Generative AI for Smart Cities: Mapping Models to Urban Applications — A Three-Axis Taxonomy

Pallavi Ranjan[1], Deepak Kumar Verma[2]
[1] School of Computer Science, University of Wollongong in Dubai
[1] Lincoln University College, Malaysia
[2] Lincoln University College, Malaysia;
[2] Marwadi University Rajkot, Gujarat, India

Email ID: PallaviRanjan@uowdubai.ac.ae; er.pallaviranjan@gmail.com; deepak.verma1980@gmail.com

Abstract: Generative artificial intelligence — generative adversarial networks, variational autoencoders, diffusion models, and large language models — has moved from research curiosity to applied tool across urban systems in under five years, now synthesizing mobility trajectories, generating household energy profiles, producing land-use designs, and functioning as spatio-temporal reasoning agents for city tasks. Despite this breadth, no existing survey classifies the field along model family, application task, and urban domain simultaneously, leaving practitioners without a principled basis for selecting a generative model for a given urban problem. This paper proposes a three-axis taxonomy — model family, task, and urban domain — documents a systematic multi-database literature search (9,763 records identified across four databases, 6,764 remaining after deduplication), presents a bibliometric keyword co-occurrence analysis corroborating the taxonomy's structure, identifies five knowledge gaps, and formulates five testable hypotheses for evaluation through comparative experimentation.

Keywords: generative artificial intelligence, smart cities, urban computing, taxonomy, digital twins, model selection, bibliometric analysis

1. Introduction

Generative artificial intelligence has moved from research curiosity to a working tool across urban systems in under five years. Generative adversarial networks (GANs) [1], variational autoencoders (VAEs) [2], denoising diffusion probabilistic models [3], and transformer-based large language models [4] now appear throughout the urban computing literature: synthesizing mobility trajectories for privacy-preserving data release [5,6], producing household smart-meter load profiles [7], generating land-use layouts and satellite imagery for planning [8], and acting as spatio-temporal reasoning agents for city tasks such as traffic signal control [9] and urban spatial cognition [10].

Three model families dominate the field, each with a distinct underlying mechanism. GANs use adversarial generator/discriminator training and remain strong on images and tabular records. Diffusion models generate through iterative denoising, offering high fidelity and strong conditional control at the cost of slower sampling. Transformers and large language models apply attention over sequences, giving them long-range dependence and a natural-language interface, though with comparatively weak numeric calibration. Applications of these families span every major urban domain: mobility and transport, energy

and utilities, the built environment, environment and climate, safety and health, and governance and services.

The volume of generative-AI-for-cities work is not matched by an ability to navigate it. A practitioner facing a concrete urban problem has no principled basis for choosing among model families, because the literature is organized by application area or by model family — seldom by both, and almost never by what the model is actually being asked to do. The phrase "generative AI for mobility," for example, covers at least four unrelated activities: synthesizing trajectories to protect privacy, forecasting demand, simulating counterfactual scenarios, and answering natural-language questions about a network. These require different architectures, are evaluated by different metrics, and sit at different levels of maturity. Collapsing them under one heading is a principal reason the field's practical guidance remains thin.

This gap is documented rather than asserted. The most recent comprehensive survey of generative AI combined with digital twins organizes 61 papers into six application areas — healthcare, industry, transportation, communication systems, energy/environment/infrastructure, and virtual scenarios [11]. The city is not among them. Urban problems are distributed across three of the six branches with no integrating category, and with no analysis of which model family suits which purpose within any of them.

Problem statement: Generative AI models are increasingly applied across urban domains, but no existing survey classifies this literature by model family, task, and urban domain together. As a result, it is unknown where the evidence for model-to-task fit is dense, where it is absent, and whether the field's own model choices track task requirements or something else.
Research question: What does the current literature reveal about which generative AI models are applied to which urban tasks and domains, and where does the resulting evidence base show gaps or inconsistency?

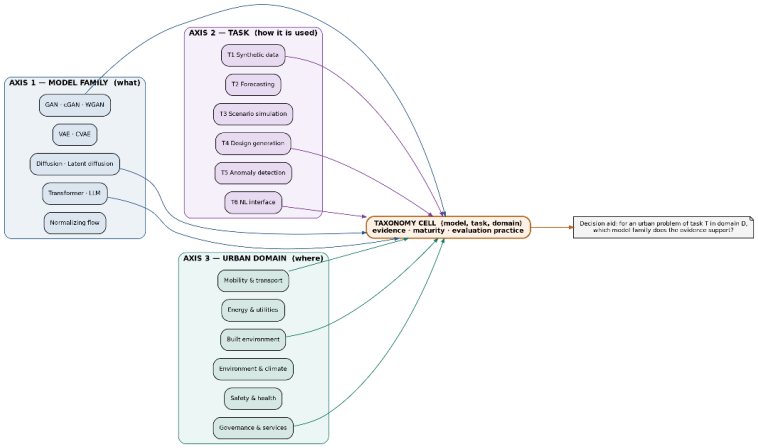


Figure 1. The three-axis taxonomy: model family (what), task (how it is used), and urban domain (where). Every surveyed paper is coded on all three axes, converting the review into a decision aid rather than a reading list.

2. Related work

2.1 Search Strategy and Screening

Four databases were searched systematically. A fifth, ACM Digital Library, was identified as relevant but could not be searched due to lack of institutional access; this is a documented limitation of the review's coverage rather than a null result and is treated as such throughout.

Database	Role	Records
Scopus	Primary	3,700
IEEE Xplore	Secondary	667
ScienceDirect	Secondary	3,927
Web of Science	Cross-check	1,469
ACM Digital Library	Not searched — no institutional access	—

2.1.1 Databases and Queries

Representative search query (Scopus, adapted to local syntax for each other database):

```
TITLE-ABS-KEY( ("generative AI" OR "generative adversarial network" OR "diffusion model" OR "variational autoencoder" OR "large language model" OR "foundation model") AND ("smart city" OR "urban" OR "built environment" OR "urban planning" OR "digital twin" OR "municipal") ) AND PUBYEAR > 2019 AND LIMIT-TO(LANGUAGE, "English")
```

Domain-specific supplementary queries were run in Scopus for each of the six urban domains individually, since a single combined query returns only surface-level coverage for six distinct domains at once - each domain has its own established vocabulary the umbrella query does not fully capture.

2.1.2 Inclusion and Exclusion Criteria

Inclusion	Exclusion
I1 Applies at least one generative model — GAN, VAE, diffusion, normalizing flow, or transformer/LLM used generatively	E1 Uses "smart city" or "urban" only as motivational framing, with no urban data or application
I2 Addresses an urban domain: mobility, energy, built environment, environment, safety, or governance	E2 Generative model applied to a non-urban domain with no explicit urban transfer discussed
I3 Reports a concrete application, dataset, or evaluation — not a purely conceptual proposal	E3 Purely discriminative or predictive model with no generative component
I4 Performs an identifiable task: synthesis, forecasting, simulation, design, anomaly detection, or NL interface	E4 Output is aesthetic or illustrative only, with no decision-support or evaluation claim
I5 Published January 2020 – present	E5 Editorial, commentary, or non-peer-reviewed report
I6 Written in English	E6 Duplicate or extended-abstract version — retain the fuller version
I7 Peer-reviewed; influential arXiv preprints admitted and flagged	E7 Full text inaccessible through available institutional access

2.1.3 Screening Flow

Deduplication was performed in Zotero across all four database exports. Of 9,763 records identified, 6,764 remained after duplicate removal — a reduction of approximately 31%, consistent with expected overlap between databases indexing overlapping journal and conference coverage. Title/abstract screening and

full-text assessment against the criteria in Section 5.2 are in progress at the time of this submission; the flow diagram below reflects the current state, with completed stages populated and remaining stages marked pending.

2.2 Generative Models for Structured and Urban Data

Modern synthetic data generation for structured data was established by CTGAN and TVAE, which introduced mode-specific normalization and conditional sampling to handle the mixed discrete/continuous columns and imbalanced categories typical of real-world records [12]. These remain standard baselines in nearly every subsequent benchmark. For urban mobility specifically, the dominant line of work is trajectory synthesis. LSTM-TrajGAN generates privacy-preserving synthetic trajectories end-to-end, introducing a TrajLoss function that balances spatial, temporal, and thematic similarity, and evaluates privacy through the trajectory-user-linking task rather than a generic tabular attack [5]. CATS extends this approach with conditional adversarial training on k-anonymized mobility matrices and attention-based global context learning, evaluated across privacy preservation, spatiotemporal characteristic preservation, and downstream utility on over 90,000 GPS trajectories [6]. The wider field of deep learning for human mobility is surveyed comprehensively in [13]. The underlying motivation for this line of work is empirical: [14] demonstrated that four spatio-temporal points suffice to uniquely identify 95% of individuals in a fifteen-month mobility dataset spanning 1.5 million people, with uniqueness decaying only as roughly the 1/10 power of spatial or temporal resolution — meaning that simple coarsening provides little genuine anonymity.

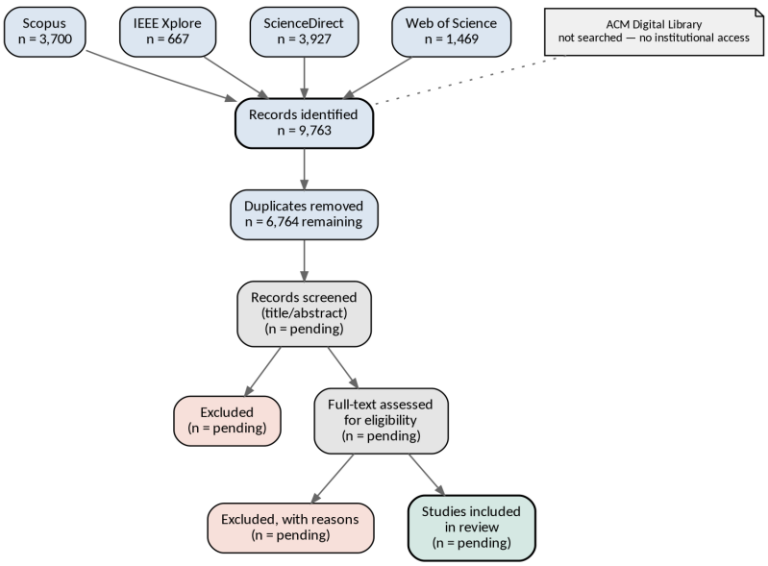


Figure 3. PRISMA-style search and screening flow. Identification and deduplication stages are complete; screening, eligibility assessment, and final inclusion remain pending at the time of this submission.

In urban energy, synthetic household smart-meter load profiles are generated to overcome the same data-sharing barrier. Faraday generates synthetic load profiles conditioned on attributes such as electric-vehicle ownership and evaluates outputs across fidelity, utility, and privacy [7]. Open building-level energy data at scale is also available: the Building Data Genome Project 2 provides hourly meter readings from

3,053 meters across 1,636 non-residential buildings at 19 sites [15], offering a candidate benchmark dataset for the energy domain that requires no access negotiation.

2.3 Generative AI and Urban Digital Twins

The most recent comprehensive survey of generative AI combined with digital twins organises 61 papers into six application areas: healthcare, industry, transportation, communication systems, energy/environment/infrastructure, and virtual scenarios [11]. Notably, the city is not one of the six categories — urban problems are distributed across transportation, energy/environment, and virtual scenarios, with no branch treating the city as an integrated system, and no analysis of which model family is best suited to which purpose.

2.4 Built Environment and Language Interfaces

Design generation in the built environment is diffusion-dominated. Recent work extends stable diffusion with ControlNet to generate satellite imagery conditioned on land-use descriptions, infrastructure, and natural environment, spatially linked to OpenStreetMap across three US cities, with generated layouts rated comparable to or better than real imagery by evaluators [8].

Language-interface applications form the newest and fastest-growing cluster. UrbanGPT introduces spatio-temporal large language models for prediction tasks [16]; CityGPT targets urban spatial cognition [10]; CityBench evaluates LLM capabilities across a battery of urban tasks [17]; and LLMLight deploys large language models directly as traffic signal control agents [9]. The subfield is surveyed in [18], and LLM-empowered agent-based simulation more broadly in [19].

2.5 Comparison of Existing Surveys

The table below compares this review against the closest existing surveys along the three axes proposed in this paper. Every prior survey fixes at least one axis — model family, urban domain coverage, or task — which is precisely why comprehensive model-selection guidance for urban generative AI does not yet exist.

Table 1. Compares this work with the related work or previous research by other researchers

Survey	Model families	Urban domains	Task axis	Key limitation
[11]	All five	Six non-urban categories	No	No city-level category; applications never mapped to models
[13]	Broad deep learning	Mobility only	Partial	Single domain; predates the diffusion and LLM waves
[18]	LLM only	Multiple	Partial	Single model family; no cross-family comparison
[20]	GAN, diffusion	Built environment	Partial	Single domain; design task only
[19]	LLM only	Cross-sector	No	Not urban-specific; one family
This	**All five**	**Six urban**	**Yes —**	**To be established**

work		domains	six tasks	

3. Bibliometric Validation: Keyword Co-occurrence Analysis

To corroborate the taxonomy independently of manual coding, a keyword co-occurrence network was generated in VOSviewer [21] from the search corpus. Each node represents a term appearing across the included literature; node size reflects frequency, and terms are grouped by colour into clusters based on co-occurrence strength.

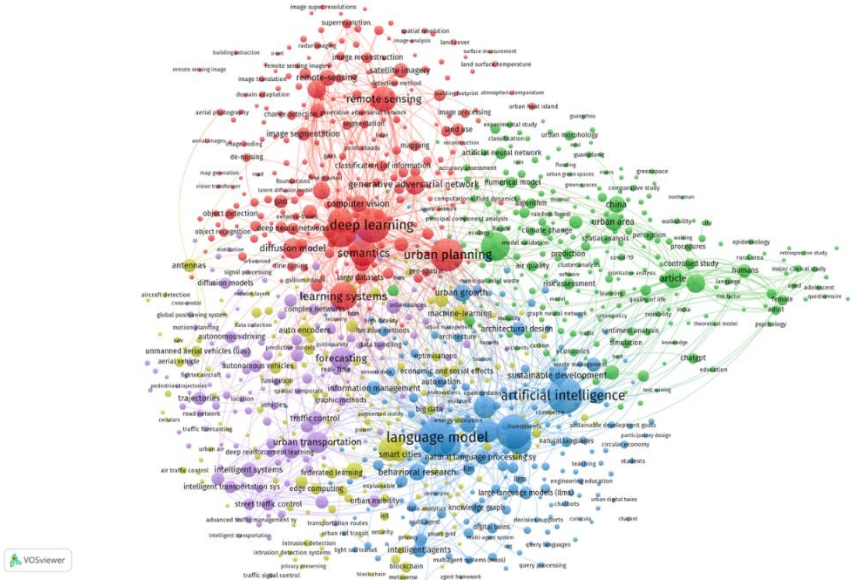


Figure 2. Keyword co-occurrence network generated in VOSviewer from the search corpus, showing five thematic clusters corresponding closely to the proposed taxonomy's urban domains.

Five clusters emerge, mapping closely onto the taxonomy's domain axis discussed in next subsection.

3.1 Thematic Clusters

. The map corroborates the taxonomy's domain structure independently of manual coding — and it reveals one striking asymmetry: five domains form dense, distinct clusters, while a sixth never coheres into one at all.

Built environment (red cluster) is the densest cluster on the map, centered on deep learning, urban planning, and remote sensing, with tight links to GANs and diffusion models. This is the domain where generative AI has moved furthest from proof-of-concept toward applied design tooling.

Environment and public health (green cluster) centers on climate change and air quality, but bleeds outward into epidemiology and human-subjects terms — humans, psychology, covid-19. Environmental applications here are frequently entangled with public-health research rather than standing as a clean domain of their own.

Governance and language interfaces (blue cluster) is built around artificial intelligence and large language models, reaching into digital twins, intelligent agents, and chatbots. Its loose connection to the older clusters marks it as the least mature domain structurally, consistent with the taxonomy's own findings.

Mobility and transportation (purple and yellow clusters) is the one point worth dwelling on. Transportation is the only domain spanning two separate clusters rather than one: purple covers trajectory-level generation (autonomous driving, trajectories), yellow covers system-level control (intelligent transportation systems, traffic signal control, edge computing). That split matters for two reasons. It signals real task diversity within one domain — data generation and real-time control are different problems, not the same one twice. And counted together, the two clusters make transportation bibliometrically larger than any single other domain, with denser internal linkage and more connections outward, including into the LLM cluster via smart mobility. That combination — task diversity plus density — is exactly what a comparative evaluation framework needs, since testing whether the best model family changes with the task requires a domain where multiple tasks already have populated, comparable methods.

Energy and utilities are the visible absence. No cluster corresponds to smart-meter data or grid forecasting; smart grid appears only as a peripheral node attached to the blue cluster. This isn't for lack of underlying work — it means that work hasn't yet reached the citation density needed to register as its own cluster, and its independent confirmation of the coverage gap identified elsewhere in this review.

Taken together, the map shows one clear center of gravity (built environment), one large but fragmented growth area (transportation), one domain entangled with an adjacent field (environment/health), one immature but emerging domain (language interfaces), and one domain invisible despite being real (energy). That structure is what makes transportation the most defensible starting point for the comparative work that follows.

Two limitations of the current stage should be stated plainly. First, ACM Digital Library was not searched due to lack of institutional access; this is a coverage gap in the search strategy, not a finding, and is disclosed as such. Second, title/abstract and full-text screening against the criteria in Section 5.2 remain in progress; the comparison of existing surveys (Section 2.4) and the knowledge gaps (Section 3) are therefore drawn from the literature identified to date and will be revisited once screening is complete, with any gap statement contradicted by the final corpus to be revised or withdrawn.

4. Conclusion

This paper has laid the foundation for a four-conference postdoctoral research programme on generative AI for smart cities. It has conducted a comprehensive literature review; proposed a three-axis taxonomy of model family, task, and urban domain; documented a systematic, multi-database search yielding 9,763 identified and 6,764 deduplicated records; corroborated the taxonomy's structure through independent bibliometric analysis; identified five knowledge gaps; and formulated five testable hypotheses to be evaluated experimentally in subsequent conferences. Immediate next steps are to complete title/abstract and full-text screening against the stated criteria, finalize the included corpus and update the comparison table and knowledge gaps accordingly, and begin designing the comparative evaluation framework (Objective O2) that will be presented at the Second Conference.

References

[1] Goodfellow, I. et al. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems (NeurIPS), 27.

[2] Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. International Conference on Learning Representations (ICLR).

[3] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. NeurIPS, 33.

[4] Vaswani, A. et al. (2017). Attention is all you need. NeurIPS, 30.

[5] Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A deep learning approach to trajectory privacy protection. GIScience 2021, LIPIcs, 177, 12:1–12:17. DOI: 10.4230/LIPIcs.GIScience.2021.I.12

[6] Rao, J., Gao, S., & Zhu, S. (2023). CATS: Conditional Adversarial Trajectory Synthesis for privacy-preserving trajectory data publication using deep learning approaches. International Journal of Geographical Information Science, 37(12), 2538–2574.

[7] Wang, S. et al. (2024). Faraday: Synthetic smart meter generator for the smart grid. arXiv:2404.04314.

[8] Wang, Q., Liang, Y., Zheng, Y., Xu, K., Zhao, J. and Wang, S., 2025. Generative AI for urban planning: Synthesizing satellite imagery via diffusion models. Computers, Environment and Urban Systems, 122, p.102339.

[9] Lai, S., Xu, Z., Zhang, W., Liu, H., & Xiong, H. (2025). LLMLight: Large language models as traffic signal control agents. Proc. 31st ACM SIGKDD, 2335–2346.

[10] Feng, J. et al. (2025). CityGPT: Empowering urban spatial cognition of large language models. Proc. 31st ACM SIGKDD, 591–602.

[11] Chiaro, D., Qi, P., Pescapè, A., & Piccialli, F. (2025). Generative AI-empowered digital twin: A comprehensive survey with taxonomy. IEEE Transactions on Industrial Informatics, 21(6), 4287–4295. DOI: 10.1109/TII.2025.3540473

[12] Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. NeurIPS, 32.

[13] Luca, M., Barlacchi, G., Lepri, B., & Pappalardo, L. (2023). A survey on deep learning for human mobility. ACM Computing Surveys, 55(1), 1–44. DOI: 10.1145/3485125

[14] de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. Scientific Reports, 3, 1376. DOI: 10.1038/srep01376

[15] Miller, C. et al. (2020). The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. Scientific Data, 7, 368. DOI: 10.1038/s41597-020-00712-x

[16] Li, Z. et al. (2024). UrbanGPT: Spatio-temporal large language models. Proc. 30th ACM SIGKDD, 5351–5362.

[17] Feng, J. et al. (2025). CityBench: Evaluating the capabilities of large language models for urban tasks. Proc. 31st ACM SIGKDD.

[18] Li, Z., Xia, L., Ren, X., Tang, J., Chen, T., Xu, Y., & Huang, C. (2025). Urban computing in the era of large language models. arXiv:2504.02009.

[19] Gao, C. et al. (2024). Large language models empowered agent-based modeling and simulation: A survey and perspectives. Humanities and Social Sciences Communications, 11, 1259.

[20] Liu, C., Li, G., Huang, C., Herzog, O., Ritter, H., Ma, S. and Ye, Y., 2026. Generative AI for Urban Planning and Design: Progress Review and Future Perspectives. Engineering.

[21] Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538.

[22] Hansen, L., Seedat, N., van der Schaar, M., & Petrovic, A. (2023). Reimagining synthetic tabular data generation through data-centric AI: A comprehensive benchmark. NeurIPS, 36, 33781–33823. arXiv:2310.16981

[23] Stadler, T., Oprisanu, B., & Troncoso, C. (2022). Synthetic data — Anonymisation groundhog day. 31st USENIX Security Symposium, 1451–1468.

[24] Yan, C., Yan, Y., Wan, Z., Zhang, Z., Omberg, L., Guinney, J., Mooney, S. D., & Malin, B. A. (2022). A multifaceted benchmarking of synthetic electronic health record generation models. Nature Communications, 13(1), 7609. DOI: 10.1038/s41467-022-35295-1