Embracing Diversity and Inclusion: A Decolonial Approach to Urban Computing

Genoveva Vargas-Solar, Chirine Ghedira-Guégan, Javier Espinosa-Oviedo, José-Luis Zechinelli-Martin

Embracing Diversity and Inclusion: A Decolonial Approach to Urban Computing

Position Paper

Genoveva Vargas-Solar[*], Chirine Ghedira-Guégan[*], Javier A. Espinosa-Oviedo[*†] and José-Luis Zechinelli-Martini[‡]

[*]CNRS, Univ. Lyon, INSA Lyon, CPE, UCBL, LIRIS, UMR5205, Lyon, France
genoveva.vargas-solar@cnrs.fr, chirine.ghedira-guegan@univ-lyon3.fr, javier.espinosa@liris.cnrs.fr
[‡]Fundación Universidad de las Américas Puebla, San Andrés Cholula, Mexico
joseluis.zechinelli@udlap.mx

***Abstract*—This vision paper underscores the technical challenges and difficulties of addressing urban computing from a Diversity and Inclusion (DEI) decolonial standpoint. Issues of DEI, which encompass factors such as gender, race, age, socio-economic status, physical abilities, and religion, necessitate a shift in how we conceptualise the design of scientific and engineering methodologies. The decolonial perspective can be employed to scrutinise the influence of dominant viewpoints on our understanding of progress, innovation, and their contribution to societal welfare. The hypothesis is that these perspectives can be integrated into the design processes that deal with solutions, knowledge, and information systems across all layers, from infrastructure to application and user interfaces. This paper exhibits research questions, challenges and possible strategies to design DEI aware urban computing solutions with a decolonial perspective.**

I. Introduction

Urban and transport planning fields face the challenge of integrating infrastructure provision, sustainability, and mobility for all. However, the intersection of sustainability and gender-fair spatial development, particularly regarding mobility, has been largely overlooked and undervalued in research and policy-making at all levels. This situation highlights the need for greater attention to these issues to create more equitable and sustainable urban environments. Gendering involves assigning traditional roles and normalising different genders' behaviours, routines, and patterns. The differing preferences of men and women accompany this. Gendering can influence many aspects of society and shape how individuals interact with the world around them. Personal security and sexual harassment are highly gendered issues that do not affect the daily mobility of men to the same extent as women. These issues can significantly impact women's everyday lives and mobility, highlighting the need for greater attention to addressing them.

Diversity and Inclusion (DEI) efforts aim to reevaluate the methodologies and design principles employed in creating technology and fostering innovation. The fundamental principle of DEI initiatives is to establish a set of guiding properties or categories that inform the strategies for data collection, preservation, and utilisation, as well as the technical conditions under which these operations are conducted and their results generated, preserved, and disseminated. These categories encompass concepts such as fairness, justice, feminism, absence of bias or ageism, decolonialism, sustainability, etc.

Existing techniques that strive to promote fairness in data analysis, technology, and innovation focus on identifying and reducing bias to ensure diverse groups characterised by various attribute values are adequately represented within data collections. For instance, when examining the mobility patterns of a group of individuals, these techniques guarantee that people of all ages and genders are represented in the data collection.

However, the focus extends beyond just the data. Ensuring equitable data handling throughout a data analysis pipeline's collection, processing, and analysis stages is crucial when implementing analytics processes. To invite all the women to the royal ball, we must also ensure that every Cinderella in the village has been considered and receives an invitation.

Current techniques strongly emphasise data collection and preparation stages, ensuring that these processes do not exclude certain underrepresented groups. For instance, if individuals with diverse racial identities over 50 are not adequately represented in a dataset, it is crucial to ensure that they are not eliminated during data cleaning. Furthermore, during the data fragmentation stages, it is essential to ensure a fair selection of data samples that will be used for model training. This can be achieved by considering that underrepresented individuals are included in the sample that will be used to train a model.

Existing techniques typically begin with selecting attributes that need protection and then employ methods to ensure fairness in the data produced during the analysis stages. A significant limitation of this approach is that it usually addresses a single protected attribute at a time. However, when examining diversity, equity, and inclusion issues, an intersectional approach often provides the most comprehensive perspective for addressing equity. This underscores our motivation to propose strategies incorporating intersectionality to tackle the problems and limitations of fairness urban computing.

This paper focuses on the challenge of organising urban spaces and cities' services provision based on data to better exploit a town's resources and offer people higher-level services. By adopting a gender-aware approach, urban computing can create more inclusive, equitable, and safe urban environments for everyone. Questions to address include how people occupy the urban spaces spatially and temporally. How do perspectives of quality of life change according to citizens' gender? How do citizens are involved or excluded from urban policies? To what extent do policies consider the needs of citizens with different experiences in how they evolve within urban spaces?

Accordingly, the remainder of the paper is organised as follows. Section II states the problem of designing and implementing DEI-aware data-driven experiments in service-based architectures. Section III enumerates and discusses the main scientific and technical challenges to address the problem organised temporally and shows the main associated implications. Section IV gives an overview of related work concerning approaches and results associated with the ambition of making data-driven service-based solutions DEI aware. Finally, Section V concludes the paper and discusses future work.

II. Towards DEI aware urban computing: problem statement

Urban computing is related to sensing the cities' status, processing harvested data and acting in new intelligent ways at different levels: people, government, cars, transport, communications, energy, buildings, neighbourhoods, resource storage, etc. The main questions to consider for integrating DEI perspectives into data-driven solutions in urbanism revolve around (1) converting qualitative DEI categories into prerequisites, invariant properties and post-constraints that can guide algorithms, data management, and processing, (2) assessing and enforcing (decolonial[1]) inclusive and fair urban computing solutions that consider and give active roles to citizens to configure urban spaces according to their expectations and requirements.

The following lines underscore the critical problems that need to be addressed. These problems are identified with the understanding that there is a need to transition from a form of technological colonialism [1], where humans are left out of the urban information systems' design process, to a decolonial approach. This approach should consider data analytics and management tasks deployed on architectures, the deployment and provision strategies of infrastructure services, and the functional properties of systems. The objective is to determine how to convert human requirements and profiles into systems potentially contributing to social good.

[1]Decolonial approaches ensure that technology (data, systems, computing resources) does not perpetuate historical injustices and inequalities but contributes to a more equitable and just society. This approach critically examines how colonial power structures continue to produce inequalities today and the changes we can make to address those inequalities https://data.org/news/decolonizing-data-for-development/.

a) DEI categories as non functional properties: The challenge of modelling qualitative properties as variables and then quantitative metrics that can be observed, computed and measured is high concerning DEI categories. The three straightforward ones are:

- Diversity is about recognising and valuing the variety of unique perspectives and skills that diverse individuals bring. It represents different identities and differences (race, ethnicity, gender, sexual orientation, socio-economic status, age, physical abilities, religious beliefs, political beliefs, etc.) within a defined setting (like an organisation or community).
- Equity involves ensuring fairness within procedures, processes, and distribution of resources. It is not just about treating everyone the same way but acknowledging that advantages and barriers exist. Equity is the approach to ensure that everyone has access to the same opportunities.
- Inclusion is about individuals feeling a sense of belonging and being valued for who they are and what they bring to the table. An inclusive environment ensures equitable access to resources and opportunities for all. It also involves removing barriers to participation and contribution.

By treating these categories as non-functional properties [2], we can ensure that the system is technically robust and socially accountable. Studies and tools in the academic world address these properties, assuming that they can be converted into bias indices that can be statistically quantified within data sets. These methodologies embrace the concept of fairness [3], understanding that low-bias indices indicate a fair distribution of data samples representing diverse groups [4], [5].

b) Data harvesting with decolonial and DEI perspectives: Human mobility data enables the study of social and community dynamics based on different data sources like traffic, commuting media, mobile devices and geotagged social media data. Gathering data about the urban environment can help improve the quality of life for people affected by cities by applying greedy algorithms to collected data and complex structures. Data harvesting techniques can unobtrusively and continually collect data on a citywide scale. Data harvesting is a nontrivial problem given the three aspects to consider: (i) energy consumption and privacy, (ii) loose-controlled and nonuniform distributed sensors, (iii) unstructured, implicit, and noise data. Data harvesting is done using different data collections: (i) the continuously harvested observations of the geographical position of individuals (that accept sharing their position) along time; (ii) the images stemming from cameras observing specific "critical" urban areas, like terminals, airports, public places and government offices; (iii) data produced by social networks and applications like Twitter, Facebook, Waze and similar.

Yixian Zheng et al. [6] identify six data types that can be harvested and represent the entities observed within urban territories according to the urban context. They refer to human mobility, social networks, geographical, environmental, health care and diversity. Figure 1 summarises the urban data types considered in urban computing: environmental monitoring data concerning meteorological data and mobile phone signals for identifying behaviours. Citywide human mobility and commuting data for detecting urban anomalies. City's functional regions and urban planning; geographical data concerning points of interest (POI), land use; traffic data; social networks data; energy data obtained from sensors; and economies regarding city economic dynamics like transaction records of credit cards, stock prices, housing prices and people's income.

Data harvesting with a decolonial and DEI perspective requires thoroughly reviewing how data is gathered, stored, and utilised. It is about ensuring that data practices do not perpetuate or amplify existing inequalities but contribute to a more equitable and inclusive society. For instance, data harvesting strategies need to consider and inform people that urban spaces are under observation, and they must be transparent about how data is anonymised to protect people's privacy. Suppose data is collected through other qualitative tools like interviews or participatory data collection. In that case, individuals must explicitly know which part of their data they are willing to share and how it will be used to answer specific analytical questions. They should also be informed about the conditions under which their data is stored, for how long, and how these guarantees are enforced. Lastly, the aspect of gender is crucial in DEI perspectives as it should inform the data harvesting strategies that are adopted. For instance, in public spaces, female individuals may feel uncomfortable with data collection. Mobility patterns, for example, could pose a problem as they could be exploited to stalk women in abusive relationships.

c) Data analytics with decolonial and DEI perspectives: Several challenges related to DEI-aware data analytics include trust and transparency regarding collecting and sharing diverse data [2]. Knowing how to communicate the results of the data analysis in a meaningful and impactful way using convenient categories and vocabulary [8].

The increasing use of Artificial Intelligence (AI) models in data analytics necessitates a careful approach to the handling, sharing, and processing raw data. This is crucial to avoid compromising privacy and the integrity of data, which should not be exploited indiscriminately. Federated learning techniques offer a viable solution for adhering to these principles. The idea is to train an algorithm across multiple independent nodes, each developing local models based on their data samples. This approach empowers the independent nodes by giving them control over the data that fuels the algorithms. These nodes then share minimal data with a coordinating node, assimilating the results to construct a comprehensive model. Selecting the servers where the models and data will be stored (location of data and replicas, duration) and executing and sharing this information with data owners can empower them to regain control. These empowerment objectives allow them to negotiate what information they are willing to share, contributing to producing and utilising specific knowledge and conclusions. Data owners can have a word on the conditions in which their data are exploited, specifying energy consumption criteria and sobriety of analytics processes. Allowing the human in the loop can reduce the possible extractivist conditions in which data analytics is performed.

How can the architecture of systems, data management strategies (including storage, fragmentation, replication, distribution), and resource allocation in target architectures be made aware of DEI? When it comes to data management strategies, one could consider the selection of server types and locations for data storage, the permanence of the data, access control, the trustworthiness of data stores, and the conditions under which data can be used and shared. Resource allocation could also prioritise workloads based on their origin and non-functional requirements like choosing servers according to their location or transparency of use or event reduced energy consumption guarantees. This would require jobs to be labelled with information about their origin, the type of project/analysis they are related to, ownership, and requirements.

d) Discussion: Our vision is to consider DEI across all the stack levels, considering the data flowing along these levels and the models produced and used to process data, thus reducing intersectional bias. Generally speaking, the research questions that can be addressed concern:

- Estimating measures that would make it possible to define an index of intersectional equity, i.e. about several attributes.
- Measuring the intersectional index on the services deployed on the layers.
- Determining the degree of compliance of the services and the whole stack with the equity index.
- Ensuring that the deployment choices, the allocation of work packages and the resources allocated do not impact the verification and respect of the fairness constraints that the service stack wishes to ensure.

DEI perspectives related to systems consider:

- Understanding implicit bias: The system should be designed in a way that it understands and mitigates implicit biases.
- Micro-aggressive content detection: The system should be capable of identifying and addressing microaggressions (gender, race, nationality, religion, socio-economic level) contained in data collections and integrated in (AI) models.
- Cultural competency: The system should be culturally competent, i.e., it should be designed keeping in mind the cultural diversity of its users.
- Social Justice Development: The system should contribute to social justice development.

We suggest reevaluating the functional aspects of urban computing solutions with new protocols to ensure DEI awareness in data harvesting, processing, and analytics processes. These protocols should comply with the CARE principles (Collective

[2]The term 'data colonialism' [1] help understand the role of data and technology in entrenching these inequalities. The goal is to challenge structures of oppression and prioritise the historical contextualisation and anti-racist critique of how statistics amplify existing micro and macro power relations.

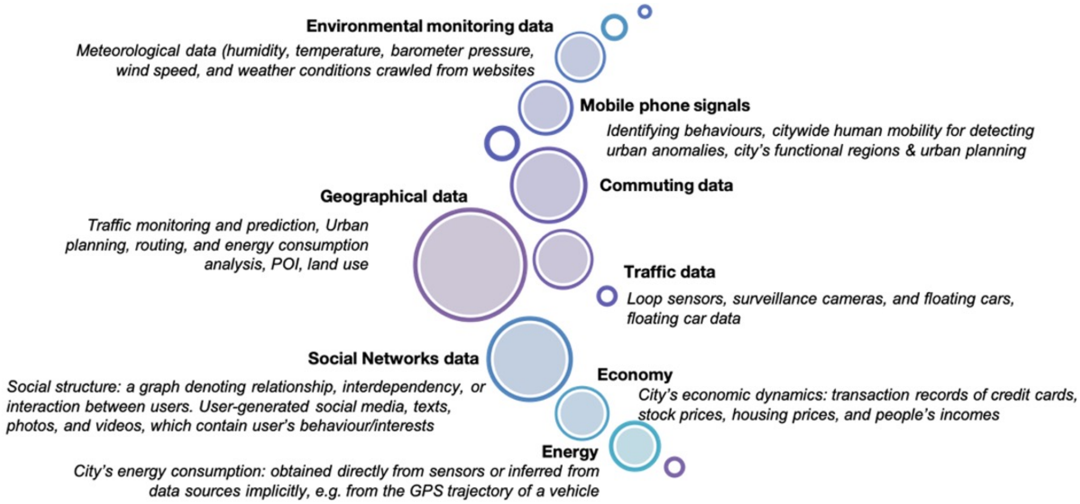

Fig. 1. An overview of urban data [7]

Benefit, Authority to Control, Responsibility, Ethics) [3], and transform human and social policy requirements into constraints and rule-based methods that can make urban computing DEI-aware.

III. Towards DEI aware Urban Computing

We assume that urban computing is a discipline driven by data and data science, which depends on highly distributed execution environments. These environments consist of infrastructure, platform, and software services with varying capacities and are deployed across extensive geographic locations. For both aspects, we associate non-functional features to data, management protocols, algorithms and analytics pipelines, deployment, and resource allocation strategies. We also add humans in the loop of decision-making and assessment to consider people's expectations for adapting and making urban computing solutions DEI aware.

Figure 2 shows a DEI urban computing architecture's general architecture. We identify the following research challenges associated with its layers, modules and functions.

a) DEI as an intersectional combination of features: DEI categories can be conceptualised as a collection of attributes (or categories) expressed as quantitative measures. These measures can be integrated to create DEI indices to evaluate inclusivity and diversity in data, processing methods, analytics algorithms, and resource allocation protocols. Depending on the context or user/application preferences, various attributes may be relevant and weighted accordingly.

The intersectional approach to tackling DEI can be framed as a multi-objective optimisation challenge, tailored explicitly for data undergoing analytical processes and resource allocation.

For data collection analytics, the issue pivots towards ensuring fairness. In this scenario, constraints are tied to specific data attributes that must be "shielded" to maintain a somewhat "equitable" statistical distribution throughout various stages of the data analytics process.

(RQ_1) The research question is: How can we measure bias in urban data across different stages of an analytics pipeline? The principle is to define a fairness index, determining to which extent a data collection is biased concerning the protected attribute. In the case of urban computing, this can be the socio-economic provenance of people or districts, girls in leisure areas, and working women in public transport. The index is computed before and after data processing (cleaning, engineering, fragmenting) to determine whether bias has been introduced. Existing techniques address one attribute at a time. The drawback is that if we privilege gender, are race, age or location harmed? Indeed, when the issues of equity, diversity and inclusion are studied, intersectional reasoning is the one that best enables an equitable perspective to be addressed. This motivates to propose intersectional strategies to address the problem.

CH_1 *The task is to refine methods that calculate a bias index, considering multiple protected attributes that collectively describe urban areas, public services and their inhabitants.*

b) DEI aware data analytics processes: Current methods primarily target the data collection and preparation phases, ensuring actions do not inadvertently omit specific underrepresented groups. For instance, if 50 years old members of the queer community are not present in a dataset, it is crucial to ensure they are not removed during data cleaning. Existing techniques like mean difference and disparate impact are used to compare input and output datasets of data preparation phases using fairness indexes. They should be adapted to deal with multiple variables.

Moreover, when fragmenting data, it is essential to guarantee a balanced selection of samples for model training. This means actively ensuring that underserved populations are included in the training sample set, thereby avoiding biasing the resulting models.

[3] https://www.gida-global.org/care

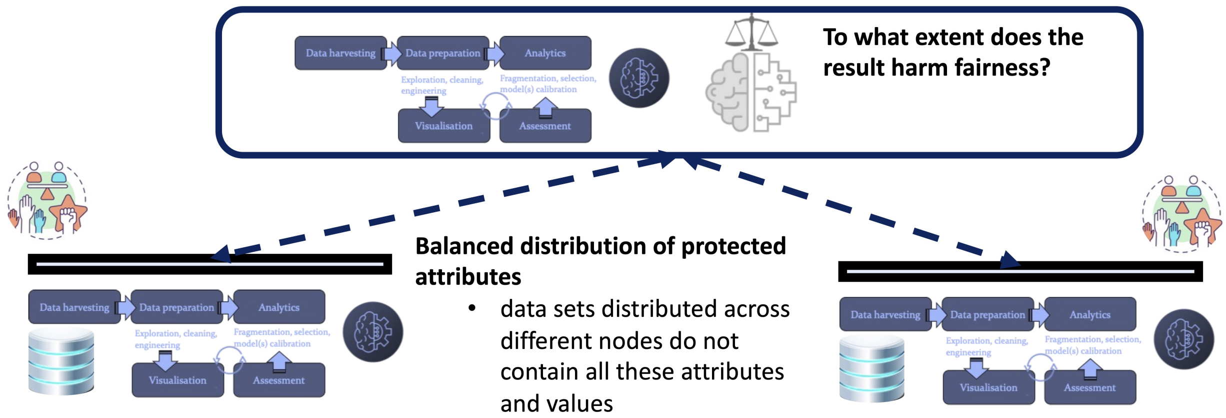

Fig. 2. DEI aware urban computing architecture

(RQ_2) The research question is: *How can we formulate a series of queries that yield samples with a specific level of fairness from one or several urban datasets with a known degree of initial bias or that can be previously transformed into non-biased datasets?*

CH_2*The challenge is thus to reevaluate query rewriting methods with the aim of (learning to) formulate a sequence of queries that ensure the creation of urban data samples that uphold intersectional fairness.*

c) DEI compliant centralised vs distributed data analytics: Data analysis through AI models increasingly requires respect for the exchange, sharing and processing of data, which risks undermining privacy and respect for data, which is not an asset that can be exploited at will.

Federated learning techniques seem to provide a solution to these rules of respect. The principle is to train an algorithm on several independent nodes by training local models on local data samples. This strategy returns control over the data that feeds the algorithms to independent nodes that exchange the minimum amount of data with a coordinator node to integrate the results and build a global model.
(RQ_3)*How to assess compliance with an expected fairness index (level of bias in data and models) in the federation?*

Given the independence of the nodes participating in a federated learning setting, each node performs a pipeline and must declare to which extent it guarantees that the data and the analytics process address a multi-objective bias level according to a specification coming from a global research question. Different nodes can manage heterogeneous data that do not necessarily have all the attributes that should be protected from bias. Therefore, the global level must be sure that aggregating the produced models' bias constraints can still be verified.

CH_3*The challenge is proposing protocols for agreement, ongoing enforcement, certification, and negotiation that ensure fairness in federated learning environments throughout the analytics process.*

IV. Related Work

Data feminism approaches [9] discuss principles and how they can be applied to improve research practices and pedagogy in geography. They explore the intersection of data and feminism, highlighting the importance of considering power dynamics and biases in data collection and analysis. The chapter [10] discusses newer approaches in gender studies that critique older ways of gathering and understanding data.

Fairness in data analytics [3], [11] involves addressing biases, ensuring equitable treatment of all participants, and developing methodologies promoting fairness at all data collection, processing, and analysis stages. It is an ongoing effort involving researchers, practitioners, policymakers, and society. Existing work addresses fairness in data analytics with approaches promoting the use of data in a way that avoids creating or reinforcing bias [12], [13]. They also address fairness and justice in data science processes through ethics of data analytics [14], and some works include feminist perspectives [15].

Fairness-aware federated learning (FAFL) [16] aims to address the fairness problem in collaborative machine learning where models trained by naive federated algorithms may be biased towards some participants and exhibit non-uniform performance across participants. Li Ju et al. [17] propose AdaFedAdam, an adaptive federated optimisation algorithm, to accelerate fair federated learning with alleviated bias. Salazar et al. [18] propose FAIR-FATE, a fairness-aware federated learning algorithm that achieves group fairness while maintaining high utility through a fairness-aware aggregation method. Ezzeldin et al. [19] propose FairFed, a fairness-aware aggregation method that enhances group fairness in federated learning while maintaining high utility. Papadaki et al. [20] propose FedMinMax, an optimisation algorithm for achieving minmax group fairness in federated learning. These approaches demonstrate improved fairness properties and outperform existing algorithms regarding fairness and convergence in federated learning. However, these approaches deal with unbalanced workload distribution and poisoning rather than considering possible biases in the data managed by nodes and how this bias contaminates the local and global models due to DEI unaware processes. Pure technical aspects in load distribution throughout nodes, interaction among a coordinator and the nodes, resulting local models and their aggregation produce biased results that can lead to unfair, unjust conclusions.

V. Conclusion and Perspectives

This position paper presents the challenges of developing new methods and philosophies to address urban computing problems from a DEI perspective. Beyond the significant technical and algorithmic challenges that already exist in urban computing, the question arises about how the social and human dimensions can be incorporated throughout the entire analytics process that often underpins urban computing solutions. With social and human elements in the loop, urban computing must ensure and reinforce a reduction in extractivist perspectives favouring fairness, explainability, participatory and transparent decision-making, and solution design. In this new perspective, people's needs should be transformed into constraints, requirements, new data management strategies, and new resource allocation protocols to ensure that urban computing benefits various groups with different requirements with an equity objective. Urban computing solutions should lead to new ways of occupying urban spaces and gender diversity-aware urban services (transportation, roads, streets, parks, parking) where people can experience a high quality of life.

References

[1] N. Couldry and U. A. Mejias, *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press, 2020.

[2] G. V. Solar and A.-S. Castillo-Camporro, "A holistic approach for measuring quality of life in "la condesa" neighbourhood in mexico city," *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, no. 39, 2020.

[3] B. Catania, G. Guerrini, and C. Accinelli, "Fairness & friends in the data science era," *AI & SOCIETY*, vol. 38, no. 2, pp. 721–731, 2023.

[4] C. Accinelli, B. Catania, and G. Guerrini, "Nondiscriminating dropout prediction beyond algorithmic fairness: Ensuring coverage in preprocessing pipelines," in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022, pp. 4642–4648.

[5] C. Accinelli, B. Catania, G. Guerrini, and S. Minisi, "A coverage-based approach to nondiscrimination-aware data transformation," *ACM Journal of Data and Information Quality*, vol. 14, no. 4, pp. 1–26, 2022.

[6] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2159–2168, 2013.

[7] G. Vargas-Solar, A.-S. Castillo-Camporro, J. Zechinelli-Martini, and J. Espinosa-Oviedo, "From data harvesting to querying for making urban territories smart," *Innovative Applications in Smart Cities, Taylor and Francis*, 2021.

[8] N. Zwiener-Collins, J. Jafri, R. Saini, and T. Poulter, "Decolonising quantitative research methods pedagogy: Teaching contemporary politics to challenge hierarchies from data," *Politics*, vol. 43, no. 1, pp. 122–138, 2023.

[9] C. D'ignazio and L. F. Klein, *Data feminism*. MIT press, 2023.

[10] M. T. Segal and V. Demos, "At the center: Feminism, social science and knowledge: An overview of this volume," *At the Center: Feminism, Social Science and Knowledge*, pp. xvii–xxv, 2015.

[11] R. Madduri, K. Chard, M. D'Arcy, S. C. Jung, A. Rodriguez, D. Sulakhe, E. Deutsch, C. Funk, B. Heavner, M. Richards *et al.*, "Reproducible big data science: A case study in continuous fairness," *PloS one*, vol. 14, no. 4, p. e0213013, 2019.

[12] M. Drosou, H. V. Jagadish, E. Pitoura, and J. Stoyanovich, "Diversity in big data: A review," *Big data*, vol. 5, no. 2, pp. 73–84, 2017.

[13] E. Pitoura, K. Stefanidis, and G. Koutrika, "Fairness in rankings and recommendations: an overview," *The VLDB Journal*, pp. 1–28, 2022.

[14] K. Martin, "Fairness and justice in data analytics," in *Ethics of Data and Analytics*. Auerbach Publications, 2022, pp. 249–253.

[15] T. C. Pargman, C. McGrath, O. Viberg, and S. Knight, "New vistas on responsible learning analytics: A data feminist perspective," *Journal of Learning Analytics*, vol. 10, no. 1, pp. 133–148, 2023.

[16] Y. Shi, H. Yu, and C. Leung, "Towards fairness-aware federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[17] L. Ju, T. Zhang, S. Toor, and A. Hellander, "Accelerating fair federated learning: Adaptive federated adam," *arXiv preprint arXiv:2301.09357*, 2023.

[18] T. Salazar, M. Fernandes, H. Araújo, and P. H. Abreu, "Fair-fate: Fair federated learning with momentum," in *International Conference on Computational Science*. Springer, 2023, pp. 524–538.

[19] Y. H. Ezzeldin, S. Yan, C. He, E. Ferrara, and A. S. Avestimehr, "Fairfed: Enabling group fairness in federated learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, 2023, pp. 7494–7502.

[20] A. Papadaki, N. Martinez, M. Bertran, G. Sapiro, and M. Rodrigues, "Minimax demographic group fairness in federated learning," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 142–159.