbook of social media*, London: Sage, pp. 233-253.

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512. <https://doi.org/10.1177/2053951715622512>.

Carlson, M. (2018). Confronting measurable journalism. *Digital Journalism*, 6 (4),, 406-417.

Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, 17, 22.

Caliandro, A., Anselmi, G., & Sturiale, V. (2020). Fake news, Covid-19 e Infodemia: un esempio di ricerca sociale in real-time su Twitter. *Mediascapes journal*, (15), 174-188.

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

Caplan, R., & Boyd, D. (2016). Who controls the public sphere in an era of algorithms. *Mediation, Automation, Power*, 1-19.

Caplan, R., & Gillespie, T. (2020). Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. *Social Media + Society*, April, 1–13.

Carson, A. B. (2015). Public discourse in the age of personalization: Psychological explanations and political implications of search engine bias and the filter bubble. *Journal of Science Policy & Governance*, 7(1).

Cheney-Lippold, J. (2011). A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28(6), 164-181.

Cherubini, F., and Nielsen, R.K. (2016). *Editorial Analytics: How News Media Are Developing and Using Audience Data and Metrics*. Oxford, UK: Reuters Institute for the Study of Journalism. Available at SSRN 2739328.

- Christin, A. (2020). The ethnographer and the algorithm: Beyond the black box. *Theory and Society*, 49(5–6), 897–918. <https://doi.org/10.1007/s11186-020-09411-3>
- Christin A (2020) *Metrics at Work: Journalism and the Contested Meaning of Algorithms*. Princeton, NJ: Princeton University Press.
- Coddington, M. (2015). Clarifying journalism’s quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331-348.
- Correll, M. (2019). Ethical Dimensions of Visualization Research (pp. 1–13). Association for Computing Machinery (ACM), <https://doi.org/10.1145/3290605.3300418> (Last accessed 20 November 2021)
- Cruz, E. G., & Harindranath, R. (2020). WhatsApp as “technology of life”: Reframing research agendas. *First Monday*. <https://doi.org/10.5210/fm.v25i12.10405>
- DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753–773.
- Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism*, 3(3), 398–415. <https://doi.org/10.1080/21670811.2014.976411>
- Dick, M. (2011). Search engine optimisation in UK news production. *Journalism Practice*, 5(4), 462-477.
- Dogrueel, L., Facciorusso, D., & Stark, B. (2020). ‘I’m still the master of the machine.’ Internet users’ awareness of algorithmic decision-making and their perception of its effect on their autonomy. *Information, Communication & Society*, 0(0), 1–22.
- Doshi-Velez, F., & Perlis, R. H. (2019). Evaluating machine learning articles. *JAMA*, 322(18), 1777-1779.
- Duffy, B. E., Poell, T., & Nieborg, D. B. (2019). Platform practices in the cultural industries: Creativity, labor, and citizenship. *Social Media+ Society*, 5(4), 2056305119879672.
- Duguay, S., Burgess, J., & Suzor, N. (2020). Queer women’s experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence*, 26(2), 237–252. <https://doi.org/10.1177/1354856518781530>
- Dylko, I., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M., & Aaziz, O. (2018). Impact of customizability technology on political polarization. *Journal of Information Technology & Politics*, 15(1), 19–33.
- El-assady, M., Jentner, W., Kehlbeck, R., & Schlegel, U. (2019). Towards XAI : Structuring the Processes of Explanations Towards XAI. In: *ACM Workshop on Human-Centered Machine Learning*.

- Eriksson, M., Fleischer, R., Johansson, A., Snickars, P., & Vonderau, P. (2019). *Spotify teardown: Inside the black box of streaming music*. Cambridge, Mass: MIT Press.
- Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahali-
os, K., ... & Sandvig, C. (2015, April). "I always assumed that I wasn't
really that close to [her]" Reasoning about Invisible Algorithms in News
Feeds. In *Proceedings of the 33rd annual ACM conference on human
factors in computing systems*, pp. 153-162.
- Farkas, J., & Schou, J. (2018). Fake news as a floating signifier: Hegemo-
ny, antagonism and the politics of falsehood. *Javnost-The Public*, 25(3),
298-314.
- Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value
sensitive design and information systems. In *Early engagement and new
technologies: Opening up the laboratory* (pp. 55-95). Springer, Dor-
drecht.
- Gerbaudo, P. (2018). Social media and populism: an elective affinity?.
Media, Culture & Society, 40(5), 745-753.
- Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and
the data-intensive web. *New Media & Society*, 15(8), 1348-1365.
- Gerrard, Y. (2018). Beyond the hashtag: Circumventing content modera-
tion on social media. *New Media & Society*, 20(12), 4492-4511. [https://doi.
org/10.1177/1461444818776611](https://doi.org/10.1177/1461444818776611)
- Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Us-
ing agent-based modelling to test a meta-theoretical framework for the
emergence of filter bubbles and echo chambers. *British Journal of Social
Psychology*, 58(1), 129-149.
- Gillespie, T. (2014). The relevance of algorithms. *Media technologies:
Essays on communication, materiality, and society*, 167.
- Gillespie, T. (2017). Governance of and by platforms. In J. Burgess, A.E.
Marwick, T. Poell (Eds.). (2017). *The SAGE handbook of social media*,
London: Sage, pp. 254-278.
- Google AI. (2017). Teachable Machine. Retrieved from: [https://teach-
ablemachine.withgoogle.com/](https://teachablemachine.withgoogle.com/) (Last accessed 23 November 2021).
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content mod-
eration: Technical and political challenges in the automation of platform
governance. *Big Data & Society*, 7(1), 2053951719897945.
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H. B. (2018). Readers'
perception of computer-generated news: Credibility, expertise, and read-
ability. *Journalism*, 19(5), 595-610.
- Gran, A.-B., Booth, P., & Bucher, T. (2020). To be or not to be algorithm
aware: A question of a new digital divide? *Information, Communication*

& Society, 1–18. <https://doi.org/10.1080/1369118X.2020.1736124>

Greene, T., & Shmueli, G. (2019). How Personal is Machine Learning Personalization?. arXiv preprint arXiv:1912.07938.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7(8), 993-1012.

Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021). Examining the consumption of radical content on YouTube. *Proceedings of the National Academy of Sciences*, 118(32).

Howard, P. N., & Kollanyi, B. (2016). Bots, #strongerin, and #brexit: Computational propaganda during the uk-eu referendum. Available at SSRN-2798311.

Howard, P. N., Bolsover, G., Kollanyi, B., Bradshaw, S., & Neudert, L. M. (2017). Junk news and bots during the US election: What were Michigan voters sharing over Twitter. *CompProp, OII, Data Memo*, 1.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146.

Jasanoff, S., & Kim, S.-H. (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press.

Jentner, W., Sevastjanova, R., Stoffel, F., Keim, D., Urgen Bernard, J., & El-Assady, M. (2018). Minions, Sheep, and Fruits: Metaphorical Narratives to Explain Artificial Intelligence and Build Trust, retrieved from: <https://bib.dbvis.de/uploadedFiles/minionssheepfruits.pdf>.

Just, N., & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, 39(2), 238–258. <https://doi.org/10.1177/0163443716643157>

Kidron, B., Evans, A., Afia, J., Adler, J. R., Bowden-Jones, H., Hackett, L., & Scot, Y. (2018). Disrupted childhood: The cost of persuasive design. *5Rights*.

Kitchin, R. & Lauriault, T. (2014) *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work*. In J. Eckert, A. Shears, J. Thatcher (Eds.), *Geoweb and Big Data*. University of Nebraska Press. Available at SSRN: <https://ssrn.com/abstract=2474112> (last accessed 20 November 2021).

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, communication & society*, 20(1), 14-29.

Kovic, M., Rauchfleisch, A., Sele, M., & Caspar, C. (2018). Digital astroturfing in politics: Definition, typology, and countermeasures. *Studies in Communication Sciences*, 18(1), 69-85.

- Kreiss, D. (2021). "Social Media and Democracy: The State of the Field, Prospects for Reform," edited by Nathaniel Persily and Joshua A. Tucker.
- Lange, A.-C., Lenglet, M., & Seyfert, R. (2019). On studying algorithms ethnographically: Making sense of objects of ignorance. *Organization*, 26(4), 598–617. <https://doi.org/10.1177/1350508418808230>
- Lasorsa, D. L., Lewis, S. C., & Holton, A. E. (2012). Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism Studies*, 13(1), 19-36.
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. OUP Oxford.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. <https://doi.org/10.1177/2053951718756684>
- Lee, A. M., Lewis, S. C., & Powers, M. (2014). Audience clicks and news placement: A study of time-lagged influence in online journalism. *Communication Research*, 41(4), 505-530.
- Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881–900. <https://doi.org/10.1177/1461444816675438>
- Lomborg, S., & Kapsch, P. H. (2020). Decoding algorithms. *Media, Culture & Society*, 42(5), 745–761.
- Lupton, D. (2014). *Digital sociology*. London: Routledge.
- MacGregor, P. (2007). Tracking the online audience: Metric data start a subtle revolution. *Journalism Studies*, 8(2), 280-298.
- Markmann, S., & Grimme, C. (2021). Is YouTube Still a Radicalizer? An Exploratory Study on Autoplay and Recommendation. In *Multidisciplinary International Symposium on Disinformation in Open Online Media* (pp. 50-65). Springer, Cham.
- Marres, N. (2012). The redistribution of methods: on intervention in digital social research, broadly conceived. *The Sociological Review*, 60, 139-165.
- Marres N. (2017), *Digital sociology: The reinvention of social research*, John Wiley & Sons.
- Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. New York: Data & Society Research Institute.
- Mauri M., Cipresso P., Balgera A., Villamira M., Riva G. (2011), Why is Facebook so successful? Psychophysiological measures describe a core flow state while using Facebook, *Cyberpsychology, Behavior, and Social*

Networking, 14(12), pp.723-731.

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042-1063.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & SOCIETY*, 35(4), 957-967.

Moeller, J., & Helberger, N. (2018). Beyond the filter bubble: Concepts, myths, evidence and issues for future debates.

Moen, G. M., Ravna, A. K., & Myrstad, F. (2018). Deceived by design. *Forbrukerrådet*.

Møller Hartley, J., Bengtsson, M., Schjøtt Hansen, A., & Sivertsen, M. F. (2021). Researching publics in datafied societies: Insights from four approaches to the concept of 'publics' and a (hybrid) research agenda. *New Media & Society*, DOI: 14614448211021045.

Musiani, F., Cogburn, D. L., DeNardis, L., & Levinson, N. S. (Eds.). (2016). *The turn to infrastructure in Internet governance*. London: Springer.

Noble S.U. (2018). *Algorithms of oppression*. New York: New York University Press.

O'neil C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Books.

Offert F., (2017). I know it when I see it. Visualization and intuitive interpretability. in *NIPS Symposium on Interpretable ML*, 2017.

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The Building Blocks of Interpretability. *Distill*, 3(3), <https://doi.org/10.23915/distill.00010>.

Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford: Oxford University Press.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. London: Penguin.

Pasquale, F. (2015). *The black box society*. Cambridge, Mass: Harvard University Press.

Poell, T., Nieborg, D. B., & Duffy, B. E. (2021). *Platforms and cultural production*. New York: Polity.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford: Oxford University Press.

Postigo, H. (2016). The socio-technical architecture of digital labor: Converting play into YouTube money. *New Media & Society*, 18(2), 332-349.

Pybus, J. (2019). Trump, the first Facebook president: Why politicians need our data too. In C. Happer, A. Hoskins, W. Merrin (Eds.), *Trump's media war*, Palgrave Macmillan, pp. 227-240. s

Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 173-182).

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020, January). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 131-141.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Retrieved from: <http://arxiv.org/abs/1602.04938> (Last accessed 20 November 2021)

Rieder, B. (2020). *Engines of Order*. Amsterdam University Press. <https://doi.org/10.5117/9789462986190>

Roberge, J., & Castelle, M. (2021). *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Palgrave Macmillan, DOI: <https://doi.org/10.1007/978-3-030-56286-1>.

Roberts, S. T. (2019). *Behind the screen. Content Moderation in the Shadows of Social Media*. Yale: Yale University Press.

Robinson, S., Deshano, C., & Kim, N. (2009). Madison commons 2.0: A platform for tomorrow's civic and citizen journalism. In J. Rosenberry, B. St John (Eds.), *Public Journalism 2.0*, New York: Routledge, pp. 174-187.

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, 35(3), 213-229. <https://doi.org/10.1177/0267323120922066>

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.

Scholz, T. (Eds.). (2012). *Digital labor: The internet as playground and factory*. London: Routledge.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 2053951717738104.

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98,

Siles, I., Segura-Castillo, A., Solís, R., & Sancho, M. (2020). Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data & Society*, 7(1), 205395172092337. <https://doi.org/10.1177/2053951720923377>

Silva, T., Mintz, A., Omena, J. J., Gobbo, B., Oliveira, T., Takamitsu, H. T., ... Azhar, H. (2020). APIs de Visão Computacional: investigando mediações algorítmicas a partir de estudo de bancos de imagens. *Logos*, 27(1), 25–54.

Smith, A. (2018). Many Facebook users don't understand how the site's news feed works. Pew Research Center, Journalism and Media.

Splendore, S. (2016). Quantitatively oriented forms of journalism and their epistemology. *Sociology Compass*, 10(5), 343-352.

Striphas, T. (2015). Algorithmic culture. *European journal of cultural studies*, 18(4-5), 395-412.

Strobelt, H., Gehrmann, S., Pfister, H., & Rush, A. M. (2018). LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 667–676.

Sunstein, C.R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, Princeton.

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social media+ society*, 7(2), 20563051211008828.

Tandoc Jr, E. C. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, 16(4), 559-575.

Tucker, J.A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B., 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature* (March 19, 2018).

Tufekci, Z. (2018). YouTube, the great radicalizer. *The New York Times*, 10, 2018.

Wang, S. (2020). Calculating dating goals: Data gaming and algorithmic sociality on Blued, a Chinese gay dating app. *Information, Communication & Society*, 23(2), 181–197. <https://doi.org/10.1080/1369118X.2018.1490796>

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report, 27.