Feminist Data Studies and the emergence of a new Data Feminist knowledge domain

by Hana Marčetić and Jan Nolin

Abstract
Mass participation in social networking sites and online life combined with the development of tracking technology facilitates gathering data on unprecedented scales. The uptake of data collecting during the 2010s coincided with the emergence of data science and data studies, along with critical perspectives such as critical data and critical algorithm studies. This paper explores one such critical perspective. Data Feminism merges the theories of intersectional feminism and critical data studies. Bibliometric text analysis of articles, conference papers, essays, and commentary was conducted in VOSviewer software, which found commonalities between terms within texts. The most prominent terms and keywords in the research area of Data Feminism identified in such a manner informed the close reading that followed. Six clusters of terms were identified, with the two largest clusters formed around the terms “big data” and “artificial intelligence” respectively. We also explored the boundaries, movements and centralities within the six clusters.

Contents

Introduction
Previous research
Feminist Data Studies and Data Feminism
Method
Results
Discussion
Conclusion

Introduction
Types of knowledge and subject areas form knowledge domains, allowing researchers to develop a specific language to talk about subject areas in unique ways (Hjørland, 2002). According to Hjørland, each domain has its own approaches, paradigms and schools, and their logic should be understood from within. This article focuses feminist theories and perspectives that study contemporary critical issues on how data
interacts with people. The starting point is that such studies serve as the setting for the emergence of a new knowledge domain within feminist studies.

It is difficult to evaluate one knowledge domain from the standpoint of another. For this reason, we approach our task with caution. Our ambition is to provide new insights into the emerging knowledge domain of Data Feminism through a combination of bibliometrics and close reading. Bibliometric tools are used to understand boundaries, tensions and movements within a field (Farooq, 2021; Flensburg and Lonburg, 2021; Hammerfelt, 2018), research subject (e.g., Mora, et al., 2017), a journal or a database (e.g., Macan, 2011).

There are several obstacles to grasping the breadth and character of the emerging Data Feminist knowledge domain. For instance, citation databases favor Western science and the English language (Albarillo, 2014; Van Leeuwen, et al., 2001), have bias issues and create the Matthew effect (Merton, 1968). The Matthew effect implies that articles and authors that are already well cited will be referenced even more. Similarly, knowledge domains that are already well represented as major subject areas will grow exponentially more than the fledgling knowledge domain. With that in mind, it is perhaps not surprising that subject indexing in major databases tends to disregard feminist research (Bergenmar and Golub, 2020). Block (2020) argued that the absence of symmetrical indexing has led to systematic misrepresentation and erasure of intersectional and feminist research.

Given that, this article provides a much-needed overview and discussion of key emerging themes of research. Extensive citation chaining was used in order to reconstruct what was not visible in key databases. This approach considerably added to conventional searches in a variety of discovery systems. We found that, on the one hand, databases have started recognizing Data Feminism as a term over the last two years. This coincides with the publication of the book *Data Feminism* (D’Ignazio and Klein, 2020a). On the other hand, different feminist approaches to studying data occurred before the term was coined.

Academic studies that contributed to feminist discussions of what is increasingly understood as a *datafied society* largely appear scattered across fields and journals. The key critical concept of datafication alludes to “the transformation of social action into online quantified data” as a means to “access, understand and monitor people’s behaviour” [1]. In 2020, Data Feminism was launched as an umbrella, offering principles in place of a programmatic agenda (D’Ignazio and Klein, 2020a). Due to the diversity of scholarly work, studies conducted before the development of the knowledge domain of and outside of this pragmatic umbrella are in the text below referred to as Feminist Data Studies. This term is used in several instances within the literature, which motivated us to choose it as such a delineator. With the emergence of Data Feminism, the outlines of a new knowledge domain are beginning to take shape. Given that, the question of how Data Feminism is shaping arises. The research question that follows from that is:

**Q1**: What issues have been the focus of the longer trend of Feminist Data Studies and the new umbrella of Data Feminism?

This study is an exploratory attempt at outlining boundaries, tensions, movements and centralities within Data Feminism and Feminist Data Studies. Publications and materials collected using scholarly databases, discovery and recommendation systems as well as citation chaining were analyzed using bibliometric software and combined with close reading. Using said software we have clustered the terms that occurred together in the texts and asked the following question:

**Q2**: What are the strengths within Feminist Data Studies and Data Feminist approach when studying *datafied society*?

In this paper, we provide an overview of previous research of critical discussions unravelling within the domains of technological research, followed by an overview of Feminist Data Studies and Data Feminism. We then describe the tools and methods used in this study. The findings are presented through five clusters.
Feminist Data Studies and the emergence of a new Data Feminist knowledge domain

of terms that were identified with bibliometric tools. Finally, in our discussion we outline the conversations happening within each of the clusters as informed by close reading, allowing us to draw careful conclusions.

Previous research

Over the past decade, a multitude of disciplines and multidisciplinary fields has critically engaged with numerous issues connected to the emergence of data-fueled targeted advertising. A shift in the utilization of user data became commercially viable already in the mid-2000s. The spread of targeted advertising across companies has led to a plethora of applications in all sectors of society (O’Neil, 2017). Targeted advertisement has been driven by big data surveillance (Zuboff, 2019), setting the foundation for a datafied society.

Critical scholars have warned about the impact of data-fueled systems on vulnerable groups and communities. Surveillance through the use of metadata has been termed dataveillance (Raley, 2013; van Dijck, 2014). There has been concern about both the character and legitimacy of the divide in knowledge between data collectors and those whose data is being harvested (boyd and Crawford, 2012; Andrejevic, 2014; Zuboff, 2019). Predictive models based on uncritically utilized mathematical models and algorithms, frequently use biased and re-purposed data to influence decisions about employment (O’Neil, 2017) and drive excessive surveillance (Zuboff, 2019). The use of predictive models for policing minorities (Browne, 2015; Alexander, 2019), as well as the continual development of technology without taking minorities and their cultures into account (Benjamin, 2019), call for rethinking and reimagining the way data-driven innovation is thought about and discussed.

The potential of the Internet as a tool for “activism, communication, media-making, and culture is often the focus of research about the Internet (...) but without sufficient power critique” (Noble, 2016). The feminist research tradition has a unique approach in criticizing and analyzing power structures, allowing for specific and unique discussions of datafication.

Feminist Data Studies and Data Feminism

Various theories and strands of feminism were utilized in the critique of datafication and automation since the 1990s, increasingly so in the 2000s. Feminist data values are expressed comprehensively for the first time in a “Manifest-no” declaration, refusing harmful data regimes and commitment to new data futures (Cifor, et al., 2019). Feminist Data Studies follow a diverse set of perspectives and theories. They have been scattered within numerous disciplines and publishing areas (to be discussed later) and with surprisingly scant co-citation between articles. Notably, across those studies there have been references to the likes of Haraway (1988) and Harding (1991, 1986). For example, standpoint and situated knowledge theories have been influential in arguing for better representation of women in tech environments as well as for more inclusion of minorities’ perspectives online (Atanasoski and Vora, 2019). Another theme common in feminist studies has been a consistent criticism of hierarchies (Yoder, 2018; Terborg-Penn, 1995). This was the context in which Data Feminism was introduced.

The concept of Data Feminism was introduced to scholarly bibliographic databases quite recently. The first time it appeared in the Web of Science and Scopus was in 2020. D’Ignazio and Klein’s (2020a) book Data feminism popularized the concept, catching the attention of scholars in various disciplines. Data Feminism was described as “a way of thinking about data, both their uses and their limits, that is informed by direct experience, by commitment to action, and by intersectional feminist thought” [2]. It remains to be seen
Feminist Data Studies and the emergence of a new Data Feminist knowledge domain

what strands and directions will influence the development and future directions of Data Feminism and Feminist Data Studies.

Method

In this study, we combined bibliometrics with a close reading of texts retrieved from databases and by citation chaining. Relevant hits were identified in the Web of Science, Scopus, ProQuest, Taylor & Francis, Sage, Emerald, EBSCO Discovery Service and Google Scholar. The initial searches were conducted in November and December of 2020. Final items were added to the collection during Spring 2021. The search terms and combinations of terms that resulted in relevant literature are listed in Table 1. The stemming of the terms is shown according to the standard in Web of Science and was adjusted to the requirements of each database when searches were conducted.

<table>
<thead>
<tr>
<th>“data femin*” or feminis* AND “surveillance studies” AND</th>
</tr>
</thead>
<tbody>
<tr>
<td>• “critical data studies”</td>
</tr>
<tr>
<td>• “data studies”</td>
</tr>
<tr>
<td>• “data science”</td>
</tr>
<tr>
<td>• “data activis*”</td>
</tr>
<tr>
<td>• “data ethics”</td>
</tr>
<tr>
<td>• datafication</td>
</tr>
<tr>
<td>• algorithm*</td>
</tr>
<tr>
<td>• “Big Data”</td>
</tr>
<tr>
<td>• “artificial intelligence”</td>
</tr>
<tr>
<td>• “machine learning”</td>
</tr>
<tr>
<td>• data AND feminis*</td>
</tr>
<tr>
<td>• “data science” AND gender</td>
</tr>
<tr>
<td>• “critical race studies” AND “data stud*”</td>
</tr>
</tbody>
</table>

Searching additional combinations and terms (e.g., data economy) did not provide any relevant hits.

After sorting the datasets, the number of relevant results amounted to two dozen, which served as a basis for further citation chaining. The final number amounted to 103 documents (see Appendix).

For the purpose of visualizing data, VOSviewer, software for the visualization of similarities in bibliographic networks, was selected in order to identify the most important concepts as well as core terms related to each other. A master file of titles, abstracts and author-assigned keywords was created. We used textual analysis functionality, introduced in 2011, to create a term map, a “two-dimensional map in which terms are located in such a way that the distance between two terms can be interpreted as an indication of the relatedness of the terms” [3]. The relatedness was determined based on the co-occurrence of the terms in the documents. The dataset was imported and the binary counting method selected. Of 3,004 terms identified, 172 met the threshold of five occurrences that we chose for the analysis. We then manually excluded the word data, which occurred in the dataset significantly more often than the rest of the terms, and those referring to the document structure instead of the topic (e.g., abstract, research, researcher, article, paper, study and author). Ultimately, we were left with 130 terms, for which we generated network
Feminist Data Studies and the emergence of a new Data Feminist knowledge domain

and density visualizations in VOSviewer. In the network visualizations, the size of a node (circle) represents the higher importance (weight) of the term it represents (van Eck and Waltman, 2020). Different colors represent clusters within the network, and each term was assigned to one cluster.

Following the analysis of keywords in VOSviewer, the collected literature was categorized according to clusters that emerged from the visualizations. We identified prominent terms within clusters and engaged in a close reading of the texts. During close reading, we identified the ways in which said terms were used and the contexts in which they figured in the texts. With these procedures, we were able to investigate the context of important concepts through both quantitative and qualitative approaches.

Results

Six clusters of terms were identified, with the two largest formed around the terms “Big Data” and “Artificial Intelligence”. Figure 1 depicts the network of terms, each color representing one of the clusters, and each node in the network signifying one of 130 terms. The term Big Data occurred most frequently within the dataset, followed by gender and ethic(s). The largest cluster, with 59 terms, is shown in red. The smallest, with 10 terms, is shown in the lighter shade of blue (Figure 1). In the following sections, we will take a closer look at each of the six clusters.

<table>
<thead>
<tr>
<th>Table 2: Six clusters with most prominent terms and number of occurrences.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 1 — Big Data</strong></td>
</tr>
<tr>
<td>Big Data (71)</td>
</tr>
<tr>
<td>gender (70)</td>
</tr>
<tr>
<td>technology (51)</td>
</tr>
<tr>
<td>algorithm(s) (43)</td>
</tr>
<tr>
<td>care (33)</td>
</tr>
<tr>
<td>feminism (33)</td>
</tr>
<tr>
<td>body (32)</td>
</tr>
<tr>
<td>knowledge (30)</td>
</tr>
<tr>
<td>bias (25)</td>
</tr>
</tbody>
</table>

| **Cluster 4 — Identity** | **Cluster 5 — Individual and societal** | **Cluster 6 — Algorithmic culture and the pandemic** |
| user (30) | process (44) | world (17) |
| woman (30) | experience (28) | culture (16) |
| person (25) | individual (25) | covid (9) |
| identity (16) | decision (16) | policy (9) |
| software | | |
Cluster 1: Big Data

Big Data was a dominant concept in the literature within this dataset. Long-standing feminist discussions on gender and minorities were applied to Big Data in this cluster (e.g., van Oost, 2020; Schiller and McMahon, 2019). Micro-sociological approaches concerned with body politics in local settings were present in research on LGBTQ+ groups and identities (e.g., Gieseking, 2018; Snapp, et al., 2016) and body autonomy (Bogers, et al., 2020). The velocity of data available to data brokers creates conditions that enable obscuring agency, context and subjectivity of “individuals who produce ‘data points’” [4]. Macrosociological approaches were also visible. Processing Big Data to garner economic gain and social control for a handful of companies was criticized as unjust and problematic (Suárez-Gonzalo, 2019; Luka and Millette, 2018). Different authors identify four to five companies reaping the majority of the benefits of datafication. Those companies are known under acronyms GAFA or GAFAM (i.e., Google, Amazon, Facebook, Apple and Microsoft), with Google and Facebook at the forefront. Feminist empiricism (McQuillan, 2016) and employment of situated knowledges (by Draude, et al., 2019; Cooky, et al., 2018; Luka and Millette, 2018) occurred within the literature as feminist tools for inspecting and challenging the imbalance in power and harms stemming from the misuse of Big Data. The micro-sociological contributions were frequently and uniquely feminist while some macro-sociological studies were similar in character as other critical discussions on Big Data where GAFA had been targeted since at least 2013 (De Filippi, 2013).

Algorithms, technology and gender were prominent terms in Cluster 1 as well. Figure 1 depicts the algorithm node as deeply intertwined with others within the network. Algorithms were discussed within the dataset as social constructs that subtly wield power. The literature within this cluster included topics of algorithmic visibility (Bishop, 2019; Agostinho, 2018), reproduction of societal values (Hayes, et al., 2020) and social relations of power (Prietl, 2019). Agostinho discussed visibility in the context of algorithmic role in making racialized bodies visible. Algorithms that were developed on biased datasets overwhelmingly represented bodies and experiences of white people, primarily men (Browne, 2015). Agostinho’s (2018) study touched on the issue of surveillance of racialized bodies on the one hand, and the lack of their visibility on the other. Bishop (2019), in turn, described how content creators managed algorithms to achieve visibility on a social media platform, and the ways in which gender played a role in strategies that were used. Reproduction of societal values occurred through algorithms and their bias, visible, for example, in talent acquisition software and a systematic lack of hiring underrepresented candidates (Yarger, et al., 2020). This provided examples of algorithms leading to the neglect of minorities. In other cases, such as predictive policing, minorities and low income communities, received disproportionately amount of attention (Hayes, et al., 2020). This has been discussed largely in an American context (Eubanks, 2018; O’Neil, 2017; Browne, 2015).
Feminist Data Studies and the emergence of a new Data Feminist knowledge domain
Bogers, et al. (2020) used a term suggested by Noble (2016), algorithmic oppression, when referring to inscription of oppression into coded systems. They remarked, “[t]oo often, computational analytics like algorithms are accompanied by the firm belief that data are intrinsically neutral and bias-free” [5]. Algorithms have the power to shape material realities and are hence of concern to feminist inquiries (Adams, et al., 2020). Feminist inquiry and reasoning support a call for more accountability in studies of talent recruitment software (Yager, et al., 2020), online representations of pregnancy (Bogers, et al., 2020), algorithmic decision-making systems (Prietl, 2019) and gender bias in social media (Schroeder, 2021).

**Cluster 2: Artificial intelligence**

The artificial intelligence (AI) node occurred as the most prominent term in Cluster 2. Critical research in this area incorporated issues of data used for training AI and reflected the intended or unintended bias of data and algorithms. If training data was selected from contexts rife with asymmetrical power structures, these would be reproduced and amplified through machine learning.

The AI node was intertwined with some of the topics that reoccurred within the literature. This was why we could see nodes such as algorithmic bias, power and inequality in this cluster as well. The issue of inequality in power distribution across society and its reflection in AI systems figured in much of the reviewed literature (e.g., Ciston, 2019; Mohamed, et al., 2020). In mainstream discussions, bias in AI was often overlooked since the “[c]ultural misunderstandings of AI allow their resulting data to assume a status of impartial fact, even as they operate by human intervention at every level” [6]. Much like algorithms, artificial intelligence has been a subject of inquiry with regards to agency and accountability. An intersectional framework was suggested in the literature to “analyze the biases and problems built into existing artificial intelligence, as well as to uncover alternative ethics from its counter-histories.” [7].

Algorithmic oppression and patriarchal oppression (Johnson, 2005) are, at least to an extent, similar in character. AI is increasingly invested with the power to sort people through algorithms while at the same time being omitted from processes of accountability. Feminist and gender research appeared to be well positioned to identify and critically discuss new and evolving forms of bias, creating room for further research.

The trajectory of AI has been notoriously difficult to grasp. Nonetheless, critical scholars have been compelled to try. Studies of the future of AI (Ferrando, 2014) and the ways in which AI and humans mutually shape each other (Krupiy, 2020; Mohamed, et al., 2020) appeared in this cluster. Ferrando stated that futures “do not appear out of nowhere”; instead “they are based on the presents, the pasts, and the ways they are being envisioned” [8]. She asserted that the future of AI was still developing under a predominantly male imagination. From such a perspective, the emergence of a new feminist knowledge domain would appear vital.

**Cluster 3: Ethics**

The topic of ethics appeared in Cluster 3. Large amounts of personal, intimate data were produced about people when they became hospitalized. Such data had traditionally been seen as highly sensitive and strictly protected. However, with the explosion of Internet-based tracking technologies, geolocation and social media there was a serious leaking of this sensitive information. Several researchers argued for the importance of feminist ethics of care for analyzing and scrutinizing the manifestations of power relations in datafied systems. Feminist ethics of care builds on Puig de la Bellacasa (2017) for whom care incorporated active involvement in maintenance and empowerment of the object of care. In Cluster 3 we noted that researchers advocated an application of such ethics to handle research data (Leurs, 2017), Big Data in social media settings (Luka and Milette, 2018) and to understand citizen data practices (Fotopoulou, 2019). Health
related data practices have received special attention since the start of the global pandemic in 2020. The adoption of surveillance technologies provoked by the pandemic has been a cause of particular concern (Taylor, 2020; Robinson and Johnson, 2021). Taylor (2020) argued that policy and science handling of the pandemic moved forward with a profound lack of data on low-wage workers, the elderly and various minorities. This appeared to be another area where feminist and gender research could make prolific contributions.

**Cluster 4: Identity**

Cluster 4 dealt with identities, containing the nodes *identity, woman, transgender* and *LGBTQ*. Data Feminist studies have and will continue to make important contributions about representation and coding gender binary in social media (Bivens and Haimson, 2016; Bivens, 2017) and software (Spiel, *et al.*, 2019). Cluster 4 was connected to Cluster 3 through the node *race*. The node was situated within the Ethics cluster, but was relevant for the Identity cluster as well. However, the articles in this cluster were concerned largely with issues related to gender and sexual orientation.

In this literature, identities were predominantly understood as complex, fluid and nuanced, which could be contrasted with the coding of Big Data categories which predominantly followed traditional binaries. The new ways of conceptualization and capturing identity through information technology was therefore creating a friction between LGBTQ+ identities and the way they have been represented (Hamidi, *et al.*, 2018; Ruberg and Ruelos, 2020). Technology driven misrepresentations and reductionism of queer and non-binary identities have been common and have real life consequences for groups and minorities pushed further to the margins through datafication. Subsequently, several articles called for reimagining and queering (Gieseking, 2020; Ruberg and Ruelos, 2020) identity-based data.

There were also concerns raised by the appearance and potential perils of automatic gender recognition systems (AGR) (Hamidi, *et al.*, 2018). Here we found an interesting widening of a long-standing criticism of societal institutions built upon limited conceptions of gender. Similar avenues of thought are intriguing in the context of automatic machine decision-making applied to gendered identities.

**Clusters 5 and 6**

Clusters 5 and 6 were the two smallest clusters in the network. Cluster 5 contained nodes including *individual, experience, decision, society* and *institution*. Nodes *covid, policy* and *health* were situated in Cluster 6, as well as *culture* and *algorithmic culture*.

Issues of ethics within institutions connected Cluster 5 to Cluster 3. Cluster 5 accentuated the interplay between the individual and institutions within the datafied society. In essence, the argument was that a new social contract between citizens and governing institutions will need to be developed. Suarez-Gonzalo (2019) argues that “it is critical to reduce the factors that give a few corporations the power to undermine citizens’ ability to act autonomously for the protection of their personal data. This requires that public institutions replace the logic of data business for the logic of fundamental rights, limiting the unbridled expansion of the markets to data” [9]. It was a strong statement but quite similar to what other critical traditions have put forward, for instance, within surveillance studies (Zuboff, 2019). What are the contemporary digital rights and obligations of the citizen?

The concern with ethics also signaled a penchant for philosophical reasoning. The institutionalization of ethics was criticized for reducing ethics to a set of operationalized practices and procedures. Subsequently, there was a tendency to resolve open-ended and irresolvable ethical questions prematurely (Metcalf, *et al.*, 2019). Building further on the private-public dichotomy, the nodes individual and experience appeared in relation to identities and tie it to Cluster 5. The notion of direct experience used by D’Ignazio and Klein should have been invoked here, to postulate how the lack of diversity in the tech industry reflected representation of diverse human experiences and minority identities. While the issue of race was often intricately connected and interwoven in the studies and arguments within our material, it was seldom the main focus.
Cluster 6, the smallest in the map of terms, contained nodes announcing topics of the latest pandemic, but also culture. The algorithmic culture node contained an instance in which the “interconnectedness and mutual shaping of society and technology” [10] was highlighted, while the pandemic related literature focuses on issues including pandemic driven technology adoption (Robinson and Johnson, 2021) and the politics of pandemic data (Taylor, 2020). The global pandemic has influenced the direction of much research being done in recent years, and that trend is becoming visible in Data Feminist research as well. In this cluster we find a suggestion for incorporating seven aforementioned principles in pandemic related research (D’Ignaizo and Klein, 2020b). Special attention is given to working with data in the face of increasing of remote work. In this way, the principles of Data Feminism are applied to and contribute to the ongoing discussion of datafication and how it has developed during the pandemic.

Discussion

The term Big Data had the most occurrences (71) in the network of terms (Figure 1). It appeared in cluster 1, the “Big Data cluster”. Cluster 1 also had a central position in the network, connecting the remaining five clusters. The term gender was also situated in this cluster, and figures as the central term of the network with 70 occurrences. Gender studies point out the prevalence of masculine concepts within tech jargon (van Oost, 2000; Harding, 1986). Big Data is one such concept, which corresponds to the (Western) masculine attempts to control nature through technology (van Oost, 2000). Notably, the Big Data feminist literature explores themes beyond gender, establishing a critical tradition that is both closely connected to the strengths of this approach and more broadly oriented towards intersectional problems on the Internet.

Trendy technological concepts of Big Data, Algorithms and Artificial Intelligence were prevalent in the material, and imported from other disciplines and technological jargon. They stemmed from techno-optimistic discourses on whose boundaries critical discussions developed. Data Feminist Studies and Data Feminism utilized and contributed to those critical discussions. Much of the material used in this study pertained to Big Data (Favaretto, et al., 2019; Cooky, et al., 2018; Hill, et al., 2016), algorithms (Adams, et al., 2020; Prietl, 2019; Yarger, et al., 2020; McQuillan, 2016) and AI (Ciston, 2019; Baker, 2018; Ferrando, 2014). There was space for reformulating these technooptimistic concepts, and good reason to reconfigure them in ways that communicated the traditional criticisms of feminist studies. Critical work on the concept of algorithm was an example of that. In mainstream technological discussions about algorithms were portrayed as neutral, driven by pure mathematics without any politics. Feminist investigations have demonstrated that skewed input into algorithms manifests itself as bias. Such work can lead to appropriation of the algorithm concept for critical discussions. Coining and utilizing terms such as algorithmic oppression and algorithmic bias (Noble, 2016; Haider and Sundin, 2019) are examples of reconfiguring the concept of algorithms to foster more critical discussions. Further development of Data Feminist studies might find more ways to contribute to such reconstructions.

Hegemonic tendencies in the developments of the Internet (Frenkel and Kang, 2021) driven by masculine thinking were intertwined with a lack of representatives of minority professionals in developing technological structures (Yarger, et al., 2020; Way, et al., 2016). Big Tech companies claim they want to do better with regards to diversity and ethics (Hoffmann, 2020; Metcalf, et al., 2019). However, employing women and minorities, in particular at jobs that require competences beside tech development rarely results in real change in company cultures. We can see this in the recent firing of co-leaders of Google’s ethical team (Simonite, 2021) and the positions minorities occupy in Big Tech in general (Yarger, et al., 2020).

Deconstruction of identity is a strength of feminist research (Cluster 4). Queer and feminist theories hold that identity changes and reshapes (Nicholas, 2014; Butler, 1990), and is not bound by biological boundaries a person is born with. Data Feminist researchers argue for queering datafied representations of gender (Hoffmann, 2020; Snapp, et al., 2016) and oppose the gender binary reinforced in datafied systems
Feminist Data Studies and the emergence of a new Data Feminist knowledge domain

(Ruberg and Ruelos, 2020; Schroeder, 2021; Bivens, 2017). This leads to an interesting dilemma. On the one hand, feminists want to rectify systematic misrepresentation of women and other minorities by algorithms. On the other hand, richer data representations of minorities expose vulnerable groups to dataveillance and policing through data (Yu, 2020; Weinberg, 2017). In other words, the issue of representation coexists with the perils of dataveillance in these texts.

The tradition of emphasizing power distribution and inequality in feminist studies is another advantage when critically studying the tech industry. Filtering the contents it delivered on individual feeds and feeding into power structures poses opportunities for feminist researchers to re-identify and deconstruct power relationships and implicit power structures.

The topics of healthcare arose from Cluster 6, particularly in the context of the global pandemic. As a female dominated discipline (Shannon, et al., 2019), healthcare research often utilized feminist theories and perspectives (e.g., Healy, 2021; Gilberg, 2018). The increase of automatisation of healthcare and medicine (Cheney-Lippold, 2017; O’Neil, 2017) makes healthcare an interesting domain for feminist inquiries. A side effect of the pandemic has been an increased interest in these research areas inside and outside of the academia. There is therefore a certain window of opportunity within which critical feminist perspectives within healthcare research may find a wider audience.

The pandemic has influenced Data Feminism in several ways. Firstly, it has drawn the attention of Data Feminist researchers to study pandemic-related topics (e.g., D’Ignazio and Klein, 2020b; Robinson and Johnson, 2021; Taylor, 2020; Yu, 2020). Secondly, a portion of scholars interested in Data Feminist research may be due to challenges posed as a result of a global pandemic (e.g., Oleschuk, 2020; Staniscuaski, et al., 2020). And lastly, while some of those interested in Data Feminism may have become hampered by the pandemic, others may still have become involved in pandemic research, rather than joining the development of this knowledge domain.

Conclusion

Internet studies have grown to involve a wealth of different approaches. Various critical knowledge domains are still evolving. This study focused on a new critical knowledge domain within feminist and gender studies. The feminist tradition of criticizing politics, technology and science under scrutiny for being male dominated has been given a new object of study in the datafied society. As this knowledge domain so far lacks conventional institutionalization, it is difficult for interested researchers to get a grasp of it. This paper has served to track boundaries, centralities, tensions and movements over time by applying bibliometrics.

Despite the issues with representation of Data Feminist and Feminist Data Studies that modern indexing systems and search tools encounter, it is possible to note an increase in the number of published studies and the outlines of a new knowledge domain. Major terms identified in our dataset include studies of Big Data, gender, algorithms, Artificial Intelligence and bias.

Our review of the literature has shown that feminist approaches make numerous and different kinds of contributions to discussions of the datafied society.

- Original epistemological contributions through revitalizing theories of standpoint epistemology (Harding, 1991; 1986) and situated knowledge (Haraway, 1988).
- Emphasizing aspects of traditional feminist critique based on an understanding of patriarchal hierarchies.
- Extending critiques developed by other fields, such as software studies (Lessig, 2006), platform studies (van Dijck, et al., 2018; Vaidhiyanathan, 2022) and surveillance studies (Zuboff, 2019).
Applying an intersectional approach, highlighting issues of data justice sexism and racism. Criticizing practices of biased, disembodied and non-binary data processing.

Will there be a merging of Data Feminism and Feminist Data Studies? Will one assimilate the other or will the two strands coexist, one with a strong program and the other that might not want to be a part of it? Whatever path Data Feminist studies take, these discussions will still be necessary and require more scrutiny. This article is merely exploratory in character, an attempt at outlining these knowledge domains for further study. This work has combined quantitative and qualitative approaches; future research could include more analyses qualitative in character.

About the authors

**Hana Marčetić** is a doctoral student at the Swedish School of Library and Information Science at the University of Borås, Sweden. Her research interests include critical data and algorithm studies, identity and social media.
Send comments to: hana [dot] marcetic [at] hb [dot] se

**Jan Nolin** is a professor at the Swedish School of Library and Information Science at the University of Borås, Sweden. His research interests include Internet studies, transparency movements, and strategies concerning academic involvement in Agenda 2030.
E-mail: jan [dot] nolin [at] hb [dot] se

Notes

2. D’Ignazio and Klein, 2020, p. 3.
6. Ciston, 2019, p. 3.
9. Suarez-Gonzalo, 2019, p. 188.

References


Feminist Data Studies and the emergence of a new Data Feminist knowledge domain


Feminist Data Studies and the emergence of a new Data Feminist knowledge domain


doi: https://doi.org/10.1057/s41305-017-0043-1, accessed 16 August 2021.

Mary Elizabeth Luka and Melanie Millette, 2018. “(Re)framing Big Data: Activating situated knowledges and a feminist ethics of care in social media research,” *Social Media + Society* (2 May).


Feminist Data Studies and the emergence of a new Data Feminist knowledge domain


**Appendix**


Feminist Data Studies and the emergence of a new Data Feminist knowledge domain


Feminist Data Studies and the emergence of a new Data Feminist knowledge domain


---

**Editorial history**

Received 30 September 2021; revised 11 January 2022; accepted 21 June 2022.

---

This paper is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Feminist Data Studies and the emergence of a new Data Feminist knowledge domain by Hana Marčetić and Jan Nolin.

*First Monday*, volume 27, number 7 (July 2022).
doi: [https://dx.doi.org/10.5210/fm.v27i7.12295](https://dx.doi.org/10.5210/fm.v27i7.12295)