Journal Pre-proofs

Perspective

The science of urban informatics

Wenzhong Shi, Anshu Zhang, Fan Shi, Michael Batty, Michael F. Goodchild

PII:	S2095-9273(25)01301-5
DOI:	https://doi.org/10.1016/j.scib.2025.12.036
Reference:	SCIB 3761

To appear in: *Science Bulletin*

Please cite this article as: W. Shi, A. Zhang, F. Shi, M. Batty, M.F. Goodchild, The science of urban informatics, *Science Bulletin* (2025), doi: https://doi.org/10.1016/j.scib.2025.12.036

The science of urban informatics

Wenzhong Shi[a, *], Anshu Zhang[a, *], Fan Shi[a], Michael Batty[b], Michael F. Goodchild[c, *]

[a] Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong 999077, China

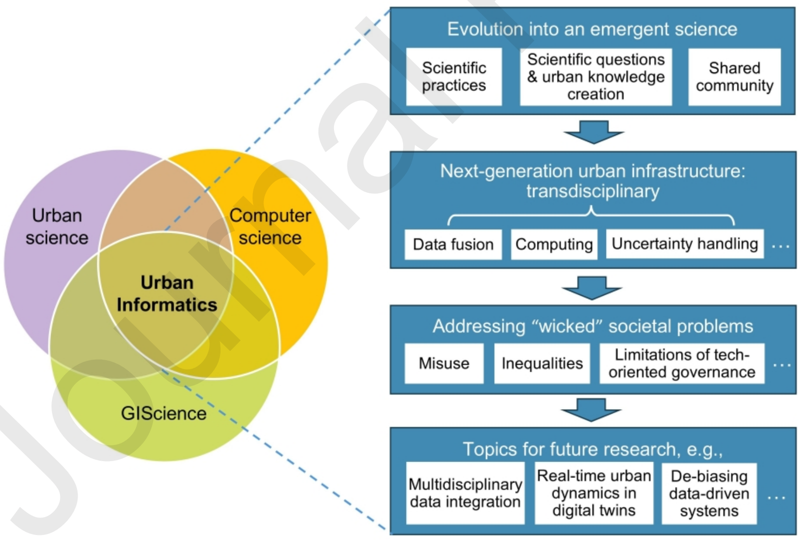
[b] Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, UK

[c] Department of Geography, University of California, Santa Barbara, CA 93106, USA

*Corresponding authors.

Email addresses: john.wz.shi@polyu.edu.hk (W. Shi), april-anshu.zhang@polyu.edu.hk (A. Zhang), good@geog.ucsb.edu (M.F. Goodchild).

Graphical abstract



As global urbanization intensifies, its challenges are becoming increasingly complex and pressing. These issues encompass problems manifest in the form and function of cities, such as traffic congestion and housing–employment imbalances, alongside escalating concerns about sustainable development and social challenges ranging from equity, inclusion, and

affordability to the unintended consequences of new technologies. The rapid spread of smartphones and Apps in the early 21st century has created a virtual electronic skin that transforms industries, land uses, mobilities, and infrastructures, which are accelerating cities to become more digitally integrated, hence "computable" [1]. Data and tools to tackle these problems are increasing, and many view technologies driven by artificial intelligence (AI) as heralding a new era of productivity and economic growth. These innovations broaden citizens' access to information and encourage public engagement, while city governments are also able to utilize such data to influence citizens' social and economic behaviors.

However, many persistent urban problems are not solvable by technology-centric interventions alone; moreover, technologies both mitigate and generate new urban challenges through unanticipated consequences. Technologically-driven changes, such as the decline in traditional commuting due to autonomous driving and working from home, as well as the redistribution of urban land uses resulting from online activities, further complicate new urban plans and solutions. Many of these complex challenges require an integrated approach that merges various technological domains and social perspectives to craft solutions that shape a brighter future for cities.

This article focuses on urban informatics [2], a rapidly growing response to many of these issues. Like many earlier efforts to engage and extend bodies of knowledge in specific domains, urban informatics examines its value in a systematic focus on information in all its senses that increasingly underpins our understanding and management of urban systems. For example, a need to focus attention on the information processed by geographic information systems (GISystems) led to the emergence of a new discipline of geographic information science (GIScience) or geoinformatics [3] in the early 1990s, which addressed fundamental questions about geographic information and spurred interest in using GISystems to enable discoveries in a range of domain-based disciplines. By the end of that decade, there was a general acceptance that this systematic focus on geographic information could constitute a science.

The various informatics fields (e.g., biomedical, public-health, community, and environmental) have surged since the early 21st century, where "informatics" is typically defined as the scientific study of data and information [4], information and communication technologies (ICT), the Fourth Paradigm for data-intensive scientific discovery, or data sciences within various application domains [5]. These new fields have, in turn, led to critical interrogations that examine their broader societal impacts, including the potential for privacy invasion, surveillance, fostering social inequality, and further diminishing the power of marginalized communities. A good example is the critique of GIScience that emerged in the 1990s and led to the survey edited by Pickles [6]. All arguments can and should be applied to urban informatics, but this paper does not attempt to comprehensively review them or to speculate on the form they will take. It simply points the way.

Instead, this article aims to assess urban informatics against the key hallmarks of an emerging science, as well as to delineate the science emerging in this field and the potential for furthering such a science. The next section explores the concept of convergent science [7], arguing that a deliberate breaking down of longstanding barriers between scientific domains can lead to dramatic advances, where a science of urban informatics could both capitalize on existing knowledge in several related disciplines and stimulate the creation of new knowledge. The next two sections illustrate the roles of urban informatics in future cities, exemplify recent advances and discoveries in the domain of urban informatics, and identify immediate research challenges. These sections also discuss the importance of human factors as embraced in the social sciences and how urban informatics can help address the persistent urban problems that involve various social concerns. The concluding remarks provide a further outlook on broad research questions for urban informatics.

Urban informatics as a convergent science. Convergent science is a science that integrates "insights and approaches from originally distinct fields" [7]. Convergence is vital for advancing sciences by bridging expertise and knowledge across domains with large epistemic gaps [8]. This approach creates solutions to the most complex of human challenges [7], enabling outcomes that surpass individual disciplines and avoiding the possible undesirable consequences of limited perspectives.

Urban informatics has followed a development trajectory analogous to GIScience (Fig. 1). Although its later stages for establishing a consolidated theoretical corpus are still unfolding (as we picture in Fig. 1c, d), urban informatics has evolved from technology-driven efforts emphasizing theory and data (Fig. 1a, b) toward the articulation of core scientific problems and the creation of new urban knowledge (Fig. 1c) [9,10]. The academic community sharing the urban-informatics connotation has grown alongside this shift, evidenced by many research laboratories and specialized degree programs, a dedicated journal, an academic society, and a community-wide research agenda in development (Fig. 1d). We anticipate continued, rapid expansion of this community over the next decade, mirroring the growth and institutionalization of GIScience since the 1990s. The Supplementary material to this paper provides a comprehensive review of the literature underpinning urban informatics as a science and the current state of its academic community.

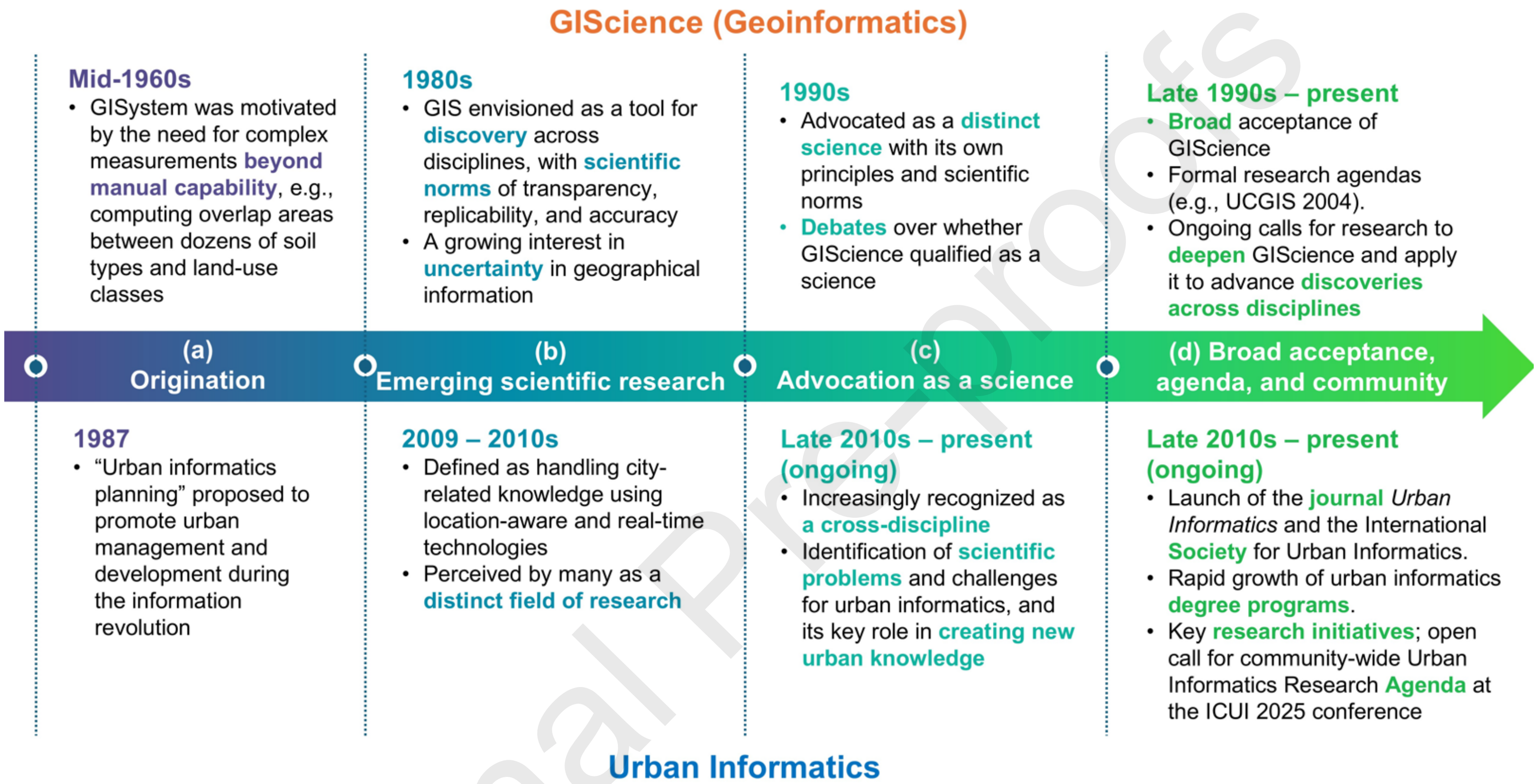


Fig. 1 The development timeline for GIScience (geoinformatics) and urban informatics. A complete version, with citations supporting each statement on the figure, is provided in Fig. S1 (online).

Our 2021 definition of urban informatics as the integration of urban science, computer science, and GIScience [2] brought it within the fold of convergent science. This integration emphasizes creating new knowledge and methods for complex urban challenges that individual fields cannot directly address, as well as deep cross-disciplinary collaboration, which are key characteristics of convergent science [7]. The term "urban" here encompasses the broader field of urban studies, rather than urban science in its narrower sense of complexity in urban morphology, social physics, and its relationships between people and the built environment [11]. For studying cities, computer science is vital to address prevalent digitalization, high-frequency urban data, and the need for advanced computational models. Meanwhile, the central role of geospatial data for cities necessitates the further integration of GIScience. The understanding and modeling of human, built environments, and urban dynamics underscore the importance of integrating urban science. These components are deeply integrated with knowledge and practice in the target domains that embrace a multitude of urban issues. This integration can not only improve the discovery, interpretation, prediction, or intervention within specific application domains like established informatics fields, but also yield generalizable solutions applicable across multiple domains with similar principles or needs.

Urban informatics as infrastructure. Digital and data infrastructures have become central components of urban infrastructure alongside traditional physical systems. Since the 1990s, public administrations in Europe and beyond have invested substantially in geographic data exchange platforms. Beyond earlier established elements focusing on data processing, transmission, management, and quality assurance, data infrastructure increasingly encompasses tools for data acquisition, modeling, analysis, skilled analysts, decision makers, and the public engaging with data products.

Urban informatics constitutes a next-generation urban infrastructure that integrates urban science, computer science, GIScience, and related domains to bridge knowledge gaps, thereby enhancing understanding of cities and solutions for urban efficiency, social outcomes, and sustainability. It will be helpful to compare this notion with those detailed in "Building Geospatial Infrastructure" [12] and connect it to more generic theories of social science.

Consequent on the definition of big data, the growing volume, velocity, spatial and temporal resolution, and heterogeneity of urban data increase difficulties in data representativeness, interoperability, decentralization, and quality control. The following interdisciplinary challenges, for example, require urban-informatics solutions:

(i) Fusing multidisciplinary data needed in addressing complex urban problems: urban datasets come with diverse structures, features, limitations, and domain-dependent interpretations. Maintaining data integrity during fusion requires knowledge from application domains, urban science, as well as data acquisition technologies in GIScience. For example, fusing datasets with different areal units commonly triggers the Modifiable Areal Unit Problem [13], that is, aggregating the same spatial measurements to different areal units can significantly alter patterns and bias analytical results.

(ii) Transdisciplinary scalable computing: real-time, low-latency applications, such as autonomous driving, require distributed architectures, efficient algorithms, and integrated solutions for security, compression, privacy, and transmission. These considerations are tied to specific data-capturing technologies in GIScience and various domain knowledge.

(iii) Transdisciplinary treatment of data quality and uncertainties: understanding and control over data quality and uncertainties become foundational science problems under the Fourth Paradigm. New data modalities and applications face the lack of mature quality standards and evaluation methods, and uncertainties from phenomena, measurement, cognition, representation, and analysis propagate through this pipeline and can severely affect decisions. Addressing these requires coordinated methods from urban science, computer science, and GIScience applied to specific urban problems (Fig. 2).

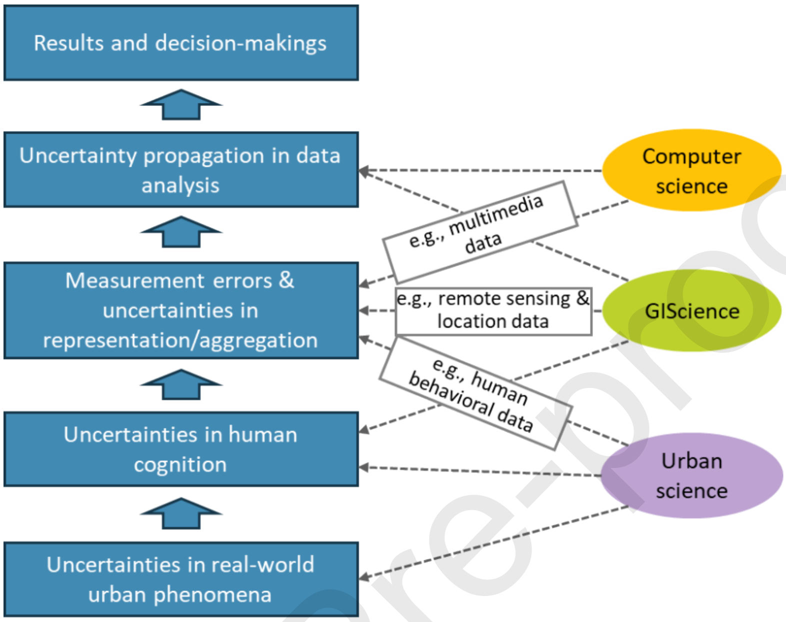


Fig. 2 The integration of knowledge, uncertainties, and methods from different disciplines and domains involved in urban informatics.

Data-driven spatial epidemiology demonstrates how urban informatics can deepen understanding of urban processes and support interventions. The 1854 Broad Street cholera investigation in London, catalogued and first interpreted by Snow [14], demonstrated that spatial analysis could greatly contribute to identifying transmission sources as well as informing mitigation measures and sanitation infrastructure. During COVID-19, studies integrated infectious-disease mechanisms with spatiotemporal modeling from GIScience, human behavior modeling from urban science, and advanced computing algorithms to clarify factors influencing epidemic spread, predict transmission patterns, and evaluate intervention strategies, making invaluable contributions to pandemic response. Meanwhile, these advances exposed persistent interdisciplinary gaps. For example, many spatial epidemiology analytics assume the independence between observations, which often does not hold due to spatial dependencies (i.e., spatial autocorrelations) in infection rates and in socioeconomic conditions over spatial administrative units. Ignoring spatial dependence can inflate statistical significance and mislead public health decisions. Addressing this requires developing novel urban-informatics methods that explicitly model spatial dependence, formulate uncertainty propagation, and provide robust confidence estimates. Important advances continue to be made in simulating how epidemic processes diffuse in the most generic of ways [15], thus providing the analytical forms that underpin various domains of urban informatics.

Urban informatics and evolving urban challenges. Many, if not most, persistent urban problems are not solvable by technologies alone, and technologies can generate new challenges concerning ethics, misuse, exacerbated inequalities, and oversimplification of urban governance. Using AI and digital twins as examples, this section explains how urban informatics, integrating scientific, technical, socioeconomic, and governance perspectives, is essential for addressing these urban issues.

As AI boosts productivity, it also threatens not only routine work but also increasingly creative and professional roles, such as medicine, education, the arts, and research processes. This results in wealth concentration among AI technology controllers; at a macroscopic scale, a minority of AI-empowered cities may accumulate political and economic power and skilled labor, while displacing millions of unemployed with lower AI literacy to “de-tech” cities that face persistent impoverishment [16]. AI systems, including large language models (LLMs), systematically inherit and perpetuate biases and discrimination from human society. Although debiasing methods through engineering data, prompts, training process, and model behaviors are under active development, they are largely pursued in resource-rich countries [17], leaving the rest of the world vulnerable to aggravated discrimination and exclusion, especially for speakers of less-resourced languages.

While the real-time monitoring and predictive modeling by digital twins can improve understanding of urban dynamics and enable more responsive policy, they do not automatically overcome the limitations of technocratic governance. Many urban issues, conflicts, disadvantaged groups, and forms of discrimination are underrepresented or “concealed” in data and analytics [18], producing biased predictions and decisions that deepen marginalization [11]. Moreover, the perceived authority of digital twins can amplify existing inequalities by privileging those with resources and expertise, while excluding others and increasing the risk of polarizing or extreme policy choices.

Urban informatics can better align technology with societal needs by decomposing urban solutions into an interdisciplinary pipeline: data capture, management, analysis, and interpretation, where each stage requires technical, social, and domain expertise and offers opportunities to assess and mitigate broader impacts. Fine-resolution earth observation data and crowdsourced geographic information are increasingly used to complement data-poor regions, notably parts of the Global South, producing less biased datasets that improve the AI training fairness and enable more effective data-driven de-biasing techniques. Effective de-biasing with these various data sources, however, requires integrating the GIScience expertise to characterize geospatial semantics, data acquisition limitations, and measurement biases; domain expertise to interpret data and results in application contexts; and social-science and policy perspectives to identify remaining biases, assess whether multi-source fusion introduces new biases, and evaluate societal acceptability. Quantitatively embedding these multidisciplinary insights into models and workflows defines a research agenda where urban informatics can substantially advance fairer AI and LLM methods.

*Concluding comments***.** Our depiction of the core of urban informatics has provided a glimpse of how this emergent field undoubtedly supports the criteria for a scientific discipline, emphasizing advances in scientific practices following the rules of repeatability, quantifiability, and transparency, with greatly successful instances such as statistics and information science. We have also illustrated how urban informatics is beginning to formulate new hypotheses and theories by integrating these advances with human factors; hence, these theories can more effectively address both persistent and emerging urban societal issues. This anticipates the eventual transition of urban informatics into the realm of an empirical science.

Subject and disciplinary domain areas are currently facing many disruptive trends, such that their subject matter is being forced to adapt and reconstruct, especially in the fields that intersect urban informatics. We are moving from well-designed, small-volume, high-quality data sources into massive but heterogeneous big data. This forces researchers to devote their attention to data handling and pattern discovery—tasks that were historically unlikely to be rated as science in themselves. New areas, including digital twins, big data, and AI, are being driven not so much by their potential for new discoveries as by their superior ability to handle new data sources.

While we might support the notion that urban information can itself be a meaningful topic of scientific discovery, an argument that was made much earlier for geographic information, we consider it unlikely that any general truths will emerge. Rather, the convergence of informatics with substantive disciplines such as urban science opens new pathways to scientific knowledge. Digital twins, for example, combine urban information with the objective of real-time simulation of complex urban environments to provide solutions to challenges of urban science. In other words, the full potential of urban information may only be realized when it is embedded in a new convergent science of urban informatics.

In this context, we suggest possible topics for future research in urban informatics, based on the essential role of information and the motivation provided by its scientific convergence:

1. Ways of organizing abundant multidisciplinary data in support of discoveries. How can data be fused, and uncertainty handled against multiple use cases?
2. What design principles underlie the user experience when multidisciplinary data are used to support advanced computational systems like AI and digital twins?
3. What forms of visualization are needed? How can uncertainty be visualized? How can scale be pictured? How should our understanding of human cognition be advanced to deal with these new forms?
4. What are the technical challenges of synchronizing dynamic digital twins with the real systems being modeled? How might the dynamic presence of traffic and pedestrians and their interactions be modeled in digital twins?
5. How should future systems like digital twins and AI be governed, owned, and managed? How do we assess and reduce biases within these systems by integrating multidisciplinary data and perspectives?

To return to the aim of this paper, we believe that this paper has clearly established the criteria for a new kind of information science and has provided overwhelming evidence in its support. We hope this paper energizes a debate where many more questions can be posed and answered.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Shenzhen Park of Hetao Shenzhen Hong Kong Science and Technology Innovation Cooperation Zone, "Theories for Spatiotemporal Intelligence and Reliable Data Analysis" (Project ID: HZQSWS-KCCYB-2024058), and the Otto Poon Charitable Foundation Smart Cities Research Institute, the Hong Kong Polytechnic University (Work Program: CD06).

References

[1] Batty M. The computable city: histories, technologies, stories, predictions. Cambridge, Massachusetts: The MIT Press; 2024.

[2] Shi W, Goodchild MF, Batty M, et al. Overall Introduction. In: Shi W, Goodchild MF, Batty M, et al, editors. Urban Informatics, Singapore: Springer; 2021, 1–7.

[3] Goodchild MF. Geographical information science. Int J Geogr Inf Syst 1992;6:31–45.

[4] Friedman CP. What informatics is and isn't. J Am Med Inform Assoc 2013;20:224–6.

[5] Frew JE, Dozier J. Environmental informatics. Annu Rev Environ Resour 2012;37:449–72.

[6] Pickles J. Ground truth: the social implications of geographic information systems. Guilford Press; 1995.

[7] Council NR. Convergence: facilitating transdisciplinary integration of life sciences, physical sciences, engineering, and beyond. National Academies Press; 2014.

[8] Petersen AM, Ahmed ME, Pavlidis I. Grand challenges and emergent modes of convergence science. Humanit Soc Sci Commun 2021;8:194.

[9] Li Q. From geomatics to urban informatics. Geomat Inf Sci of Wuhan Univ 2017;42:1–6.

[10] Liu C, Ye X, Yuan X, et al. Roundtable discussion: progress of urban informatics in urban planning. Front Urban Rural Plan 2023;1:23.

[11] Shi W, Goodchild M, Batty M, et al. Prospective for urban informatics. Urban Inform 2022;1:2.

[12] Dangermond J, Goodchild MF. Building geospatial infrastructure. Geospat Inf Sci 2020;23:1-9.

[13] Openshaw S. The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography (CATMOG), 1984.

[14] Johnson S. The ghost map: a street, an epidemic and the hidden power of urban networks. Penguin UK; 2008.

[15] Smallman-Raynor M, Cliff AD, Ord JK, et al. A Geography of infection: spatial processes and patterns in epidemics and pandemics. Oxford: Oxford University Press; 2022.

[16] Munoz JM, Naqvi A. Artificial intelligence and urbanization: the rise of the Elysium city. J Econ Polit Econ 2017;4:1–13.

[17] Yogarajan V, Dobbie G, Keegan TT. Debiasing large language models: research opportunities. J R Soc New Zealand 2025;55:372–95.

[18] Jiang H, Geertman S, Witte P. Smart urban governance: an alternative to technocratic “smartness.” Geo J 2022;87:1639–55.