

Ghosts of white methods? The challenges of Big Data research in exploring racism in digital context

Big Data & Society
 July-December: 1–12
 © The Author(s) 2021
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/20539517211048964
journals.sagepub.com/home/bds



Kaarina Nikunen 

Abstract

The paper explores the potential and limitations of big data for researching racism on social media. Informed by critical data studies and critical race studies, the paper discusses challenges of doing big data research and the problems of the so called ‘white method’. The paper introduces the following three types of approach, each with a different epistemological basis for researching racism in digital context: 1) using big data analytics to point out the dominant power relations and the dynamics of racist discourse, 2) complementing big data with qualitative research and 3) revealing new logics of racism in datafied context. The paper contributes to critical data and critical race studies by enhancing the understanding of the possibilities and limitations of big data research. This study also highlights the importance of contextualisation and mixed methods for achieving a more nuanced comprehension of racism and discrimination on social media and in large datasets.

Keywords

Big data, racism, social media, topic modelling, critical race studies, white methods

Research based on large datasets extracted from different social media platforms, such as Twitter and Facebook, has provided new studies on migration, the refugee crisis and the circulation of racism. Using big data analytics, researchers have been able to identify prominent themes, topics and agents in the public debate connected to migration and the intertwined discourses of migration and racism (Poole, 2019; Siaperä et al., 2018). Several studies based on large datasets also explored the rise of digital racism (Siaperä, 2019; Awan, 2016) and the role of digital technology in enhancing the spread of racial discourse (Daniels, 2018; Farkas et al., 2018; Matamoros-Fernandez, 2017; Noble, 2018). Conducted in the context of Finland, our study explored the dynamics between the mainstream news media and the social media and the ways in which these dynamics shape and strategically amplify, for example, the different racialised understandings of migration and refugee issues in society (Laaksonen et al., 2020; Pantti et al., 2019; Pöyhkäri et al., 2019).

Since big data provides powerful new measures of social life, it is vital to critically examine the potential and the limitations of big data approaches, particularly from the perspective of marginalised groups. Big data can show dominant trends and vocabularies connected with migration, such as framing refugees as a threat and the increase

of racist expressions and hate speech in connection to migration (Ferra and Ngyen, 2017; Siaperä et al., 2018); however, critical race studies and feminist research have identified various limitations to what can be achieved with big data. This paper interrogates the potential and the limitations of big data research by introducing three methodological approaches from our own big data research. The paper contributes to critical data and critical race studies by enhancing the understanding of the possibilities and limitations of big data research (Cooky et al., 2018; Gillborn et al., 2018; Iliadis and Russo, 2016; boyd and Crawford, 2012).

Before discussing the case itself, I address the question of situated knowledge and research positionalities, which are rarely taken up in data studies (Matamoros-Fernandez and Farkas 2021); however, critical data and critical race studies find these aspects to be essential. I write this in my role as the head of a multidisciplinary research

Faculty of Information Technology and Communication, Tampere University, Finland

Corresponding author:

Kaarina Nikunen, Faculty of Information Technology and Communication, Tampere University, Finland.
 Email: Kaarina.nikunen@tuni.fi



project, as an afterthought and reflection of the research process, and as a scholar who has worked several years in the field of migration research and feminist research. As a white woman, who works and lives in a predominantly white Nordic society, I am shaped by my privilege that reflects and limits my research in certain ways. These reflections come out of the attempt to take situated knowledge seriously as part of big data research and to interrogate the attempts to expand the knowledge interest in big data research. In what follows, I discuss the challenges posed by critical data studies and critical race studies to big data methods and then I introduce the case study. The paper discusses these issues in the context of social media studies.

Critical data studies meets critical race studies

Critical data studies have extensively analysed the various limitations of big data methodologies (Cooky et al., 2018; Crawford et al., 2014; Hargittai, 2015; Trottier, 2014). Datasets are always historical, constructed and laden with political and ethical values (Metcalf and Crawford, 2016). Metcalf and Crawford call for more nuanced and ethical research processes and for a better understanding of the multiple ways in which big data processes can cause not only individual but networked harm (Metcalf and Crawford, 2016, 2–3).

Topics that concern marginalised groups or experiences of discrimination and racism pose particular challenges for big data research. As many scholars have pointed out, although large social-media datasets are often considered to represent the general public, in fact social-media data emphasise the contributions of those actors who are active and in powerful positions on social-media networks (Cooky et al., 2018; Kennedy and Moss, 2015). Also, different social-media platforms have different users. For example, in various national contexts, Twitter has been the platform favoured by influencers, politicians, journalists and activists (Wojcik and Hughes, 2019; Larsson and Moe 2011; however see Brock, 2012; Sharma, 2013; Clark 2020), whereas some populations are underrepresented in most social media. In other words, the demographic and socioeconomic factors matter when choosing social media sites for analysis (Hargittai, 2015). This is not a particular problem if these limitations of and biases in selection are clearly acknowledged. Illustrative of this point is the Great British Class Survey conducted by Burrows and Savage (2014) through a BBC-hosted online platform. The authors reflect on how big data gathered via social media provided self-referential performativity indicators to educated upper and middle classes by functioning as a new mirror in which to look at themselves. Other, more marginal populations appeared as ‘ghostly’ or invisible figures in this research (Burrows and Savage, 2014, 4). Critical race scholars argue that quantitative analyses,

such as big data analysis, lack a critical race-conscious perspective and tend to repeat existing racial inequalities and to disguise or even normalise racial inequity (Gillborn et al., 2018). Minorities get easily lost in big data analytics, and rather than ignoring minorities in big data, researchers could focus on ‘the statistical outliers’ or complement big data with qualitative approaches to make up for the biases (Crawford, 2013; Welles, 2014).

The ‘loss of meaning’ is another recurring problem in big data analysis on social media. Big data is extracted from its context, and, therefore, the contexts of the meaning, understanding and practice cannot be read from the data (Hand, 2014, 15; Baym, 2013; Andrejevic 2020). In other words, what is left out in big data analysis is crucial for understanding the nature of the data and the contexts in which the data were produced. From the perspective of critical race studies, pointing out how the meanings of race and racism are produced in particular discursive fields is critical for identifying the manifestations of racial closure, a process that has important implications for interpreting data (Bulmer and Solomos, 2004, 9). For example, Gillborn et al. (2018, 173) showed that when race was taken as a prior category, racist patterns of inequality (failures in education, criminal prosecution, etc.) were often interpreted as being caused by ethnic origin, whereas critical race scholars explored the ways in which these patterns and outcomes were connected to structural racism. A lack of such contextual understanding of data or sensitivity to complex and multifaceted experiences can result in unproblematic collation of racialised subjects with socially problematic behaviour (Gillborn et al., 2018). This tendency is evident in much of the research on race in the social sciences that focuses on social problems and crime. Large datasets can further stigmatise groups by means of categorisations and profiling connected to credit-worthiness or criminal justice (Richardson et al., 2019; Zook et al., 2017). Gangadharan (2012) pointed out that big data practices and algorithmic profiling impact the political freedom (racial profiling), economic well-being (redlining) and health (medical profiling) of people from different social classes and circumstances.

Bonilla-Silva and Zuberi (2008) use the terms *white logic* and *white methods* to describe the tendency to fortify racial inequalities via research design. These concepts refer to the dominance of whiteness in the social sciences, as part of the social structures that reproduce white privilege (Frankenberg, 1993). Bonilla-Silva and Zuberi argue that for a long time, the social sciences have been numerically, logically and methodologically led by white male scholars. White logic refers to ‘a context where white supremacy has defined the techniques and processes of reasoning about social facts’ (2008, 17). It refers to the basis of the techniques used to analyse empirical reality and the basis of the reasoning used to understand society. They argue that white logic assumes objectivity and

is the anchor of Western imagination, granting centrality to the knowledge, history, science and culture of the predominantly white world and experience. Bonilla-Silva and Zuberi describe white methods as the practical tools that are used in empirical research that reproduces racial inequality and stratification. The argument by Bonilla-Silva and Zuberi points out whiteness as a norm that dominates the contexts where research is done.

Whiteness can be approached as a structural position rooted in history of colonialism, imperialism and capitalism (Du Bois, 2003; Garner, 2007; Lipsitz, 2006). Central characteristics of whiteness is the way its' discriminatory power is not recognized by white people or the advantages gained from being white are not identified (Dyer, 1997). These advantages are maintained through norms, traditions, and institutions. Daniels (2015) argues that such normative whiteness is an essential part of technology industry with white leadership, racial inequalities and ideology of color-blindness. Consequences of such normative whiteness become illustrated in examples of technology design where white face and skin are set as default (Benjamin, 2019; Buolamwini and Gebru, 2018; Constanza-Chock, 2020).

Similarly, Chakravarty et al. (2018) and Ng et al. (2020) demonstrate that the field of communication studies is dominated by the ideology of white supremacy: non-white scholars continue to be under-represented in communication studies. These inequalities are not only connected to methods but are re-enacted in citations, editorial positions, publications and conferences, reflecting the histories of colonial power, especially among Western academia, and the dominance of the English language (Ng et al., 2020; see also Hedge and Shome 2002). These imbalances in research are furthered by the data divide: technology companies have become increasingly data rich, whereas activists, associations and researchers in public institutions have few possibilities to work with similar datasets (boyd and Crawford, 2012). Furthermore, data require resources and the ability to unravel black boxes and designs to create new approaches. This is often possible only for scholars doing research in wealthy Western universities or institutions (boyd and Crawford, 2012; Cooky et al., 2018, 9; Crawford et al., 2014).

Whiteness, as constructed and multifaceted positionality (Garner, 2007) operates in different ways in a range of national, social and racial contexts. In the Nordic countries whiteness appears as pervasive social reality that has been historically strengthened with race science (Andreassen, 2014). While in contemporary multicultural societies Nordic whiteness is increasingly contested, it is an integral part of the societies and academic life taking shape in colour blindness (Ahmad, 2020; Hübinette and Tigerwall, 2009; Keskinen, 2018; Lundström and Teitelbaum, 2017; Rastas, 2007). As argued by Rastas and Seye (2018, 4) 'in predominantly white social spaces, racism and anti-

racism are often still understood as issues that concern only non-white people and other racialized minorities'.

When discussing the shortcomings of traditional methods for understanding racial experiences and inequalities, Bonilla-Silva and Zuberi (2008) state that 'if we begin with a racially biased view of the world, we will end with a racially biased view of what data have to say'. The knowledge that is based on the behaviour of dominant groups tends to lead to knowledge that reproduces that order. Mohanty (2003) echo this notion by arguing that, if we form our understanding of a just society on the basis of the experience of the privileged, we are not likely to recognise the structures that further inequalities, whereas the experience of the marginalised 'allows for a more concrete and expansive vision of universal justice' (Mohanty, 2003, 510).

The understanding of big data as an automated process that supposedly produces 'neutral' data for public institutions, policy makers or markets, without questioning the composition of the data and its contexts, invisibilities and origins, indicates the practices of white methods. This does not mean that big data is automatically 'white method' but adheres to contexts (technology studies and industry) and practices (de-contextualization, correlation) that can consolidate existing ideology of colour-blindness and be incapable of addressing issues of race and racial inequality (Daniels, 2015). In other words, researchers should carefully consider the ways in which big data research may contribute to racial inequalities and discrimination and explore the means to challenge this. This also means exploration of construction of whiteness in research contexts (Du Bois, 2003).

What are the ways to challenge white logic and white methods? Critical race studies and feminist theory grant epistemic power to experience as a source of knowledge in order to produce transformative knowledge in the interest of social justice (Cooky et al., 2018). Particularly for marginal groups, sexual minorities, and ethnic and racial minorities, the ability to bear witness, to provide evidence from experience, has been an important way of shedding light on lived inequalities shaped by structures of discrimination. Essed (2004) argue that the people who experience racism can produce knowledgeable insight into how racism manifests itself in everyday life. Therefore, it would 'make sense to listen to what ethnic minorities have to say, to explore through probing and questioning what life felt like in white dominant society, to see dominant society through the eyes of those who were considered not to belong, not to be part of the norm' (Essed, 2004, 125). This is a way to 'upset' something that has been taken for granted (Scott, 1991). At the same time feminist theory has problematized the essentializing force of experience and the way it can operate as a category. Scott argues that while experience can make individuals as the starting point of knowledge, it can also naturalize their categories as black,

white, heterosexual, homosexual etc. (Scott, 1991, 782). To avoid this Scott proposes that experience should not *explain* but, rather, create space for *reflexive interrogation* (2001). Scott's work is critical of homogenizing identity categories (discussed as fantasy scenarios) and argues the need for exploring their diffractions, variations and differences. In other words experience of inequality is different within and across different contexts and groups.

To bring together the ideas by Bonilla-Silva & Zuberi and Scott, the methodological challenge is therefore multifaceted: to be able to uncover the contexts of dominant whiteness in big data research design as well as ability to address the complexity of inequalities connected with multiplicity of identities and experience. As pointed out by McCall (2005) intersectional complexity is difficult to capture through traditional social scientific methods and requires combination of approaches.

Such space can emerge in contextualized research that identifies the complexities of knowledge production and experience, focusing on the processes where categories and inequalities connected to them, are being produced, experienced and resisted (McCall, 2005): 'Different contexts reveal different configurations of inequality' (McCall, 2005, 22).

To address these challenges, I introduce our research project Racisms and Public Communication in the Hybrid Media Environment (Hybra) Based on the experiences of our project, I identified the following three approaches to explore racism in the context of social media with big data: 1) using big data analytics to identify the dynamics of how racist discourse is produced online 2) complementing big data and going beyond big data via qualitative approaches and 3) using big data to question or reform the infrastructures that foster hate and racism – in other words, questioning the data-based logics of racism and discrimination. At the end of the paper, I discuss lessons learned from these different approaches and ways to enhance big data research with insights from critical race studies and feminist scholarship.

Big data analytics of racist discourse

Hybra research project focused on racist discourse and the dynamics of the refugee debate in Finland. Hybra was funded by the Academy of Finland as a consortium of three universities, with researchers from media studies with expertise in migration and critical race studies, social sciences and computational science. The multidisciplinary nature of the project was both its strength and the main challenge. The scholars with significant experience in racial and migration studies had very little expertise in big data research, whereas the computational scientists were not familiar with critical race studies or migration research, although they had expertise in social sciences.

The research project was born in a politically tense situation in Finland with growing public expressions of racism and rise of the anti-immigrant movement. While political groups with openly anti-immigrant agenda had been organizing from early 2000s on, the so-called European refugee crisis¹ in 2015 intensified public debates and increased racism towards migrants particularly from the Middle East and the African continent. The anti-immigrant movement targeted strongly all asylum seekers, however anti-Black racism is unquestionably also part of the society². According to European survey on the experience of racism among African immigrants and descendants of African immigrants in 12 European countries, prevalence of racist harassment was the highest in Finland (EU-Midis, 2018).

The researchers in the group shared concern over the increased racism in Finland and saw this project as a way to explore its dynamics and influence on society. In this sense the researchers were in accordance, however some were more actively engaged in anti-racist activism than others. Building a common understanding of the key concepts, understanding the definitions of race and racism and figuring out how to use big data to meaningfully explore these questions required time and many conversations.

Our aim was to explore the emergence of racist discourse in public debates during the refugee crisis. To achieve this, we wanted to look at the public debates on migration and refugee issues by using large datasets to get a bird's-eye view of the phenomenon. This seemed like a great opportunity to disclose the dynamics of spread of racist discourse that we had encountered in our previous work in smaller qualitative settings. The focus was on spread of racist discourse, rather than on the experience of it, which seemed logical and timely. To be honest, we were all excited about the opportunity to conduct this research that was first of its kind in Finland. However, looking back, foregrounding the spread of racism over experience of it can also be derived from our own positionalities as scholars who had no first hand experience of racism.

The research team was all white, which is not exceptional in Finland, actually, it is more of a rule in academic projects. This reflects whiteness as a structural discriminatory power in the Finnish academia and illustrates concerns raised by Bonilla-Silva and Zuberi in terms of white logic. The small number of non-white academics cannot be explained by demographic factors alone, not even in predominantly white societies, such as Finland. Structural racism inevitably affects study choices, career paths and researchers' work opportunities (Rastas and Nikunen, 2019). In one of our early discussions, one of the researchers addressed this problem of not having any non-white scholars in the team. While this was acknowledged as a limitation, it was not really addressed until later in the process. I return to this question after introducing the case study.

At the time, topic modelling was gaining popularity as an accessible method to be applied to large datasets. Despite coming from different research traditions, we found topic modelling to be a method we all understood. Therefore, we conducted two case studies using latent Dirichlet allocation (LDA) (Blei, 2012), an increasingly popular method in the humanities and the social sciences for studying textual data (e.g. Jacobi et al., 2016). Both case studies had their strengths and specific problems.

The first study investigated the refugee debate in Finland by identifying the patterns, dominant agents and dynamics between the news media and the social media during the so-called refugee crisis. As we were looking for ways to explore the dynamics of the refugee debate in Finland, we combined different approaches with topic modelling, such as hyperlink analysis and network analysis, to see what big data might show us. These decisions reflect both the uncertainty and expectations that are, perhaps, typical of many big data projects, in which researchers rely on technology-based empiricism (Mazzocchi, 2015), collect the data and hope to find something (Nelimarkka, 2021; Laaksonen et al., 2020).

The data were collected from Finnish news media, social media and online discussion forums consisting of 27,504 online news articles and 1,082,815 unique messages from a variety of social-media platforms, including Twitter, Facebook, Instagram, Google Plus and YouTube, and hundreds of Finnish online discussion forums and thousands of blogs provided by Futusome (Pöyhtäri et al., 2019).

For the social-media data, the final analysis was based on 58 topics that seemed to have the highest relevance to the refugee debate or the broader discussion on immigration. For the news-media data, we identified 38 topics relating to the refugee and migration debates. Previous experience in migration research provided valuable insight when interpreting the topics and grouping them thematically (Isoaho et al., 2019).

We found that both in the news media and on the social media, refugee issues were connected with negative connotations; however, the social-media debates had overtly negative, anti-immigrant framings that were characterised by racist discourse, hostile expressions and negative stereotypes. Crime (allegedly committed by asylum seekers and other immigrants) was a prevalent theme both in news media and on social media, but the social-media debates were characterised by an anti-immigration stance, with a focus on the increasing number of sexual assaults and other crimes allegedly committed by asylum seekers (Pöyhtäri et al., 2019).

To further understand the relationship between the refugee discussion and the wider media ecology, we analysed hyperlink sharing on different platforms. On Facebook and Twitter, news was used to provide contextual background or positive views on the refugee crisis, whereas on discussion forums (particularly Suomi24 and

Hommaforum), the news and news-like content was mostly circulated to spread negative views motivated by anti-immigration ideology. The tendency to share content from alternative media with an anti-immigrant stance was evident (Pöyhtäri et al., 2019).

This case study was not ideal because we did not develop the conceptual approach to the research question but, instead, tried different approaches to data. While topic modelling provided convincing results, such results were not necessarily novel in migration research. However, the hyperlink analysis opened up an interesting view regarding traces of manipulation and purposefully circulated content on social media, but due to time and resource constraints, this approach could not be explored further.

Overall, the study confirmed discoveries from previous research (dominant frames of threat, crime and management) and what had been observed in various qualitative studies (Chouliaraki and Stolic, 2016; Nikunen, 2019a; Van Dijk, 1991; van Leeuwen and Wodak, 1999; Horsti, 2008). The big data approach enabled us to ‘verify’ how crime-related issues are purposefully connected to refugees and how the hybrid media environment is used by anti-immigrant groups. This finding reflected the contemporary socio-historical climate, characterised by the rise of populism and the intensifying debate over immigration with racist undertones in the digital-media context. In the end, we saw the racist undertones and the networks emerging in our data, but to understand the dynamics of racism, we should have taken more time to connect the main concepts and theories of digital racism with the code and its work. Developing a more integrated combination of computational and qualitative approaches would have required much more time from the whole research team. Several issues prolonged the process and kept our attention in technical matters. For example, gathering data in machine-readable form required several negotiations with the leading national newspaper. Concerns over the access of data and challenges to understand it, felt urgent and took time from developing more integrated approaches.

The second case study evolved more gradually. We had several discussions on the object of research, reaching the understanding that racism cannot be approached as a fixed thing that we can capture using big data; instead, we decided to turn the research question around and investigate how racism was understood and discussed in public. The idea was to ‘map macro-level discursive contexts in which racism is discussed in Finnish news media and online discussion forums by using topic modelling’ (Pantti et al., 2019). This time, the social-media data was well known existing academic database, also used in previous research. The social media data comprised the Suomi24 discussion forum (Lagus et al., 2016). Due to its size, the forum can be considered a significant platform in the Finnish mediascape even though it has lost popularity

in recent years. The data consisted of 5262 news items and 113,410 social-media posts. Again, we conducted topic modelling with 86 initial topics. After excluding incoherent and nonmeaningful topics, we ended up with 51 topics. We then organised the remaining manually validated and labelled topics into higher-order categories to show the overall discursive contexts in which the term ‘racism’ was used in the Finnish media. Our analysis yielded 13 different categories.

The study showed that the discussion of racism on the online forum was predominantly employed by the supporters of the Finns Party and the anti-immigrant movement. For example, the strategy of reversing the victim–oppressor positions, typical of populist right-wing rhetoric, was highly prominent. The Finnish-speaking Finns were positioned as the racialised others in opposition to the Swedish-speaking minority population. Individual topics represented different perspectives on this victimised collective identity, such as ‘language racism’ referring to Swedish as a mandatory and official language in Finland. The dilution of the term ‘racism’ in the Finnish public discourse was evident, with terms such as ‘age racism’, ‘obesity racism’, ‘ugliness racism’ and ‘health racism’ (Pantti et al., 2019). Overall, the study demonstrated that different topics were framed differently in the online forum and the news media. Individual stories about the experiences of racism by people of colour in Finland were mostly addressed in the context of news on anti-racist campaigns. By contrast, on the discussion forum, the notion of racism was both de-historicised and re-historicised by expanding its meaning to include reverse racism and producing new expressions. This finding echoed previous studies that discussed how racism takes on a variety of forms, finding new vocabularies and appearing to be debatable (Bulmer and Solomos, 2008; Song, 2014; Titley, 2019). It also illustrated the mobility of racism that cannot be captured and fixed in a simple way.

Both studies were able to demonstrate how right-wing populist ideology and actors are using the social-media environment to circulate racist discourse. This is done by connecting refugees and asylum seekers with crime and threat, often using unconfirmed sources and denying accusations of racism while appropriating reverse-racism discourse to accuse the elite and the media of discriminating white Finns. In both cases, there were biases in the data that limited the research and caused epistemological problems that critical race studies have pointed out and that need to be acknowledged.

The first approach: addressing the bias

In the first case, the social-media data were provided by a data-analytics firm. As the data were collected by an external company, we were faced with the classical black-box problem: we could not be entirely certain of how the data

were collected and what they consisted of. In the second case, the social-media data, though vast and collected by our team, was hardly representative of the general population or even the social-media discussions in general. Suomi24 used to be widely popular among the Finnish population but has become known for uncivil discussions around migrant issues. According to the latest user survey, middle-aged males make up the largest user group on Suomi24 and this population group is also active in the anti-immigrant movement (Harju, 2018). These limitations were acknowledged but not really discussed. How does this bias shape our understanding of the discussions on racism or refugee and migration issues? What does it mean that the agents talking through our data are telling us a story that emphasises crime and threat in the context of refugees and migrants and speaks about reverse racism and discrimination of white Finns?

The landscape painted by this research approach raised various ethical and epistemological issues. Ethically, it is crucial that while trying to explore digital racism, we should also try to avoid re-circulating racist stereotypes and perceptions through our research. Focusing on the racist discourses produced by hateful groups may inadvertently give visibility and power to these discourses, and while this is often unavoidable, it should be minimised. Furthermore, epistemologically, we should try to unpack the structures of racial inequalities rather than embrace them. Focusing on the mainstream media and the platforms with a strong presence of anti-immigrant and right-wing populist groups does not exactly challenge the structures that feed and host racial inequalities and racist discourses.

In terms of understanding the experience of racism in the context of migration, large datasets from social media are not necessarily helpful as minority groups tend to be invisible in big data (Welles, 2014). Even if research can demonstrate the ways in which anti-immigrant ideologies spread in the public debate, the crucial question to ask is this: are these the actors that we want to focus on, when we are exploring a phenomenon such as racism? Leaning on these datasets, we inadvertently emphasised the connections between racialised subjects and crime, and focused on views of anti-immigrant groups. Even though this was done with a highly critical approach, it seemed evident that we needed to expand our exploration to include experience of and response to racism among and with racialized minorities.

Clearly, the research design would have benefitted from a team accompanied by scholars of colour, right in the beginning. This could have impacted the approaches and choices made, for example what questions to ask, which platforms to explore and how. It is likely that scholars of colour would have had better capacity to identify and challenge approaches that operate on white context. This is not only about recognizing limitations (that can often be expressed in tokenist way) but a deeper question of

epistemological value of multiple perspectives in research and central in countering white logic, as argued by Bonilla-Silva and Zuberi (2008). Actually our research setting in many ways illustrates how white logic works in academic institutional setting. As often is the case in externally funded research project, the pressure to produce results and publications can override the focus on developing and re-thinking research approach. Most members of the team were early career scholars in precarious, temporary positions who had additional pressure to to publish and meet the criteria of possible university posts. These structural aspects that favour speedy publishing and straight forward research settings, influence the research process and knowledge production profoundly. Working in an academic environment that is predominantly white and operate as colour-blind, limits capacity to identify how whiteness works. The choices made in the research project, were shaped by all of this: the combination of researchers' positionalities as white scholars in predominantly white universities, the challenge and excitement of big data methods, the sense of urgency to produce results and publish, as well as the difficulty to create time for re-thinking the research setting.

Informed by feminist and critical race research, we understood that in the next phase of the research we should include more proactive approaches in finding the spaces of experience and agency of racialised subjects. This insight didn't just land on us suddenly. Rather, it was an underlying sense of something that was missing in our approach. This sense was perhaps felt most strongly by those researchers who had prior experience in qualitative and participatory research, however, also the computational scientists of the team shared and supported the need of multiple methods. During this time new collaborative research projects on racism and anti-racist activism were established in Finland. Collaborations and exchanges with scholars of colour such as Leonardo Custódio, highlighted the relevance of experience and agency in researching racism. Custódio was invited to introduce their anti-racist work in ARMA alliance, that has developed participatory research practices and contributed to the anti-racist struggles with workshops, film and book productions (Custódio and Gathuo, 2020). In addition, British scholar Francesca Sobande was invited to introduce her work on experiences of Black people on digital racism to a public seminar and she contributed to a special issue on digital racism (Sobande, 2021). This work put our own approaches in perspective and showed how knowledge from experience may open up new insight to big data research.³ Strong focus on knowledge from experience in understanding racism was also present in art exhibition 'African presence in Finland' as well as in AfroEuropean conference organized in Tampere that pointed to the same direction in research (Rastas, 2020).

However, combining approaches from such qualitative and ethnographic research with big data studies was far from simple since they seemed to operate on different epistemological grounds. Focus on experience required stepping away from big data analysis, expanding the investigation as suggested by Crawford (2013) and combining big data with ethnography or interviews – that is, not just confirming what we see in the data but complementing data sources with qualitative research. This is what I identify as the second approach: complementing the data with interviews and case studies to account for a different angle regarding the circulation of racist discourse than what came out of big data research.

The second approach: complementing data with experience

To expand the big data approach regarding public debate, we focused on cases where migrants and refugees actively challenged the anti-immigrant discourse circulating on social media and mainstream media. The first case study focused on the Once I Was a Refugee social-media campaign on Facebook and Twitter (#ennenolinpakolainen). The study included an analysis of the campaign and interviews with the campaign's founder and participants.⁴ The study adopted a multi-sited approach, using interviews and digital ethnography to investigate both the patterns and the meanings of the campaign, with 175 postings by mostly people of colour who arrived in Finland as refugees. The campaign provided space for counter-narratives that questioned the stereotypical representations of refugees and the growing racist discourse in the public. The study revealed frustration regarding the stereotypical and racist discourse that people who are non-white experience on a daily basis in Finland (Nikunen, 2019b). Therefore, the study focused on the everyday struggles with racism and the affective circulation of hate and fear through the media and social media debates. By means of the campaign, individuals could counter the claims that refugees are useless and threatening non-humans – claims that were circulating in the public domain. The participants also experienced vulnerabilities caused by digital visibility. Many of them contemplated how their participation would put their family members at risk of hostile attacks or expose information that could be harmful to them. This reflected the complexities that come with the social-media context, algorithmic power and oppositional audiences. For people in vulnerable positions, the way social media platforms collapse multiple contexts and complicate profile management or restrict audiences can be particularly risky (Marwick and boyd, 2010).

The case concerned an emerging network of anti-racist activism in Finland that was forming outside existing

political groups and organisations. However, this emerging movement was completely invisible in our big data analysis, and we would not have encountered it without research expertise in migration. This is an important way to *curate* and complement big data with scholars who have particular expertise and experience in the field.

Another example is making big data small (Welles, 2014) by concentrating on a particular group of ‘outliers’ in big data to explore the group in more detail. This approach brings visibility to groups that would otherwise be ignored and enables a deeper understanding of experiences of racist encounters and discrimination in digital space. The case study was based on the data collected from public Facebook groups by means of a custom-built tool as part of the Hybra project <https://github.com/HIIT/hybra-someloder>. The data consisted of 3643 Facebook posts in four anti-racist and four overtly racist groups. Data was fully anonymized and used to explore the different dynamics between different group discussions. No direct quotes were used from the data or references to any individuals. This approach was complemented with interviews and focused on one particular public campaign, a Right to Live. This group, formed by migrant protesters from Iraq and Afghanistan and their Finnish allies, gained public attention in Finland through their on-going demonstration against deportations (Haavisto, 2020). The interview analysis⁵ explored the different ways in which the migrant protesters and asylum seekers took part in public debates as multi-lingual ‘fact providers’ and ‘experience experts’ in the difficult and uncertain situation, for example under the threat of being deported. As in the case discussed above, this study also shows how the group faced hostility and racist attacks on social media, to the extent that they had to change the name of the group. Participants had to create new Facebook identities to be able operate without continuous attacks from anti-immigrant and overtly racist groups.

These two studies focused on the experiences of racism and showed that the people who are the targets of the debate can react, discuss and respond to it on social media. Therefore, the case studies provided counterviews to the big data picture that we had drawn from the larger datasets. They also emphasised the different contexts and agency of the racialised subjects in encountering racism and the particular vulnerabilities that the digital-media context produces in this struggle.

The third approach: training the machine

The third approach went beyond big data studies on public communication by exploring, in the form of action research, the infrastructures and technologies of public communication with the aid of big data (Ruckenstein 2019). It used the data captured in previous case studies to intervene in and restructure the public discussion on social media and

the infrastructures that host it. This approach identified new logics of racism and discrimination, emerging due to the ways that digital media and data are used to profile and categorise participants, create different social worlds and produce the new vulnerabilities that some of the interviewed people in our research referred to. These logics are connected to the hidden structures of algorithmic profiling that organise the views, connections and flows of information in digital space. In practice, the third approach involved a study where the researchers of the Hybra project were involved in creating a hate-speech detector based on the data collected in the study (Laaksonen et al., 2020). Initiated by two NGOs, the governmental office of the Non-Discrimination Ombudsman and one software company, the study was part of a pilot project on hate-speech detection to be carried out in Finland for the municipal elections in spring 2017. The aim was to design technical infrastructures to automatically filter potential hate speech based on large social-media monitoring data and to develop measures to react to detected hate speech. A database of openly racist Finnish-language public Facebook groups was used to train the hate-speech detection algorithm. During the process, the research team discussed and reflected on the multiple challenges regarding automated hate speech detection and the unrealistically high expectations of automated hate speech detection by the officials and the NGOs.

The pilot project revealed inconsistencies connected to the automated processes that require ‘rudimentary scales for classifying and a reliance on word-based approaches, while in reality hate speech is a linguistic and social phenomenon with various tones and forms’ (Laaksonen et al., 2020). The researchers concluded that the standard state-of-the-art procedures in machine learning are ill-fitted to capture evolving social and contextual phenomena. In this sense, the project could not be described as an undeniable success; however, it was an eye-opener for the officials and NGOs involved in the process regarding the constraints and limitations of algorithms and automated tools and the value of human-centred practices in identifying complex, contextual phenomena, such as racism and hate speech (Laaksonen et al., 2020). Overall, the case study opened an avenue for considering how big data could be used in research projects, not just as data but as a way to explore and detect biases and mechanisms of discrimination in social-media infrastructures and to question the neutrality of big data itself (Haapoja et al., 2020).

More research needs to be done on the new forms of discrimination and doubled marginality that emerge through data-driven practices. Lack of transparency and access to public databases also increase the risks of misuse and discrimination, as argued by Brucato (2017) in the case of police killings. This third approach can highlight the ways in which data-driven platforms benefit financially from racism (Matamoros-Fernandez, 2017; Noble, 2018; Farkas

et al., 2018) as well as how mechanisms of data-based marginalisation can be resisted through counter-data (Kidd, 2019) and data-justice projects (Dencik et al., 2019). Clearly, data are increasingly used as raw materials for automated decision making and management in various areas of society. This third approach reveals ways to explore and unravel the biases and inconsistencies involved in these practices.

Conclusions

Racism operates in multiple ways across different contexts (Gillborn et al., 2018). In a society that is structured by racial domination, the impact of racism needs to be explored by many different indicators at the same time. While examining racist discourse in the context of migration and refugee issues, research needs to be aware of the ways in which methods may be embedded, draw on and enhance racial inequalities and, therefore, operate as white methods. Drawing on the work by Bonilla-Silva and Zuberi and Scott, the aim has been to explore the methodological challenges of big data research from the perspective of critical race studies and feminist studies.

Critical race studies feminist studies find common ground with critical data studies in pointing out some of the recurring problems with big data such as unproblematised categories and biased and decontextualised data. They emphasize the importance of contextualisation and mixed methods for a more nuanced understanding of experiences and the conditions that shape experience. This paper argues that while big data provides possibilities to identify the dynamics and networks of racism, it has major challenges in understanding issues connected with inequalities and the experience of marginalised subjects that usually require different epistemological approaches.

The three approaches introduced in this paper to explore racism in the context of refugee debate are epistemologically different. The first approach of big data analytics points out the dominant power relations and the circulation of racist discourse by using an inductive approach. It relies on the dominant data structures, thus leaving its possible biases unquestioned. The second approach of complementing big data with qualitative case studies, ethnography and small data emphasises the experience and agency of racialised subjects as valuable reflective knowledge while also understanding that knowledge from experience should avoid essentialist re-categorizing. The third approach of unravelling the new logics of racism questions the epistemological basis of big data by revealing the ways that discrimination may be embedded in big data production and in the use of data for machine learning.

What can we learn from these case studies and their epistemic differences? How insights from critical race studies

and feminist studies can be incorporated to enhance big data methodologies?

The first approach points out the need to open up data for reflexive interrogation of categories and possible built-in biases, as argued in critical race studies. This means that evaluating datasets is crucial for understanding the nature of data and whose views are presented through data. What kind of knowledge interest the composition of data serves? What do we want to understand and why? The first approach also showed the importance of careful conceptual work together with coding: eventually the decision to explore how racism was debated and defined, stemmed from the understanding of the contested nature of the term, as discussed in critical race studies.

One of the main lessons learned concerned the emancipatory importance of agency and the ethical value of anti-racist action in this particular research topic. In the context of marginalization, knowledge from experience can offer valuable insight on the structures, sites and experiences of racial injustice. Therefore it is important to actively expand the research scope for example, to recognise the agency of those who are at the centre of the inquiry and provide space for collaboration with scholars of colour if they are not already included in the research team.

In terms of exploring knowledge from experience, the main challenge is how to incorporate knowledge from experience to big data studies in an anti-essentialist way: how to allow complexity of categories and multiplicity of experience in such research setting? Patterns of inequalities exist but they may operate in different ways in different contexts. Research needs to explore these differences and by doing that, also point out the complexities of social categories.

One way to do this is to move from big data set towards deeper, more contextualized, qualitative approach to shed light on the complexities within and across experience as the second approach showed. Another way is to curate big data from the beginning with different insights from experience, instead of starting from the main stream media and main social media datasets. These insights can inform the different ways in which racism is experienced in different sites and responded to by different racial minorities and groups.

The fact that anti-racist approaches are not in the centre of social scientific tradition and big data studies reflect problems pointed out by Bonilla-Silva and Zuberi (2008). They appear regularly as additional or marginal questions, often reacted to in the course of research rather than in the beginning or are dealt with tokenistic manner. All case studies in this paper highlight the significance of academic institutional context for anti-racist research and for countering white logic. Without emphasis on diversity and its' epistemic value in research it may be difficult to recognize institutional racism and the ways in which it becomes embedded in research. To be able to produce nuanced and

reflexive research is also greatly connected to these conditions of the research. The growing demands of efficiency, speedy publishing, precarious research positions and structural racism in academia, are part of the capitalist white logic, that work against complexities, time-consuming multiple methods, errors, trials, dialogue and collaborations all needed in careful multidisciplinary research to capture the different dimensions of complex phenomenon, such as racism. The use of big data in above mentioned research economy context, adds the tendency to favour decontextualized data from mainstream datasets with fixed categories, without time and space to explore, question and challenge. It follows that it becomes difficult to reflect and react on the epistemological limitations of ones' own research and therefore to be able to combine and develop methods that originate from different disciplines with different epistemological grounds.

No single method can answer or solve all our problems. We need multiple approaches to account for the different articulations of racial power (Chakravartty et al., 2018) and develop computational methods to study racism; this cannot be done without the approaches from critical race studies that can enhance big data studies by offering ways to combine knowledge from experience in an anti-essentialist manner with contextual analysis of the wider social relations, political cultures and technological conditions that shape and enhance racism in society.

Acknowledgments

I would like to thank Leonardo Custódio, Matti Nelimarkka, Salla Maria Laaksonen, Jesse Haapoja, Camilla Haavisto, Jenni Hokka, Mervi Pantti, Reeta Pöyhtäri, Juho Pääkkönen and Gavan Titley for their comments and valuable work preceding this article. I would also like to thank Minna Ruckenstein and the anonymous reviewers for their insightful comments.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Academy of Finland (grant number 295948).

ORCID iD

Kaarina Nikunen  <https://orcid.org/0000-0002-5747-4093>

Notes

1. Scholars have been critical to refer to the arrival of 1,3 million migrants in Europe as European crisis, since the conditions had

been building up years and affected more profoundly countries outside Europe (Nikunen, 2019a).

2. While racism in Finland is often discussed in context of increased migration from 1990s, structural racism has long history in discrimination against Roma and Samí minorities and people of colour (Seikkula, 2019). Finland does not keep official statistics on race and ethnicity. Instead statistics categorize according to language, country of birth and descendance. Largest ethnic minorities originate from Russia, Estonia, Sweden, Iraq and Somalia. All foreign born residents comprise around 7% of the population (https://www.stat.fi/tup/suoluk/suoluk_vaesto_en.html).
3. The project wanted to hire Custódio in 2019, however due to university merge in Tampere the administrative hiring process prolonged and frustrated both the project leader and Custódio. During this time Custódio was granted with post-doctoral fellowship in another university. This is a mundane example of the fragility of critical race studies and multidisciplinary collaboration in academic institutions in the face of institutional changes.
4. The interviews were anonymised, and the participants were referred to in terms of gender and age.
5. The interview data was anonymised and analysed with Atlas.ti programme.

References

- Ahmad A (2020) When the name matters: An experimental investigation of ethnic discrimination in the Finnish labor market. *Sociological Inquiry* 90(3): 468–496. doi:10.1111/soin.12276.
- Andreassen R (2014) The search for the white nordic: Analysis of the contemporary New nordic kitchen and former race science. *Social Identities* 20(6): 438–451.
- Andrejevic M (2020) *Automated Media*. London: Routledge.
- Awan I (2016) *Islamophobia in Cyberspace. Hate Crimes Go Viral*. London: Routledge.
- Baym N (2013) Data not seen: The uses and shortcomings of social media metrics. *First Monday* 18: 10.
- Benjamin R (2019) *Captivating Technology*. Durham: Duke University Press.
- Blei DM (2012) Probabilistic topic models. *Communications of the ACM* 55(4): 77–84.
- Bonilla-Silva E and Zuberi T (2008) Toward a definition of white logic and white methods. In: Zuberi T and Bonilla-Silva E (eds) *White Logic, White Methods: Racism and Methodology*. Rowman & Littlefield pp: Plymouth, 3–27.
- boyd D and Crawford K (2012) Critical questions for Big data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Brock A (2012) From the blackhand side: Twitter as a cultural conversation. *Journal of Broadcasting and Electronic Media* 56(4): 529–549. <https://doi.org/10.1080/08838151.2012.732147>.
- Brucato B (2017) Big data and the new transparency: Measuring and representing police killings. *Big Data & Society* 4(1): 1–5.
- Bulmer M and Solomos J (2004) Introduction: Researching race and racism. In: Bulmer M and Solomos J (eds) *Researching Race and Racism*. London: Routledge, 1–15.
- Buolamwini J and Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* 81: 1–15.

- Burrows R and Savage M (2014) After the crisis? Big data and the methodological challenges of empirical sociology. *Big Data & Society* 1(1): 1–6.
- Chakravarty P, Kuo R, Grubbs V, et al. (2018) #Communicationswhite. *Journal of Communication* 68(2): 254–266.
- Chouliaraki L and Stolic T (2016) Rethinking media responsibility in the refugee “crisis”: A visual typology of european news. *Media, Culture & Society* 39(8): 1162–1177.
- Clark M (2020) Black twitter: Building connection through cultural conversation. In: Rambukkana N (eds) *Hashtag Publics: The Power and Politics of Discursive Networks*. New York: Peter Lang, 205–217.
- Constanza-Chock S (2020) *Design Justice. Community-Led Practices to Build the Worlds We Need*. Cambridge: MIT Press.
- Cooky C, Linabary J and Corple D (2018) Navigating big data dilemmas: Feminist holistic reflexivity in social media research. *Big Data & Society* 5(2): 1–12.
- Crawford K (2013) The hidden biases in big data. *Harvard Business Review*. April 2013.
- Crawford K, Milner K and Gray ML (2014) Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication* 8: 1663–1672.
- Custódio L and Gathuo M (2020) Connections with paulo Freire’s Legacy in anti-racism media activist collaboration in Finland. Commons. *Revista de Comunicación y Ciudadanía Digital* 9(2): 133–158.
- Daniels J (2015) My brain database doesn’t see skin color”. color-blind racism in the technology industry and in theorizing the Web. *American Behavioural Scientist* 59(11): 1377–1393.
- Daniels J (2018) The algorithmic rise of the ‘alt-right.’. *Contexts* 17(1): 60–66.
- Dencik L, Hintz A, Redden J, et al. (2019) Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society: Data Justice* 22(7): 873–881.
- Du Bois WEB (2003) /1920) *The Souls of White Folks*. In *Darkwater: Voices From Within the Veil*. By W. E. B. Du Bois, 55–74. Amherst: Humanity Books.
- Dyer R (1997) *White: Essays on Race and Culture*. London: Routledge.
- Essed P (2004) Naming the unnameable: Sense and sensibilities in researching racism. In: Bulmer M and Solomos J (eds) *Researching Race and Racism*. London: Routledge, 119–133.
- EU-Midis II (2018) Second European Union Minorities and discrimination survey: Being clack in EU. European Union agency for fundamental rights. Luxembourg: European Union.
- Farkas J, Schou J and Neumayer C (2018) Cloaked facebook pages: Exploring fake islamist propaganda in social media. *New Media & Society* 20(5): 1850–1867. <https://doi.org/10.1177/1461444817707759>.
- Ferra I and Ngyen D (2017) Migrant crisis: “tagging” the european migration crisis on twitter. *Journal of Communication Management* 21(4): 411–426.
- Frankenberg R (1993) *White Women, Race Matters: The Social Construction of Whiteness*. Minneapolis: University of Minnesota Press.
- Gangadharan S (2012) Digital inclusion and data profiling. *First Monday* 17(5–7).
- Garner S (2007) *Whiteness. An Introduction*. London & New York: Routledge.
- Gillborn D, Warmington P and Demack S (2018) Quantcrit: Rectifying quantitative methods through critical race theory. *Race Ethnicity and Education* 21(2): 158–179.
- Haapoja J, Laaksonen S and Lampinen A (2020) Gaming algorithmic hate-speech detection: Stakes, parties, and moves. *Social Media and Society* 6: 2. <https://doi.org/10.1177/2056305120924778>.
- Haavisto C (2020) Impossible Activism and the Right to be Understood: The Emergent Refugee Rights movement in Finland. In O. C. Norocel et al. (eds.), *Nostalgia and Hope: Intersections between Politics of Culture, Welfare, and Migration in Europe*, IMISCOE Research Series, https://doi.org/10.1007/978-3-030-41694-2_11.
- Hand M (2014) From cyberspace to dataverse: Trajectories in digital social research. In: Hillyard S and Hand M (eds) *Big Data? Qualitative Approaches to the Digital Research*. Emerald: Bingley, 1–30.
- Hargittai E (2015) Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science* 659: 63–76. <https://doi.org/10.1177/0002716215570866>.
- Harju A (2018) Suomi24-keskustelut kohtaamisten ja törmäysten tilana. *Media & Viestintä* 41(1): 51–74.
- Hegde RS and Shome R (2002) Postcolonial approaches to communication: Charting the terrain, engaging the intersections. *Communication Theory* 12(3): 249–270.
- Horsti K (2008) Europeanization of public debate: Swedish and Finnish news on African migration to Spain. *Javnost–The Public* 15(4): 41–54.
- Hübinette T and Tigerwall C (2009) When racism becomes individualised: Experiences of racialisation among adult adoptees and adoptive parents of Sweden. In: Keskinen S, Tuori S, Irmi S and Mulinari D (eds) *Complying with Colonialism. Gender, Race and Ethnicity in the Nordic Region*. Farnham: Ashgate 119–135.
- Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 1–7.
- Isoaho K, Moilanen F and Toikka A (2019) A big data view of the european energy union: Shifting from ‘a floating signifier’ to an active driver of decarbonisation? *Politics and Governance* 7(1): 28–44.
- Jacobi C, Atteveldt W and Welbers K (2016) Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism* 4(1): 89–106.
- Kennedy H and Moss G (2015) Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society* 2(2): 1–11.
- Keskinen S (2018) The crisis of white hegemony, neonationalist femininities and antiracist feminism. *Women’s Studies International Forum* 68(May-June): 157–163.
- Kidd D (2019) Extra-activism: Counter-mapping and data justice. *Information Communication and Society* 22(7): 954–970.
- Laaksonen S, Pantti M and Titley G (2020) Broadcasting the movement and branding political microcelebrities: Finnish anti-immigration video practices on YouTube. *Journal of Communication* 70: 2.
- Laaksonen S-M, Haapoja J, Kinnunen T, et al. (2020) The datafication of hate: Expectations and challenges in automated hate speech. *Frontiers in Big Data* 3: 3.

- Larsson A and Moe H (2011) Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media and Society* 14(5): 729–747.
- Lagus K, Pantzar M, Ruckenstein M and Ylisiurua M (2016) Suomi 24. Muodonantoa aineistolle. Valtiotieteellisen tiedekunnan julkaisuja. 2016:10. Helsinki university.
- Lipsitz G (2006) *The Possessive Investment in Whiteness*. Philadelphia: Temple University Press.
- Lundström C and Teitelbaum B (2017) Nordic whiteness: An Introduction. *Scandinavian Studies* 89(2): 151–158.
- Matamoros-Fernandez A (2017) Platformed racism: The mediation and circulation of an Australian race-based controversy on twitter, facebook and YouTube. *Information, Communication & Society* 20(6): 930–946.
- Matamoros-Fernandez A and Farkas J (2021) Racism, hate speech, and social Media: A systematic review and critique. *Television and New Media* 22(2): 205–224.
- Marwick A and boyd D (2010) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience *New Media & Society* 13(1): 114–133.
- Mazzocchi F (2015) Could Big data be the end of theory in science? *Science and Society EMBO Reports* 16(10): 1250–1255.
- McCall L (2005) The complexity of intersectionality. *Signs* 30(3): 1771–1800.
- Metcalf J and Crawford K (2016) Where are human subjects in Big data research? The emerging ethics divide. *Big Data & Society* 3(1): 1–14.
- Mohanty C (2003) “Under western eyes” revisited: Feminist solidarity through anticapitalist struggles. *Signs: Journal of Women in Culture and Society* 28(2): 499–535.
- Nelimarkka M (2021) Social Science Coding. Forthcoming.
- Ng K, White C and Saha A (2020) #Communicationsowhite: Race and power in the academy and beyond. *Communication, Culture & Critique* 13(2): 143–151.
- Nikunen K (2019a) *Media Solidarities: Emotions, Power and Justice in the Digital Age*. London: Sage.
- Nikunen K (2019b) Once a refugee: Selfie activism, visualized citizenship and the space of appearance. *Popular Communication: The International Journal of Media and Culture* 17(2): 154–170.
- Noble S (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Pantti M, Nelimarkka M, Nikunen K, et al. (2019) Meanings of racism: Public discourses about racism in Finnish news media and online discussion forums. *European Journal of Communication* 34(5): 503–519.
- Poole E, et al. (2019) Contesting #StopIslam: The dynamics of a counter-narrative against right-wing populism. *Open Library of Humanities* 5(1): 1–39. <https://doi.org/10.16995/olh.406>.
- Pöyhtäri R, Nikunen K, Nelimarkka M, et al. (2019) Refugee debate and networked framing in the hybrid media environment. *International Communication Gazette* 81: 4.
- Rastas A (2007) *Rasismi Lasten ja Nuorten Arjessa. Transnationaalit Juuret ja Monikulttuuristuva Suomi*. Tampere: Tampereen yliopisto.
- Rastas A (2020) Involving diaspora communities through action research. A collaborative museum exhibition on the african presence in Finland. In: Espinoza Garrido F, Koegler C, Nyangulu D and Stein MU (eds) *Locating African European Studies*. London: Routledge, 79–91.
- Rastas A and Nikunen K (2019) Introduction: Contemporary african and black diasporic spaces in Europe. *Special Issue in Open Cultural Studies* 3(1): 207–218.
- Rastas A and Seye E (2018) Music and anti-racism: Musicians’ involvement in anti-racist spaces. *Popular Music and Society*. 42(5): 592–610. DOI:10.1080/03007766.2018.1526527.
- Richardson R, Schultz J and Crawford K (2019) Dirty data, bad predictions: How civil rights violations impact police data, predictive policing and justice. *New York University Law Review* 94(May): 192–233.
- Ruckenstein M (2019) Tracing medicinal agencies: Antidepressants and life-effects. *Social Science & Medicine*: 235 (2019):1123682.
- Scott J (1991) The evidence of experience. *Critical Inquiry* 17(4): 773–797.
- Scott J (2001) Fantasy echo: History and the construction of identity. *Critical Inquiry* 27(2): 284–304.
- Seikkula M (2019) Adapting to post-racism? Definitions of racism in non-governmental organization advocacy that mainstreams anti-racism. *European Journal of Cultural Studies* 22(1): 95–109.
- Sharma S (2013) Black twitter? Racial hashtags, networks and contagion. *New Formations* 78(1): 46–64.
- Siapera E (2019) Organised and ambient digital racism: Multidirectional flows in the Irish digital sphere. *Open Library of Humanities* 5(1): 1–34. <https://doi.org/10.16995/olh.405>.
- Siapera E, Boudourides M, Lenis S, et al. (2018) Refugees and network publics on twitter: Networked framing, affect, and capture. *Social Media + Society* 4(1): 1–21. <https://doi.org/10.1177/2056305118764437>.
- Sobande F (2021) Spectacularized and branded digital (re)presentations of black people and blackness. *Television and New Media* 22(2): 131–146.
- Song M (2014) Challenging a culture of racial equivalence. *The British Journal of Sociology* 65(1): 107–129.
- Titley G (2019) *Racism and Media*. London: Sage Publications.
- Trottier D (2014) Big data ambivalence: Vision and risks in practice. In: Hillyard S and Hand M (eds) *Big Data? Qualitative Approaches to the Digital Research*. Emerald: Bingley, 51–74.
- Van Dijk TA (1991) *Racism and the Press: Critical Studies in Racism and Migration*. London: Routledge. <https://doi.org/10.1145/3240431.3240436>.
- Van Leeuwen T and Wodak R (1999) Legitimizing Immigrant control: A Discourse-Historical Analysis. *Discourse Studies* 1(1): 83–118.
- Welles B (2014) On minorities and outliers: The case for making Big data small. *Big Data & Society* 1(1): 1–2.
- Wojcik S and Hughes A (2019) Sizing up Twitter users. Pew Research Center. Available at: file:///C:/Users/kn27827/AppData/Local/Temp/twitter_opinions_4_18_final_clean.pdf.
- Zook M, Barocas S, boyd d, et al. (2017) Ten simple rules for responsible big data research. *PLoS Computational Biology* 13(3): 1–10. <https://doi.org/10.1371/journal.pcbi.1005399>.