ON PURSUING A CRITICAL REASONING OF THE SCIENCE OF CITIES AND THEIR USE OF BIG DATA AND AI

by Ana María Jaramillo & Nandini Iyer

ijurr

The following essay reflects our perspective as two researchers: one who studies the social practices of scientists conducting research in Science, and the other who examines housing insecurity and structural segregation in cities. Both of our works are grounded in theories from the sociology of science and employ computational methods. With this essay, **we seek to raise questions about the scientific practices of researchers studying cities using Big Data and AI, rather than concentrating solely on the content of their scientific output**. We will present our opinions, based on factual evidence, regarding the ethical concerns that arise when studying cities with Big Data and AI in the absence of established ethical frameworks. Specifically, we examine three interconnected problems: the uneven geographic distribution of urban research, which creates epistemological blind spots; the methodological invisibility in big data approaches; and the lack of interdisciplinary collaboration, which can lead to ineffective policy recommendations.

The nuances of urban sustainability

As outlined in Sustainable Development Goal (SDG) 11, which focuses on "Sustainable Cities and Communities," the target is to ensure that by 2030, everyone has access to adequate, safe, and affordable housing, as well as transportation and participatory planning for human settlements. This goal also highlights the importance of reducing the environmental impact of cities and creating resilient community spaces that can withstand natural disasters ("Goal 11 |

Department of Economic and Social Affairs", n/d). In response to this pressing need, **the study of cities as a complex research object has grown significantly, even giving rise to the concept of a "Science of Cities"** (Batty, 2013). This emerging discipline has attracted many pSTEM researchers, particularly physicists specializing in complex systems and computer scientists, who are drawn to its interdisciplinary framework.

However, the Science of Cities is often approached from narrow disciplinary perspectives, without grappling with the fundamental tensions that exist within and across SDGs. As urban researchers, we must confront difficult questions: When improving transportation accessibility conflicts with reducing emissions, which takes priority? When housing affordability necessitates urban expansion that compromises environmental preservation, how do we make decisions? These require extending beyond addressing technical problems and considering ethical frameworks for assessing the impacts and trade-offs of different policy scenarios. Yet, much of current urban science overlooks these stark contradictions, focusing on optimizing individual metrics rather than developing the analytical frameworks to help decision-makers navigate these complex sustainability dilemmas.

Urban big data

In the age of Big Data and AI, we now have the computing resources to analyze vast datasets involving millions of people at rapid speeds—typically within hours or a few days—something that was previously unimaginable.

While most of these studies adhere to the scientific method with rigor, in this essay, we want to highlight some ethical implications that need to be addressed within this new Science of Cities. These concerns range from **the lack of positionality among researchers and potential colonial practices in data extraction to inadequate critical interpretations of results and implications**, particularly in studies related to segregation and migration.

Cities are generally hotspots for both internal migration within countries and international migration, particularly in major metropolitan centers that serve as hubs for global economic and cultural exchange (Sassen, 2013). It is estimated that by 2030, 84% of the population in high-income countries will live in urban areas, while in middle- and low-income countries, those percentages are projected to be 59% and 38%, respectively (United Nations, Department of Economic and Social Affairs, 2020). As cities continue to expand both horizontally and vertically (Frolking et al., 2024), the challenges for government agencies and citizens grow in tandem with the increasing collection of personal data. This data encompasses a range of information, including census details, residential locations, health records, environmental features, crime statistics, and mobility patterns tracked through smartphones, social media, and camera surveillance.

The rising volume of data presents opportunities for data-driven interventions in urban environments. Still, it also raises similar issues encountered by AI practitioners worldwide: the absence of ethical, legal, and economic frameworks for effectively using data and applying computational techniques. In Europe, the AI Act, which is set to be implemented in phases starting in 2024, aims to address these concerns. However, its protections extend only to data concerning European citizens, leaving international residents of the continent—such as expats, immigrants, and refugees—as well as tourists, unprotected by the AI Act (Madiega et al., 2019).

The ethical stakes of urban data

The ethical concerns of urban data collection become clear when we conceptualize cities as data extraction mines, where individuals' consent is often overlooked as their information is collected in government data centers or cloud servers (Crawford, 2021). While this data collection is typically justified by the potential benefits for citizens' well-being through data-driven urban interventions, **this information can also be seen as a form of capital that can be extracted, accumulated, and exploited** (Sadowski, 2019), potentially serving as a means of control over civilians (Pasquinelli, 2023).

Researchers in the Science of Cities must be vigilant against misleading uses of individual data, which has previously led to individual persecution and killings in the pre-internet era (Véliz, 2020). Today, we face alarming instances of data misuse in the AI era, including: i) profiling marginalized communities as part of criminal groups based on biased data rather than real crimes (Hutson, 2011), ii) using facial recognition technology alongside personal information to target immigrants for arrest or deportation (Joseph, n/d), iii)

restricting the democratic right to protest by anticipating demonstrations and making pre-emptive arrests (Booth et al., 2011), and iv) employing facial recognition algorithms trained on images of inmates captured without their consent, which are then used in drone warfare against human targets (O'Neil, 2016), including the AI-assisted genocide in Palestine (Al Jazeera, n/d).

Three critical problems in the Science of Cities

GEOGRAPHICAL INEQUALITY — Returning to the topic of scientific practices: Science is one of the most unevenly distributed resources worldwide (Xie, 2014), with a small number of countries and institutions holding the majority of research resources. This issue is also evident in the Science of Cities; "global cities" concentrate numerous scientific institutions and have access to vast amounts of individual data. This data enables an exploration of the complexities within these cities and helps policymakers understand urban dynamics and design effective interventions.

However, a significant disparity exists between cities worldwide. A brief literature search in Open Alex for the global cities of London, Tokyo, and New York, alongside the term "City," reveals a total of 130,000 scientific publications. In contrast, other cities—such as Berlin, Bogotá, Nairobi, Tehran, and Beijing—each have fewer than 5,000 scientific publications. The different attention, in terms of scientific publications, given to the cities is not necessarily related to their size, but also to the resources of those cities, the number of universities, and

available datasets. While focusing on these cities can offer interesting insights, it could also lead to overlooking other regions that have more pressing issues that need to be addressed.

From our perspective, these dynamics reveal three main issues in relation Science of Cities. First, there is an epistemological stagnation due to the extensive focus on specific cities, driven by the availability of public and private datasets. This results in the overuse of the same information and the repeated answering of similar questions. Second, some cities are understudied because of the lack of available public and private data (Hagen-Zanker et al., 2023). This often leads to more severe problems remaining unaddressed, particularly since many of these cities are located in countries that have not reached the United Nations' Sustainable Development Goals (SDGs) for 2030 (Sustainable Development Solutions Network, 2024). Third, for those cities or countries that are understudied but do have available public datasets, researchers affiliated with institutions in high-income countries may fall into the "white savior" trap. They might study complex phenomena in these cities without collaborating with local researchers, governments, or the communities from which the data was sourced (Das, 2024).

METHODOLOGICAL INVISIBILITY — A key dimension of criticality in the Science of Cities lies in how findings are interpreted in relation to surrounding populations. Much of the field depends on big data (remote sensing, anonymised mobility trajectories, modeled indicators), onto which demographic attributes are often imputed. While these approaches enable comparative, large-scale analyses, they only reflect specific urban experiences,

highlighting gaps in planning, infrastructure, housing, or service provision. Equally significant are the populations who remain invisible in such data because their practices, vulnerabilities, or needs are less detectable, irregular, or systematically excluded. These gaps in representation underscore the need for mixed-methods approaches that combine mobility trajectories with travel diaries, surveys, and other qualitative data sources. Although such methods come with their own biases, such as scalability, they provide perspectives that are otherwise excluded, offering a more comprehensive assessment of how urban systems are experienced across different populations.

LACK OF INTERDISCIPLINARY COLLABORATION – As our final point, we want to address the limited number of interdisciplinary and mixed-methods studies in some research related to the Science of Cities. While interdisciplinary research is not strictly required, it is particularly important in applied subfields like the Science of Cities, where i) the definition of "cities" is a legal term, ii) citizens and non-citizens coexist based on their legal status, and iii) individuals are categorized by geographical metrics such as socio-demographics, which often leads to their segregation within different urban areas (Useche et al., 2024). Studies using Big Data and computational methods to analyze these individuals must be grounded in the theoretical frameworks of the various fields that contribute to the understanding of cities. This is especially critical when the Science of Cities aims to influence the design of interventions that can significantly impact the well-being of citizens.

In most cases, computer scientists do not receive training in critical thinking, reasoning, or ethics during their education. Consequently, when studies are not co-designed and results are not co-interpreted with experts from fields such as law, history, political science, sociology, or geography, as well as with the communities impacted, the research can lack validity and, more importantly, pose potential harm to those involved in the study—including individuals represented in secondary or public datasets (Noble, 2018).

Specifically concerning the Science of Cities, while research findings may aim to inform public policy, if they are not approached carefully and critically—taking into account historical data and insights drawn from previous qualitative evidence—they can inadvertently reinforce stereotypes about vulnerable and marginalized populations, such as those from low-income backgrounds, migrant communities, or BIPOC groups (Pasquinelli, 2023; Crawford, 2021; Hutson, 2021; Noble, 2018; O'Neil, 2016). As a result, these communities could suffer harm as a secondary effect of the research.

Toward more critical approaches

These methodological concerns point toward alternative approaches. Alongside methodological innovation, community-led approaches offer another pathway toward critical Science of cities. Residents may not possess the technical expertise to analyze extensive data or the longitudinal memory to track urban trends. Still, they often have an attuned sensitivity to everyday changes in their surroundings (shifts in traffic, housing pressure, access to public space, etc.) Integrating these experiential insights

with data science not only grounds analysis in lived realities but also challenges the top-down assumptions embedded in purely quantitative models. Participatory mapping, citizen science initiatives, and co-produced datasets can bridge the gap between big data and situated knowledge, improving the interpretation of urban data science and the design of more equitable interventions.

In conclusion, while this essay may raise controversy among researchers in the Science of Cities, **it also has the potential to spark doubt and inspire concrete actions that promote more interdisciplinary and community-based scientific practices**. As a first step toward more ethical research, we advise scholars in the Science of Cities to provide statements of positionality regarding their work: clarifying where, why, and by whom their studies are conducted, as well as increasing efforts for capacity building in the cities they study. For example, they could follow the model of projects like LAC-Urban Health and SALURBAL (Diez Roux et al., 2019), where interdisciplinary teams of researchers collaborated with civil society to evaluate health and various other dimensions in cities across Latin America, with implications for policymaking.

* **Opinions written in the text are of our own responsibility and do not reflect the opinion of our institutions.**

References

Al Jazeera. (2024, April 4). 'AI-assisted genocide': Israel reportedly used a database for Gaza kill lists. Al Jazeera. https://www.aljazeera.com/news/2024/4/4/ai-assisted-genocide-israel-reportedly-used-database-for-gaza-kill-lists

A. V. Diez Roux, S. C. Slesinski, M. Alazraqui, W. T. Caiaffa, P. Frenz, R. J. Fuchs, J. J. Miranda, D. A. Rodriguez, O. L. S. Dueñas, J. Siri, A. V. Vergara, Global Challenges 2019, 3, 1800013. https://doi.org/10.1002/gch2.201800013

Batty, M. (2013). The new science of cities. The MIT Press. https://doi.org/10.7551/mitpress/9399.001.0001

Booth, R., Laville, S., & Malik, S. (2018, May 22). Royal wedding: police criticised for pre-emptive strikes against protesters. The Guardian. https://www.theguardian.com/uk/2011/apr/29/royal-wedding-police-criticised-protesters

Crawford, K. (2021). Chapter 3: Data. In The Atlas of AI. Yale University Press. https://doi.org/10.2307/j.ctv1ghv45t

Das, D. (2024). Identity Decolonization Amid the Coloniality of Computing [PhD Dissertation, University of Colorado at Boulder]. https://core.ac.uk/download/635937582.pdf

Data company directly powers immigration raids in workplace | WNYC | New York Public Radio, podcasts, live streaming radio, news. (n.d.). WNYC. https://www.wnyc.org/story/palantir-directly-powers-ice-workplace-raids-emails-show/

Frolking, S., Mahtta, R., Milliman, T. et al. Global urban structural growth shows a profound shift from spreading out to building up. Nat Cities 1, 555–566 (2024). https://doi.org/10.1038/s44284-024-00100-1

Goal 11 | Department of Economic and Social Affairs. (n.d.). Department of Economic and Social Affairs, Sustainable Development. Retrieved August 19, 2025, from https://sdgs.un.org/goals/goal11#targets_and_indicators

Hagen-Zanker, J., Agyemang, F., Mallah, H. B., Memon, R., Paller, J., Wolf, L. J., & Fox, S. (2023, July 18). Practical reflections on conducting survey research in cities in low and middle-income countries. ODI: Think Change. https://odi.org/en/publications/practical-reflections-on-conducting-survey-research-in-cities-in-low-and-middle-income-countries/

Hutson, M. (2021, March 8). Artificial intelligence could identify gang crimes—and ignite an ethical firestorm. Science | AAAS. https://www.science.org/content/article/artificial-intelligence-could-identify-gang-crimes-and-ignite-ethical-firestorm

Madiega, T. & European Parliamentary Research Service. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. In EPRS | European Parliamentary Research Service [Report]. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

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

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, USA.

Pasquinelli, M. (2023). The eye of the master: A social history of artificial intelligence. Verso Books.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. Big Data & Society, 6(1). https://doi.org/10.1177/2053951718820549

Sassen, S. (2001). The Global City: New York, London, Tokyo (REV-Revised). Princeton University Press. https://doi.org/10.2307/j.ctt2jc93q

Sustainable Development Report 2024. (2024, June 17). Sustainable Development Solutions Network. https://www.unsdsn.org/resources/the-sustainable-development-report-2024/

United Nations, Department of Economic and Social Affairs, Population Division. (2020). Population Facts, No. 2020/2. Population Facts, 1–2. www.un.org.development.desa.pd/files/undes_pd_2020_popfacts_urbanization_policies.pdf

Useche, A. F., Sarmiento, O. L., Álvarez-Rivadulla, M. J., Medina, P., Higuera-Mendieta, D., & Montes, F. (2024). Spatial segregation patterns and association with built environment features in Colombian cities. Cities, 152, 105217. https://doi.org/10.1016/j.cities.2024.105217

Véliz, C. (2021). Privacy is power. Brooklyn: Melville House.

Xie, Y. (2014). "Undemocracy": inequalities in science. Science, 344(6186), 809–810. https://doi.org/10.1126/science.1252743

About Ana María Jaramillo

Ana has been a postdoc at the Complexity Science Hub since June 2023 and works mainly for the project on Multi-attribute, Multimodal Bias Mitigation in AI Systems (MAMMOth).

Ana finished here PhD in Computer Science at the University of Exeter. During her doctoral training she studied the effect of segregation and diversity on scientific impact in Computer Science, applying sociology of science theories with computer science methodologies.

Webpage:

https://csh.ac.at/ana-maria-jaramillo/

About Nandini Iyer

Nandini Iyer is a postdoctoral research assistant at the Complex Connections Lab, Northeastern University London. She holds a BSc in Computer Science from the University of Illinois at Urbana-Champaign and completed her PhD at the University of Exeter's BioComplex Lab. Her research explores the intersection of socioeconomic inequality, human mobility, and public transport, focusing on transport poverty and data-driven analysis of socioeconomic disadvantage.

Webpage:

https://www.networkscienceinstitute.org/people/nandini-iyer