

<https://doi.org/10.1038/s42949-026-00377-2>

# Bridging urban theory and artificial intelligence: a multi-agent recommendation system for sustainable city development

Jiawei Tong<sup>1,2</sup>, Shuihua Wang<sup>1</sup> ✉, Guangyu Wang<sup>3</sup>, Yan Wang<sup>4</sup> & John Moraros<sup>1</sup> ✉

As cities increasingly rely on AI for sustainability challenges, a critical gap emerges: AI applications in urban planning and safety predominantly proceed without explicit guidance from established urban theories that have guided sustainable development for decades. Our analysis reveals that technology-driven research dominates the field, while problem-driven approaches addressing genuine urban needs remain minimal. To bridge this theory-practice disconnect, we develop a Large Language Model (LLM)-based multi-agent recommendation system [publicly available online] that realigns AI development with sustainable city principles. The system employs specialized agents to recommend appropriate theoretical frameworks, AI methods, and data sources for urban challenges, drawing from classical urban theories. Through diverse case studies, we demonstrate how our approach transforms technology-focused solutions into theory-grounded interventions that address sustainability's interconnected dimensions. Our framework fundamentally shifts the question from “what can algorithms do?” to “what does this urban challenge require for sustainable outcomes?”—ensuring AI amplifies rather than replaces the theoretical wisdom essential for creating resilient, equitable, and livable cities that contribute to global sustainability targets.

Artificial intelligence (AI) offers unprecedented opportunities for optimizing urban systems as cities worldwide race to meet sustainability targets. Yet, our large-scale empirical analysis of 1123 AI applications in urban planning and safety across global contexts reveals a critical misalignment: current AI applications in urban safety and planning areas proliferate in a theoretical vacuum, prioritizing technical metrics over the social, environmental, and spatial dynamics essential for sustainable cities<sup>1,2</sup>. Without grounding in established urban wisdom, even the most sophisticated AI system may inadvertently compromise the resilience, equity, and livability that future generations depend upon<sup>3,4</sup>. Recent frameworks further demonstrate AI's potential to operationalize sustainability targets through systemic approaches addressing emissions, inequality, and exclusion simultaneously<sup>5</sup>, yet such integration requires explicit theoretical grounding to avoid techno-solutionist pitfalls<sup>6</sup>.

This misalignment stems from AI development processes that bypass decades of urban theoretical knowledge. Established urban theories—from Jacobs' social surveillance principles<sup>7</sup> to Newman's spatial security

frameworks—demonstrate that sustainable cities emerge from integrated social-spatial-economic systems. But as AI proliferates in urban sustainability efforts—from carbon reduction to equitable resource distribution—few applications engage with this legacy. Comprehensive reviews of AI, IoT, and big data convergence in sustainable smart cities<sup>8</sup> confirm this pattern: despite impressive technological sophistication, theoretical integration remains minimal. Consequently, these applications risk replicating a pattern that scholars have critiqued: optimising for efficiency while neglecting the holistic principles of sustainable urbanism<sup>6,9,10</sup>.

Quantitative analysis confirms limited explicit theoretical integration. Among 1123 studies representing 151.6-fold growth since 2008, only 1.16% explicitly cite established urban theories by name. Nearly half (47.3%) are technology-driven, focused on algorithmic capabilities rather than urban crises (8.6%)<sup>11</sup>. This imbalance demonstrates that urban AI research predominantly asks “what can AI do?” rather than “what do cities need for sustainable development?”—signaling a fundamental misalignment between how AI is being developed and how sustainable cities actually function<sup>12,13</sup>.

<sup>1</sup>Department of Biosciences and Bioinformatics, Suzhou Municipal Key Lab AI4Health, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, China.

<sup>2</sup>Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Liverpool, UK. <sup>3</sup>Department of Computer Science, Data Science, and Engineering, New York University Shanghai, Pudong, China. <sup>4</sup>School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China.

✉ e-mail: [shuihua.wang@xjtlu.edu.cn](mailto:shuihua.wang@xjtlu.edu.cn); [john.moraros@xjtlu.edu.cn](mailto:john.moraros@xjtlu.edu.cn)

**Table 1 | Algorithm and data source grouping summary**

Algorithm groups		Data source groups	
Group	Group name	Group	Group name
0	Optimization and Decision Methods	0	Environmental and Climate Data
1	Statistical and Spatial Analysis	1	Sensor and Internet of Things (IoT) Data
2	Machine Learning (ML) and Deep Learning	2	Geographic Information and Social Survey Data
3	Generative and Pre-trained Models	3	Composite Data
		4	Regional and Experimental Data
		5	Image and Remote Sensing Data

This misalignment has intensified dramatically over time, with particularly concerning implications for urban sustainability. Since 2020 alone, over 60% of all urban AI applications have emerged, coinciding with the proliferation of pre-trained models and accessible AI frameworks. The acceleration is striking: newer algorithms achieve near-instantaneous adoption. Contrastive Language-Image Pre-training (CLIP) and Generative Pre-trained Transformer 3 (GPT-3) were applied to urban problems within the same year of their release, reflecting intense pressure to demonstrate innovation regardless of sustainability outcomes<sup>14</sup>. Consequently, the field increasingly prioritizes algorithmic novelty over addressing genuine urban sustainability challenges, further widening the gap between technological capabilities and urban needs<sup>5,16</sup>.

We address this gap through a theory-first AI recommendation framework that prioritizes urban sustainability requirements over algorithmic capabilities. This approach advances recent proposals for human-AI symbiosis in urban informatics<sup>17</sup>, positioning AI not as an autonomous solution but as a collaborative partner that amplifies human expertise. Specifically, our system employs five specialized agents working sequentially—from problem structuring to theory retrieval, algorithm matching, data selection, and integration validation. Through this process, the framework transforms the core question from “what can this algorithm do?” to “what does this urban challenge require?”<sup>18</sup>. To achieve this, the system maps alignments between 46 classical urban theories and contemporary AI methods, thereby generating recommendations that integrate sustainability’s social, environmental, and economic dimensions<sup>19</sup>. We validate this approach through three case studies spanning food security, urban heat mitigation, and disaster resilience. The results demonstrate how theory-grounded AI generates solutions aligned with Sustainable Development Goal (SDG) 11 targets, ultimately enabling cities to transcend technical performance metrics and achieve genuine sustainability outcomes<sup>20,21</sup>.

## Results

### The technology-first paradigm

Our empirical investigation examines four key dimensions of AI adoption in urban planning and safety domains: algorithmic preferences, data source utilization, theoretical integration, and research motivations. Our systematic analysis reveals the mechanisms driving the technology-first paradigm. Such evidence establishes the foundation for understanding the theory-practice disconnect in urban AI applications.

We retrieved an initial corpus of 2797 papers from the Web of Science (WoS)<sup>22</sup> and Scopus<sup>23</sup> databases using the targeted search query (“urban planning” OR “urban safety”) AND “AI”, with no time restriction, and performed initial data cleaning. After further data cleaning and duplicate removal, we conducted systematic screening to ensure quality. A BERT-based classifier performed domain-specific classification (see Supplementary Material 8, Table S10). We then filtered the results to peer-reviewed research articles<sup>24</sup>, ultimately identifying 1123 genuine AI applications in the urban planning and safety domains. This BERT-based approach follows established practices in natural language processing<sup>25</sup>, with similar methods recently applied to the classification of urban sustainability literature, achieving comparable accuracy levels<sup>26</sup>. These 1123 validated papers form the foundation

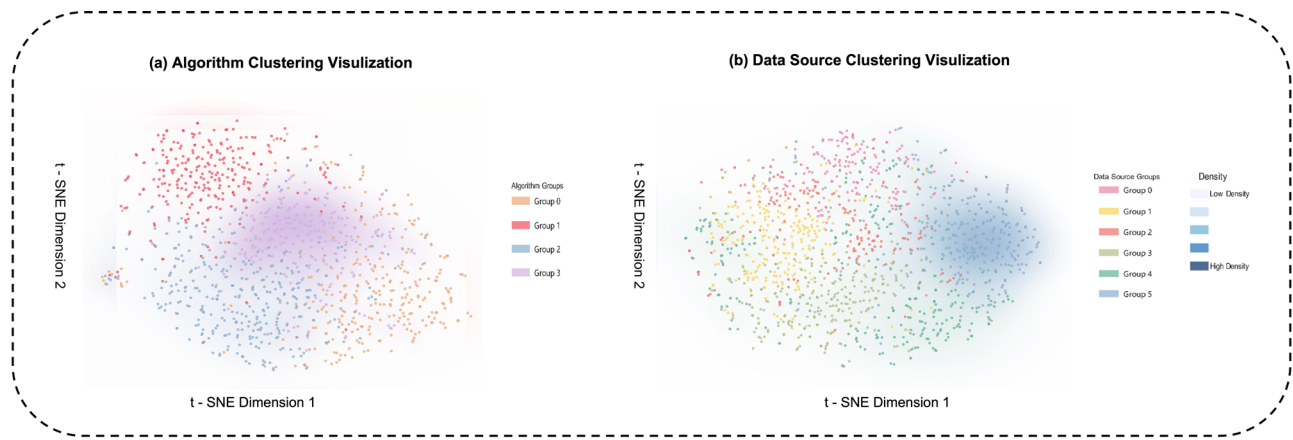
for examining the current state of AI integration in urban planning and safety management.

Four distinct algorithm groups and six data source categories emerged from clustering analysis. Dimensionality reduction using t-distributed Stochastic Neighbour Embedding (t-SNE) reveals distinct clustering patterns in both algorithmic approaches and data sources utilised in urban planning and safety AI research (Table 1, Fig. 1). The t-SNE method has proven effective for visualizing high-dimensional urban data structures<sup>27,28</sup>, with recent theoretical advances supporting its reliability for cluster detection in complex datasets<sup>29</sup>.

Machine Learning (ML) and Deep Learning methods (Group 2) dominate the algorithmic landscape, comprising 42.7% of all methods employed and forming the densest cluster (Fig. 1a)—a finding consistent with recent systematic reviews documenting the proliferation of ML approaches in urban applications<sup>30,31</sup>. Traditional optimization and statistical approaches (Groups 0–1: Optimization & Decision Methods, and Statistical & Spatial Analysis) demonstrate established methodological foundations, while the emergence of Generative & Pre-trained Models (Group 3) reflects recent advances in AI capabilities for urban applications.

Environmental & Climate Data and Sensor & Internet of Things (IoT) Data (Groups 0–1) exhibit the highest utilization among data sources, accounting for 68.3% of all data usage (Table 1, Fig. 1b). This pattern reflects the maturity and accessibility of environmental monitoring systems and sensor networks in urban contexts<sup>32,33</sup>. Composite Data, Regional & Experimental Data, and Image & Remote Sensing Data (Groups 3–5) account for only 19.2% of combined usage (Fig. 1b), despite all 46 theories requiring at least one of these categories at the recommended relevance threshold ( $\geq 0.6$ ), with 69.6% requiring two or more. Meanwhile, Environmental & Sensor data (Groups 0–1) is the only grouping where observed usage exceeds theory-derived demand (68.3% vs. 58.7%), consistent with technology-driven selection favoring readily available automated sources<sup>34</sup>. Geographic Information & Social Survey Data (Group 2) occupies an intermediate position, bridging traditional geospatial approaches with contemporary data collection methods. The difference in utilization between established environmental/sensor data sources and emerging composite data approaches was statistically significant ( $\chi^2 = 156.4$ ,  $p < 0.001$ ).

Only 13 of 1123 papers (1.16%) demonstrated explicit theoretical integration by citing established urban planning or safety theories by name. This analysis reveals limited visible theoretical communication between AI applications and established urban frameworks—a pattern that recent critiques have identified as problematic for sustainable urban development<sup>35</sup>. None of the 63 Random Forest applications analyzing urban spatial patterns reference spatial interaction theories. Similarly, Convolutional Neural Network (CNN) - based urban image analyses ( $n = 47$ ) contain no citations to established visual assessment frameworks from urban design literature<sup>36</sup>. Among the 13 papers that did engage with theory, 10 were authored by interdisciplinary teams including at least one urban planning or safety expert, suggesting the importance of cross-disciplinary collaboration<sup>37</sup>. This theoretical vacuum is particularly concerning given the diversity of algorithmic approaches identified (Table 1), each of which requires domain-specific adaptation.



**Fig. 1 | Algorithm and data source grouping visualisation using t-SNE dimensionality reduction.** **a** Four algorithm groups identified through clustering, with Machine Learning & Deep Learning (Group 2) showing highest density (42.7% of

papers). **b** Six data source groups revealing clear separation between traditional (Groups 0–1, 68.3% combined) and emerging data types (Groups 3–5, 19.2% combined).

The distribution of research motivations (see Supplementary Material 3) (reveals a technology-first paradigm dominating urban planning and safety AI studies (Fig. 2a). Technology-driven papers constitute 47.3% of the corpus (531 papers), focusing on exploring new technological applications, algorithm development, and method transfer—a trend multiple reviews have identified as misaligned with urban planning needs<sup>38,39</sup>. Method-driven research accounts for 14.0% (157 papers) and addresses the limitations of existing solutions through performance improvements or efficiency enhancements. In stark contrast, problem-driven research initiated by specific urban crises or real-world events comprises only 8.6% (97 papers)—examples include Beijing’s traffic congestion incidents, New York City’s flooding events, and Tokyo’s earthquake evacuation crises. General needs-driven studies account for 17.0% (191 papers) and address broad urban requirements without specific triggering events. The remaining 147 papers (13.1%) exhibit unclear or mixed motivations (Table 2).

Algorithm selection criteria analysis reveals a distinct hierarchy in researchers’ priorities (Fig. 2b): technical novelty (59.0%), algorithmic performance (48.7%), computational efficiency (51.3%), ease of implementation (10.3%), domain-specific applicability (12.8%), and geographical generalizability (5.1%). This distribution demonstrates a clear preference for technical metrics over practical relevance in urban planning applications. Rigorous domain validation analysis reveals significant gaps in verification practices. Only 12.8% of studies conduct meaningful domain-specific verification. More critically, comprehensive multi-method validation is completely absent (0.0%). Geographical generalizability represents another critical gap. This factor is essential for urban planning algorithms intended for diverse global contexts, yet remains minimally addressed. Only 5.1% of studies conduct rigorous cross-city validation, while 97.4% fail to address cultural adaptability.

### Validation of theoretical integration measurement

Our 1.16% finding captures explicit theoretical citations, but may not reflect the full extent of theoretical engagement. To evaluate this measurement’s validity and explore implicit theoretical foundations, we conducted a three-method validation protocol on a stratified subsample ( $n = 200$ , see Section “Validation of Theoretical Integration Measurement” and Supplementary Material 10). Automated classification using GPT-4-turbo to detect theoretical engagement from introduction sections achieved only 43.0% accuracy (86/200 correct; Cohen’s  $\kappa = 0.35$ ) against expert coding. Among 114 misclassifications, primary errors included over-interpretation of technical descriptions (36 cases, 31.6%), missing implicit theoretical cues (32 cases, 28.1%), and surface citation misjudgment (29 cases, 25.4%). For example, the model classified Xiang et al.’s water resource management study<sup>40</sup> as “theory-driven” by misinterpreting sustainability terminology, while missing implicit travel behavior theory in Xin et al.’s mobility analysis<sup>41</sup>. This poor performance demonstrates why explicit citation criteria became

necessary—automated semantic analysis cannot reliably detect domain-specific theoretical foundations at scale.

Team composition analysis revealed paradoxical patterns. Contrary to expectations, disciplinary diversity showed no significant correlation with theory-driven intensity (Pearson’s  $r = 0.08$ ,  $p = 0.267$ ; Spearman’s  $\rho = 0.09$ ,  $p = 0.221$ ). However, pure urban planning teams scored highest ( $0.65 \pm 0.24$ ), followed by interdisciplinary teams ( $0.39 \pm 0.23$ ) and pure computer science teams ( $0.24 \pm 0.18$ ). To further validate these team-level patterns and explore implicit theoretical engagement, semantic theoretical density (STD) analysis using an 847-term theoretical lexicon revealed moderate correlation with expert assessment ( $r = 0.62$ ), slightly higher than explicit citation ( $r = 0.58$ ), though not significantly different (Steiger’s  $Z = 0.54$ ,  $p = 0.59$ ). Critically, all three measurement approaches demonstrated identical rank ordering across team types (Table 3), with semantic density showing Pure Planning teams scoring 65.7% higher than Pure CS teams ( $0.58$  vs  $0.35$ ;  $t = 6.83$ ,  $p < 0.001$ , Cohen’s  $d = 1.34$ ). Among papers lacking explicit citations ( $n = 165$ , 82.5%), semantic density varied substantially (range: 0.18–0.74,  $SD = 0.19$ ). Manual review of top-quartile cases ( $n = 31$ ) revealed that 87.1% (27/31) demonstrated sophisticated domain knowledge through problem framing and methodological choices.

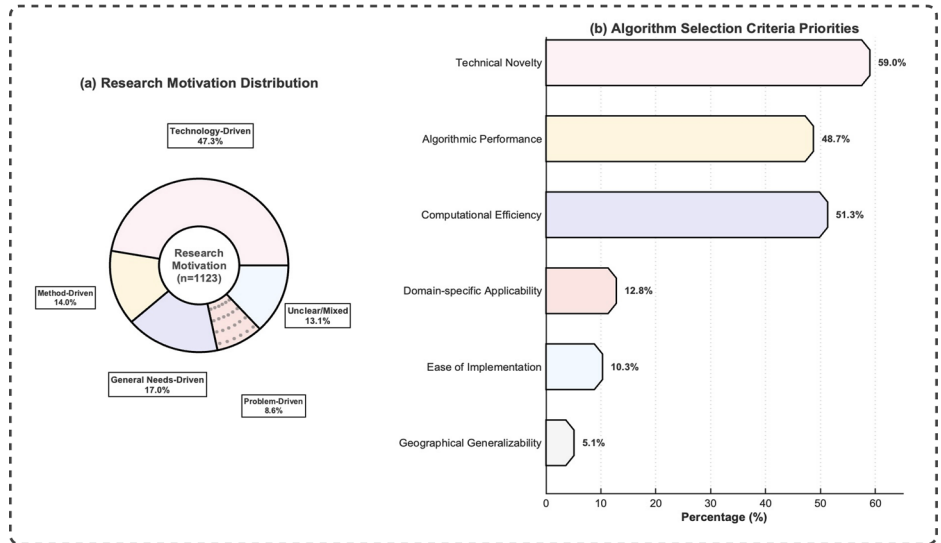
### Temporal evolution: from problem-pull to technology-push

To understand the drivers of AI adoption in urban planning and safety, we analyze the temporal gap between algorithm development and domain application across all 1123 papers. Our corpus captures self-labeled AI research—papers explicitly using “AI” terminology—rather than the complete history of computational methods in urban domains. The timeline visualization (Fig. 3) reveals an important distinction about algorithmic adoption. Rather than showing when methods first entered urban research, it captures when they became reframed under the AI label. Such reframing documents shifts in disciplinary discourse, consistent with Rogers’ diffusion of innovation theory<sup>42</sup>.

Our analysis reveals a dramatic transformation in AI adoption patterns. Prior to 2008, merely 14 papers (0.7%) employed AI methods in urban planning and safety contexts. Following 2008, applications surge to 1109 papers (99.3%)—a 151.6-fold increase visible in Fig. 3a. This transformation coincides with the broader AI renaissance, driven by computational advances and data availability<sup>43</sup>, rather than with emerging urban challenges requiring AI solutions. Critically, this growth reflects the proliferation of AI terminology and framing: many algorithms shown in Fig. 3 (particularly Groups 0–1) were established in planning practice decades earlier<sup>44,45</sup>, their appearance here marking terminological repackaging rather than initial methodological introduction.

Temporal evolution unfolds in three distinct phases clearly delineated in Fig. 3. Critically, the 2008 inflection point labeled “AI Terminology

**Fig. 2 | Analysis of research paradigms in urban planning and safety AI studies.** **a** Distribution of research motivations among 1,123 papers, showing the dominance of technology-driven approaches (47.3%) over problem-driven research (8.6%). **b** Algorithm selection criteria priorities, revealing emphasis on technical novelty (59.0%) and computational efficiency (51.3%) over domain-specific applicability (12.8%) and geographical generalizability (5.1%).



**Table 2 | Research motivation distribution in urban planning and safety AI studies (N = 1123)**

Research Motivation	Papers	Percentage
Technology-Driven	531	47.3%
Method-Driven	157	14.0%
General Needs-Driven	191	17.0%
Problem-Driven	97	8.6%
Unclear/Mixed	147	13.1%
Total	1123	100.0%

Adoption Surge” in panel (a) marks systematic reframing under AI terminology following deep learning breakthroughs, not algorithm development or latency occurrence—this initiates Phase 1 (2008–2012), characterized by sparse, experimental adoption of existing algorithms. Subsequently, Phase 2 (2012–2017) witnesses a machine learning revolution penetrating urban domains following AlexNet’s breakthrough demonstration<sup>46</sup>, exemplifying technology-push wherein innovations diffuse based on capability rather than domain need<sup>2</sup>. Finally, Phase 3 (2018–2025) represents an unprecedented explosion visible in panel (b): 2020–2025 alone contributes over 60% of all applications, driven by pre-trained models and accessible AI frameworks.

The latency between algorithm development and urban application reveals fundamental patterns in technology transfer consistent with diffusion theory<sup>42</sup>. Rather than absolute latency values, what matters most is their dramatic compression across time: Group 3 algorithms demonstrate near-instantaneous adoption—CLIP achieves same-year development and application (2021), while GPT-3 enters urban planning within months of release<sup>47</sup>. Such compression from decades (Groups 0–2) to essentially zero (Group 3), illustrated by converging lines in Fig. 3a, reflects intensifying pressure to demonstrate AI innovation regardless of domain relevance. Within-group heterogeneity (CV) decreases from 237% (G0) to 16% (G3) (Table 4), demonstrating convergence toward uniform, immediate adoption. Group-specific patterns shown in Fig. 3c–f further reveal mounting innovation pressure across the field.

Building on the latency patterns identified above, the 2020–2025 period represents an inflection point in algorithm density and diversity (Fig. 3b). The AlexNet moment of 2012<sup>46,48</sup>—when deep learning achieves a 15.3% top-5 error rate compared to 26.2% for traditional methods—catalyzes widespread AI adoption across domains. This “winner-take-all” performance gap creates what innovation theorists call a “technology

imperative”<sup>42</sup>, in which adoption is driven by the fear of being left behind rather than genuine utility.

Group 2 algorithms dominate the landscape with Random Forest (RF)<sup>49</sup>, Support Vector Machines (SVM)<sup>50</sup>, eXtreme Gradient Boosting (XGBoost)<sup>51</sup>, and Convolutional Neural Networks (CNNs)<sup>43</sup> (28) applied across disparate urban challenges, as detailed in Fig. 3e. The algorithm distributions reveal concerning patterns of methodological misalignment: CNNs designed for image recognition<sup>43</sup> are repurposed for non-visual planning problems, while natural language models process numerical safety metrics.

Group 3’s emergence, visualized in Fig. 3f, exemplifies the innovation imperative driving the field. Generative Adversarial Networks produce urban designs (17 applications) without engaging design theory, BERT analyzes planning documents while ignoring planning discourse analysis frameworks<sup>25</sup>, and GPT models automate report generation absent planning communication principles. From 2021 to 2025, virtually every newly released AI algorithm finds immediate application in urban contexts, irrespective of theoretical fit or practical necessity—a pattern Rogers<sup>42</sup> identifies as characteristic of hype-driven diffusion.

Simultaneously, Group 1 shows distinct patterns (Fig. 3d) with statistical and spatial analysis methods. While Group 0 maintains a steady presence (Fig. 3c), primarily in optimization tasks where mathematical foundations naturally align with planning objectives. These patterns, consistent with the latency characteristics shown in Table 4, demonstrate differential adoption dynamics across algorithm groups.

**Bridging theory and practice: a computational mapping framework**

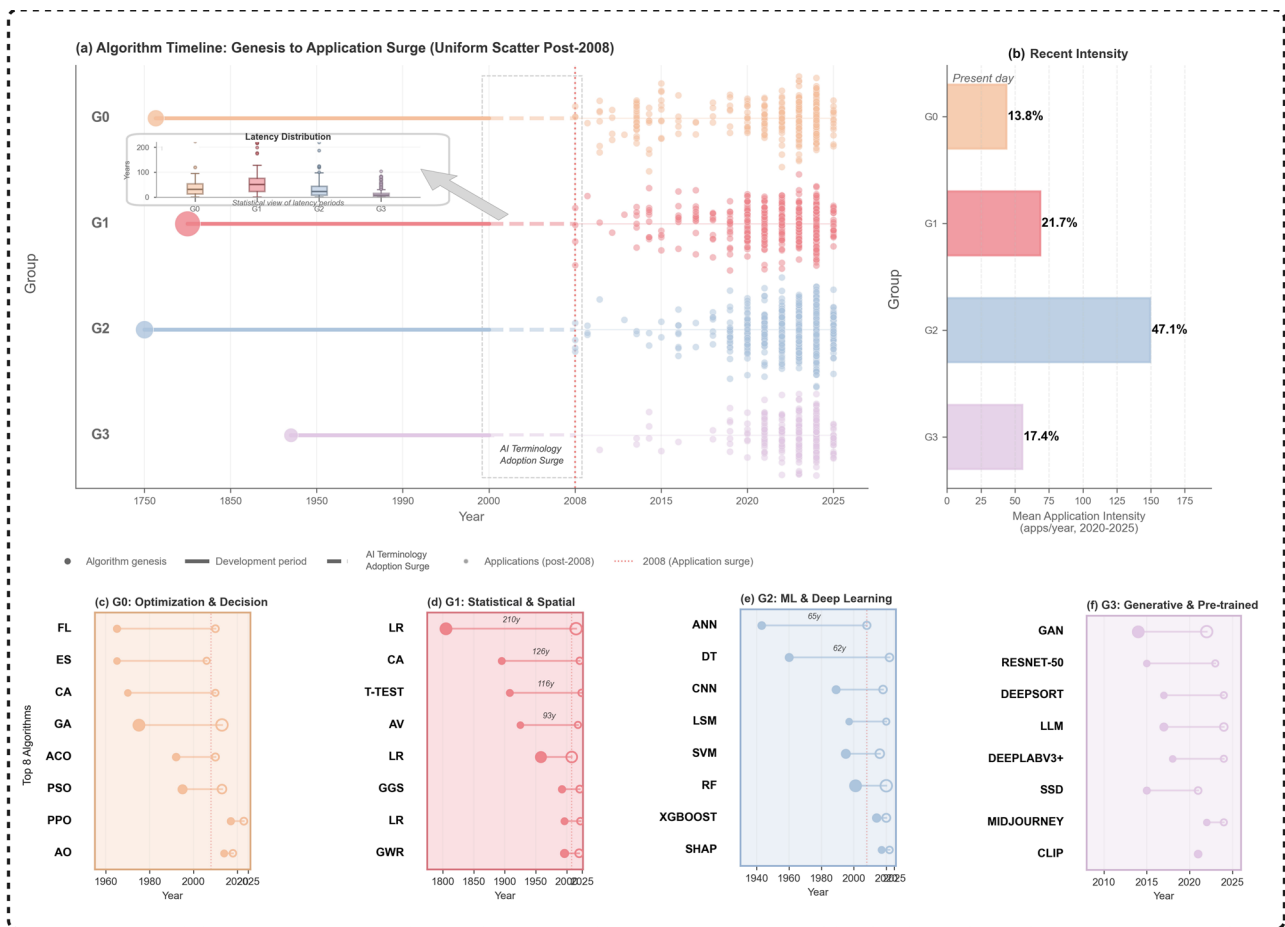
To address the theory-practice disconnect identified in our analysis, we develop a computational framework that systematically maps urban theories to appropriate AI methods and data sources. Our approach transforms the question from “what can this algorithm do?” to “what theoretical principles should guide this urban intervention?” Through systematic analysis of 46 classical urban theories, we establish explicit connections between theoretical frameworks and computational approaches, providing a foundation for theory-driven AI development in urban contexts.

We employ NLP (Natural Language Processing) techniques to systematically identify and categorize 46 classical theories in urban planning and safety (see Supplementary Material 5, Table S6). Table 5 presents the most influential theories based on citation frequency and temporal persistence in the literature. Our computational linguistics framework extracts 127 distinct computational principles from the 46 urban theories (see Supplementary Material 5), revealing how classical theories inherently

**Table 3 | Three-method validation results: measurement convergence analysis (N = 200)**

Measurement approach	Overall Mean ± SD	Pure planning	Inter-disciplinary	Pure CS	Correlation with expert assessment
Explicit citation	0.18 ± 0.38	0.35 ± 0.48	0.11 ± 0.31	0.00 ± 0.00	$r = 0.58^{***}$
Semantic density	0.46 ± 0.21	0.58 ± 0.18	0.42 ± 0.19	0.35 ± 0.22	$r = 0.62^{***}$
Expert assessment	0.45 ± 0.29	0.65 ± 0.24	0.39 ± 0.23	0.24 ± 0.18	—
<i>Pairwise Measurement Correlations</i>					
Citation ↔ Semantic density		$r = 0.54^{***}$			
Citation ↔ Expert assessment		$r = 0.58^{***}$			
Semantic ↔ Expert assessment		$r = 0.62^{***}$			

\*\*\* $p < 0.001$ . All three measures show identical rank ordering across team types. Steiger's Z-test: Z(Citation vs Semantic) = 0.54,  $p = 0.59$ .



**Fig. 3 | Temporal evolution of AI algorithm adoption in urban planning and safety research (2008–2025).** **a** Group-level timelines from development (left genesis points) to AI-labeled application (right scatter points). Genesis circles mark the earliest algorithms; c–f reveal within-group heterogeneity (Table 4). The “AI Terminology Adoption Surge” (2008) marks when methods became systematically framed as AI, not when developed. Inset boxplot shows the latency distribution; for

Groups 0–1, the means reflect all algorithms, not early outliers. **b** Application share 2020–2025 by group: G0 (13.8%), G1 (21.7%), G2 (47.1%), G3 (17.4%), showing G2 dominance. Algorithm-level heterogeneity: (c) G0, CV = 237%; (d) G1, CV = 38%; (e) G2, CV = 32%; (f) G3, CV = 16%. Color intensity represents application frequency. Dates reflect the first AI-labeled appearance; many methods existed in urban research decades earlier.

establish connections to modern AI methods. We select 11 representative theories spanning all four clusters for detailed mapping analysis against algorithms and data sources identified in the 1123 AI application papers (Fig. 4).

The mapping demonstrates how urban theories provide natural computational frameworks (Fig. 4). Taking CPTED as an illustrative example<sup>52,53</sup>, its “natural surveillance” principle translates to G1 spatial

analysis for viewed calculations, G2 machine learning for activity pattern recognition, and G0 optimization for sight-line configurations, with alignments formalized through logical consistency constraints (Supplementary Material 4, Table S5). CPTED’s “territorial reinforcement” principle further requires diverse data sources—environmental data for lighting, geographic data for layouts, image data for surveillance, and experimental data from field observations. Yet current implementations reveal systematic

**Table 4 | Algorithm adoption patterns in urban AI research**

Algorithm group	Mean Latency <sup>a</sup>	Range (years)	CV (%)	2020–25 Share	Adoption Pattern
G0: Optimization and Decision	30.9	11–210	237	16.0%	Steady integration
G1: Statistical and Spatial	48.1	13–55	38	22.9%	Delayed adoption
G2: ML and Deep Learning	28.0	3–24	32	44.5%	Accelerating diffusion
G3: Generative and Pre-trained	13.2	0–5	16	16.5%	Immediate transfer

<sup>a</sup>Mean latency reported in years. Groups 0–1 latencies primarily reflect AI terminology adoption rather than initial methodological introduction. Critical finding: Progressive heterogeneity compression (CV: 237% → 16%) demonstrates adoption pattern convergence—contemporary algorithms (G3) achieve near-zero latency with minimal variance (CLIP: 2021 → 2021), while classical methods (G0) span centuries with extreme variability (Linear Regression: 210y vs. PSO: 11y).

simplification<sup>54,55</sup>: at the *instance level*, G2 methods comprise only 36% of 89 algorithm applications from 47 CPTED papers, though *paper-level analysis* shows 72% employ at least one G2 method (Table S11)—indicating widespread but shallow adoption. More critically, only 18% of papers integrate three or more data types, substantially below the theory-derived demand of 78.3% for multi-modal integration. Neighborhood Unit theory<sup>56,57</sup> reveals complementary patterns: while Perry’s framework emphasizes service accessibility requiring socio-demographic data, current applications predominantly employ G1 statistical analysis (42% of papers) with environmental data (35%), reflecting research priorities focused on environmental assessment rather than population-centered optimization.

#### Case a: problem-driven research—urban food waste crisis

As illustrated in Fig. 5, the multi-agent system processes research inputs through three steps: theory identification, algorithm matching, and data source selection. We evaluate recommendations using five metrics (Table 6) that assess scenario complexity, theory-practice alignment, implementation feasibility, and solution robustness. Three representative cases demonstrate the system’s application across different research motivations: Case A (problem-driven) exemplifies the complete transformation process, while Cases B (method-driven) and C (technology-driven) highlight key variations. Importantly, these cases represent scenarios where preliminary assessment suggested AI exploration was warranted. The system does not claim universal AI applicability; many urban problems—particularly those centered on trust-building, political negotiation, or community empowerment—may require non-technical interventions outside our system’s scope.

A major US city faces the paradox of discarding 10,000 tons of edible food annually<sup>58</sup> while significant populations experience food insecurity. Consequently, the original approach employs continuous approximation methods to optimize vehicle routing, yet focuses solely on minimizing transportation costs. Initially, the Scenario Analyzer Agent decomposes the food waste crisis into seven structured dimensions. This decomposition reveals multiple stakeholders (donors, recovery organizations, vulnerable populations), temporal dynamics, and spatial considerations. As a result, the complexity score  $\xi(S) = 0.82$  indicates a multi-faceted challenge requiring integrated solutions. Subsequently, the Theory Retriever Agent identifies Urban Metabolism Theory<sup>59,60</sup> ( $\sigma = 0.91$ ), which conceptualizes food waste as systemic flow disruption, alongside Environmental Justice Theory<sup>61</sup> ( $\sigma = 0.87$ ), which reveals inequitable access patterns. Together, these complementary frameworks guide the design of equitable resource circulation.

Building on these theoretical foundations, the Algorithm Matcher Agent maps requirements to computational methods: specifically, Multi-objective Optimization with Decomposition (MOEA/D)<sup>62</sup> for efficiently balancing multiple competing objectives through problem decomposition, Graph Neural Networks (GNN)<sup>63</sup> for spatial equity analysis, and Temporal LSTM<sup>64</sup> for demand prediction. Consequently, the integrated framework achieves a capability score  $\text{cap}(A, r) = 0.89$ . In parallel, the Data Source Selector Agent prioritizes social vulnerability indices<sup>65</sup> ( $Q = 0.92$ ), followed by food supply data ( $Q = 0.85$ ), environmental data ( $Q = 0.81$ ), and transportation networks ( $Q = 0.78$ ). This expansion beyond routing data therefore encompasses equity-relevant sources critical to addressing food insecurity. Finally, the

Integration Validator Agent achieves robustness score  $R = 0.91 > R_{\min} = 0.85$  through Monte Carlo simulations. Moreover, expert review confirms effective integration of efficiency and equity objectives, particularly highlighting the vulnerability-weighted distribution mechanism that ensures food reaches high-need populations. Table 7 summarizes this transformation from technical optimization to systemic solution.

#### Case B: method-driven research—urban heat island prediction

Researchers addressing limitations in Urban Heat Island (UHI) prediction initially focus on improving accuracy from  $R^2 < 0.8$  to  $R^2 = 0.95$  using new stereoscopic urban morphology metrics. Recent studies have demonstrated the effectiveness of 3D urban morphological indicators<sup>66</sup> and building volume information<sup>67</sup> combined with XGBoost models, showing superior performance in predicting land surface temperature<sup>68</sup> and analyzing urban heat island drivers<sup>69</sup>. Our system identifies two relevant theoretical frameworks to expand this approach. Urban Climate Theory<sup>70</sup> ( $\sigma = 0.93$ ) provides a foundational understanding of heat island formation<sup>71</sup>, while Compact City Theory<sup>72</sup> ( $\sigma = 0.82$ ) reveals that compact cities, despite their sustainability goals, face environmental quality and heat stress challenges. This theoretical grounding leads to algorithmic integration. Physics-Informed Neural Networks (PINN) embed thermodynamic constraints by incorporating physical laws into neural network training<sup>73</sup>, making them suitable for heat transfer problems<sup>74</sup> and scientific machine learning applications<sup>75</sup>. Spatial-GCN captures neighborhood heat interactions through graph-based spatial-temporal modeling<sup>76,77</sup>, while SHAP<sup>78</sup> provides interpretability through game-theoretic explanations.

Data sources expand significantly beyond basic morphology and meteorology to include social vulnerability data, recognizing that heat exposure disproportionately affects vulnerable populations<sup>79,80</sup>, as well as dynamic urban factors. This transformation elevates a narrow focus on accuracy into a comprehensive planning tool that generates vulnerability maps and actionable design guidelines, shifting the focus from technical performance to practical urban interventions (see Supplementary Material 6, Table S7).

#### Case C: technology-driven research—AI applications in disaster management

A GRU-CNN architecture for urban applications is reframed by our system, which identifies Urban Resilience Theory ( $\sigma = 0.94$ ) as the primary framework<sup>81,82</sup>. This theory emphasizes systemic reconfiguration and collective agency—principles that guide our multi-method integration. Based on this framework, the system recommends augmenting GRU-CNN with three complementary approaches. Agent-based modeling (ABM) enables population response simulation<sup>83</sup>, Network Analysis captures infrastructure interdependencies<sup>84</sup>, and Reinforcement Learning supports adaptive strategies<sup>85</sup>. To support these integrated methods, data requirements expand from basic time-series and hazard maps to include infrastructure networks, social communication patterns, and historical event data. The integration of GRU-CNN architecture has shown promising results in urban environmental monitoring<sup>86</sup>.

Beyond technical enhancements, this transformation fundamentally redefines the technology’s role in urban contexts. The system shifts from a

capability demonstration to a community-centered resilience platform. More importantly, residents become active participants in resilience-building rather than passive recipients of alerts. Accordingly, system functions extend to cascading failure prediction<sup>87</sup>, evacuation modeling, and resource optimization. Table 8 illustrates this comprehensive transformation (see Supplementary Material 7, Table S8 and Supplementary Material 6, Table S7).

### Discussion

Only 1.16% of papers in urban planning and safety AI applications demonstrate explicit theoretical integration by citing established urban theories (see Fig. 4). This striking gap is not merely a citation oversight but reflects three deeply interconnected systemic barriers that perpetuate the theory-practice disconnect. First, cross-domain expertise remains critically scarce. Among 948 papers with identifiable author backgrounds, merely 11.7% of first authors possess knowledge spanning both AI/computing and

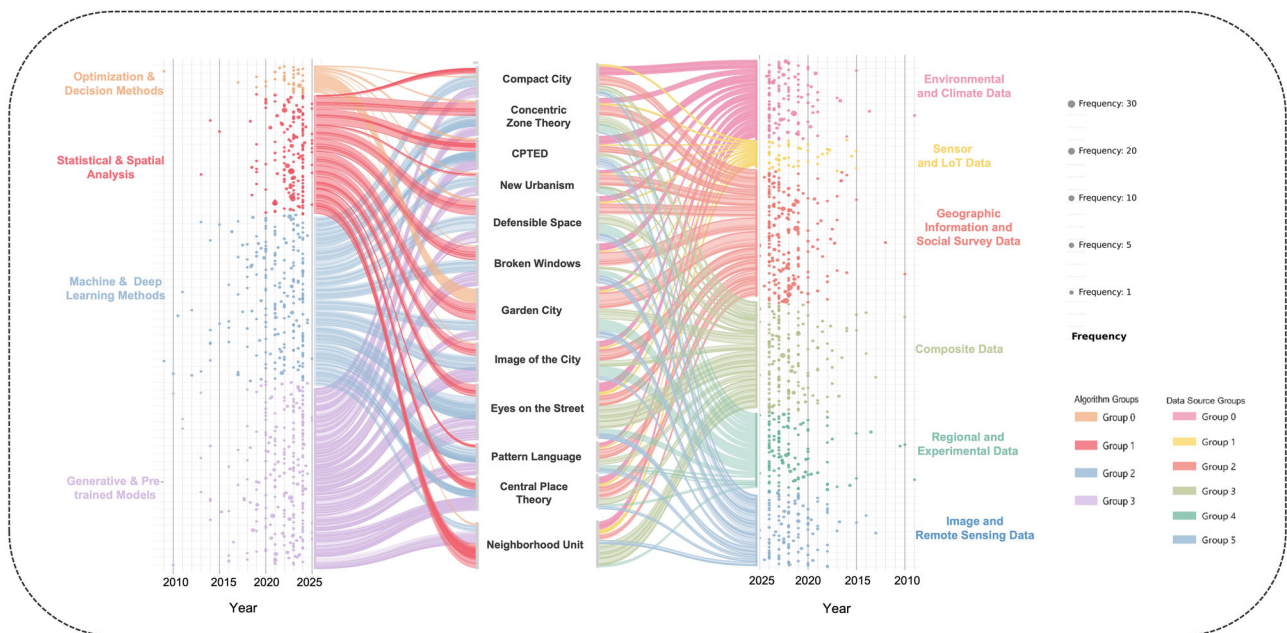
urban domains. Furthermore, only 33.3% of research teams include members from multiple disciplines<sup>35</sup>. This expertise gap creates research environments where algorithmic capabilities dominate problem framing, often at the expense of urban theory considerations. Second, and closely related, data availability increasingly dictates research priorities rather than theoretical importance.

Traditional Urban and Geospatial Data (Groups 0–1) dominate at 68.3%, while emerging data sources remain largely untapped at 19.2%—substantially below the theory-derived demand, where all 46 theories require at least one such source. This imbalance creates what we term “data opportunism”—a phenomenon where research questions arise from measurability rather than theoretical significance<sup>88</sup>. In other words, readily accessible datasets determine which urban problems researchers address, rather than theories driving data collection. For instance, abundant sensor data shapes traffic optimisation studies, while harder-to-measure social equity dimensions receive less attention despite their theoretical significance. Third, specialised research communities develop implicit domain expertise that remains invisible to cross-disciplinary evaluation. Domain expertise becomes highly context-dependent, acquired through sustained professional practice rather than formal articulation<sup>89,90</sup>. Travel behaviour researchers, for example, may apply decades of accumulated knowledge about built environment-transport relationships without explicitly citing spatial interaction theories. Such sophisticated theoretical understanding remains invisible to automated detection or evaluation by researchers from other disciplines<sup>91</sup>.

These three barriers do not operate in isolation; rather, they collectively create a self-reinforcing cycle that significantly impedes interdisciplinary communication. Knowledge structures across disciplines exhibit substantial heterogeneity, leading to ambiguity and misunderstandings when computer scientists collaborate with urban planners<sup>92,93</sup>. When domain expertise operates implicitly, the theoretical foundations underlying apparently technical work remain obscured<sup>94</sup>. Indeed, our team composition analysis reveals a counterintuitive pattern that quantifies this challenge. Pure urban planning teams score highest on theoretical engagement (0.65), while interdisciplinary teams achieve only moderate scores (0.39). This paradox demonstrates that theoretical knowledge does not automatically transfer

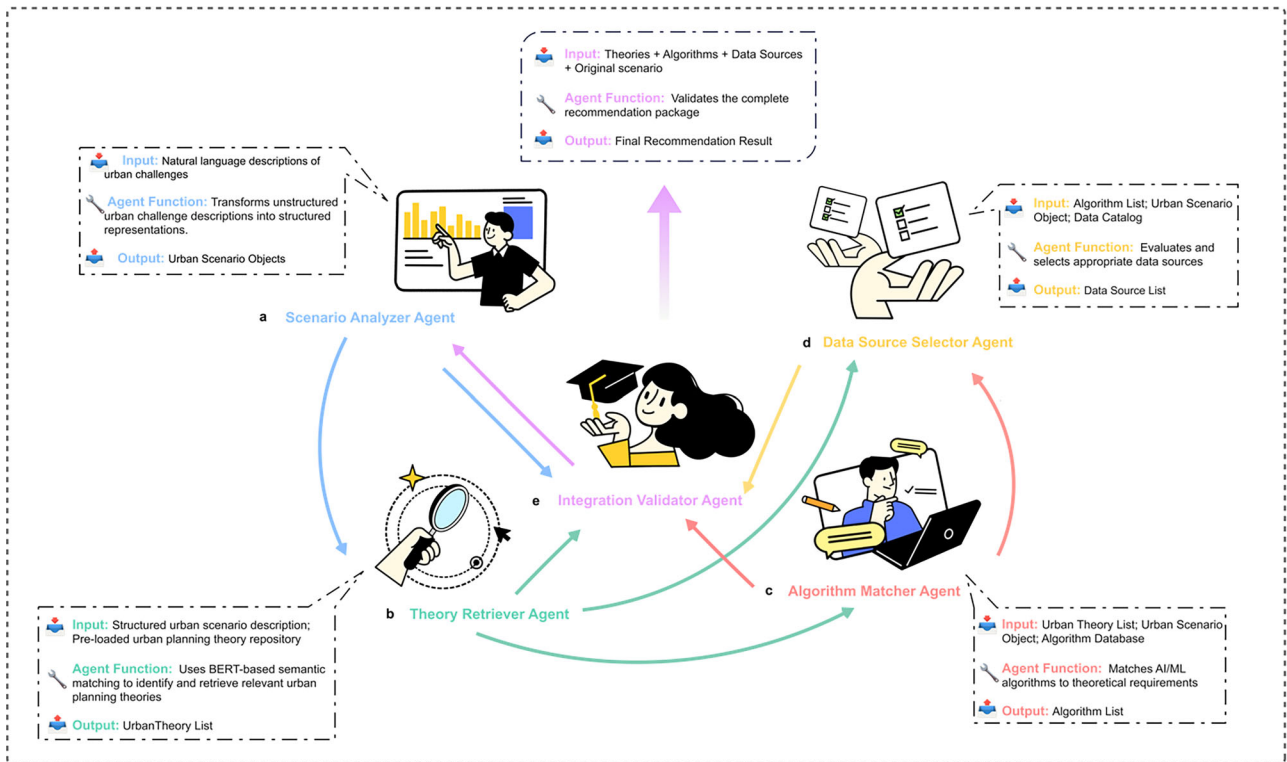
**Table 5 | Most influential theories in urban planning and safety fields**

Theory cluster	Theory name	First appearance
Design	Garden City	1898
	Neighborhood Unit	1929
	New Urbanism	1980s
Spatial structure	Central Place Theory	1933
	Concentric Zone Model	1925
	Compact City	1973
Safety	Crime Prevention Through Environmental Design (CPTED)	1971
	Defensible Space	1972
	Eyes on the Street	1961
Perception	Image of the City	1960
	Pattern Language	1977



**Fig. 4 | Theory-algorithm-data mapping framework (instance-level).** Based on 1123 papers (2008–2025) yielding 2847 algorithm instances. **Left:** Algorithm groups (G0: Optimization; G1: Statistical/Spatial; G2: ML/DL; G3: Generative) with instance frequencies. **Center:** Eleven representative theories from 46-theory knowledge base.

**Right:** Data source categories with usage distribution. **Flow interpretation:** Thickness represents proportion of algorithm instances, not papers. Example: CPTED’s G2 flow = 36% of instances (32/89), while 72% of papers use G2 (Table S11). Multi-method papers contribute to multiple flows. See Table S11 for paper-level statistics.



**Fig. 5 | Multi-agent recommendation system architecture for urban scenario analysis (implemented as publicly accessible web application at [https://llm-based-multi-agent-recommendation-system-eljs4zovvgefco8gzj.streamlit.app/]).** The system comprises five specialized agents: **a** Scenario Analyzer Agent transforms unstructured urban challenge descriptions into structured representations; **b** Theory

Retriever Agent uses BERT-based semantic matching to identify relevant urban planning theories from a pre-loaded repository; **c** Algorithm Matcher Agent matches AI/ML algorithms to theoretical requirements from an algorithm database; **d** Data Source Selector Agent evaluates and selects appropriate data sources; and **e** Integration Validator Agent validates the complete recommendation package.

**Table 6 | Key metrics for theory-driven recommendations**

Metric	Description, interpretation, and low-score guidance
<i>Scenario Assessment</i>	
$\xi(S) \in [0, 1]$	<b>Scenario Complexity:</b> Quantifies challenge complexity across 7 dimensions. $\xi > 0.7$ : complex (multi-method required); $0.5 \leq \xi \leq 0.7$ : moderate; $\xi < 0.5$ : simple. <i>Low-score signal:</i> $\xi < 0.3$ suggests problem may not warrant AI complexity – consider descriptive statistics, GIS mapping, or expert consultation.
<i>Theory-Practice Alignment</i>	
$\sigma \in [0, 1]$	<b>Theory-Scenario Similarity:</b> Semantic alignment via BERT cosine similarity. $\sigma > 0.85$ : strong; $0.70 \leq \sigma \leq 0.85$ : moderate; $\sigma < 0.70$ : weak. <i>Low-score signal:</i> Persistent $\sigma < 0.70$ across all theories indicates problem may require novel theoretical development or non-computational approaches (participatory planning, stakeholder workshops, qualitative research).
<i>Implementation Feasibility</i>	
$cap(A, r) \in [0, 1]$	<b>Algorithm Capability:</b> Algorithm $A$ 's ability to fulfill requirement $r$ . $cap > 0.85$ : strong match; $0.70 \leq cap \leq 0.85$ : augmentation needed; $cap < 0.70$ : mismatch. <i>Low-score signal:</i> No algorithm achieving $cap \geq 0.70$ suggests technical solutions may be premature – consider foundational data collection, pilot studies, or capacity building.
$Q(D) \in [0, 1]$	<b>Data Source Quality:</b> Evaluates source $D$ across 6 dimensions. $Q > 0.80$ : high priority; $0.65 \leq Q \leq 0.80$ : acceptable; $Q < 0.65$ : requires enhancement. <i>Low-score signal:</i> Systematic $Q < 0.65$ indicates data infrastructure gaps requiring institutional investment before AI deployment.
<i>Solution Validation</i>	
$R \in [0, 1]$	<b>Integration Robustness:</b> Proportion of perturbed scenarios (Monte Carlo, $N = 1000$ ) maintaining performance above $\tau = 0.70$ . Approval requires $R > 0.85$ . <i>Low-score signal:</i> $R < 0.85$ indicates solution instability – consider adaptive management, phased implementation, or human-in-the-loop systems instead of fully automated approaches.

All metrics normalized to [0, 1]; higher values indicate stronger performance. Low-score signals should prompt reconsideration of whether AI is appropriate for the specific urban challenge.

across disciplinary boundaries—diverse teams require intentional strategies to bridge epistemological gaps. Combined, these factors produce a self-reinforcing cycle: scarce interdisciplinary expertise, data-driven research design, and invisible theoretical communication collectively ensure that technical accessibility determines research trajectories rather than theoretical importance<sup>95</sup>. As a result, the field consequently exhibits technical

sophistication while remaining theoretically impoverished, failing to engage with complex socio-spatial realities that urban theories have long addressed and ultimately limiting both real-world impact and broader relevance<sup>96</sup>.

Building on these structural barriers, we now examine how they manifest in researchers' motivations and methodological choices. Two interconnected patterns explain why 67.8% of articles pursue technology-

driven research while only 1.3% address concrete urban problems (Table 2). First, data accessibility creates a self-reinforcing cycle that privileges certain research directions. Traditional and Geospatial Data dominate at 68.3% (exceeding theory-derived demand of 58.7%), while Urban-generated Data accounts for only 19.2% (Fig. 1b), as readily available datasets enable shorter research cycles<sup>97,98</sup>. In contrast, emerging sources—social media, mobile trajectories, crowdsourced platforms—require institutional partnerships, ethical approvals, and extended timelines that exceed standard project cycles<sup>99,100</sup>. Consequently, analytical frameworks favor static administrative data while dynamic user-generated data are less frequently incorporated.

Second, the 2020–2025 period intensified this technology-first orientation through unprecedented adoption speed. During these years, 60% of all applications emerged (Fig. 3b), with algorithms like CLIP and GPT-3 achieving the same-year development and application. Notably, algorithm selection criteria reveal misaligned priorities: performance (37.4%) and novelty (26.0%) drastically outweigh domain applicability (8.2%). This misalignment leads to methodological mismatches—CNNs designed for image recognition process non-visual planning data, while NLP models analyze numerical safety metrics. Moreover, Group 3 algorithms demonstrate mean latency of merely 13.2 years compared to 48.1 years for traditional methods (Table 4), reflecting a rush to adopt cutting-edge techniques regardless of their appropriateness. These combined patterns create problematic selection effects favoring researchers skilled in rapid implementation over those focused on urban problem-solving<sup>101</sup>. In practice, this means researchers optimize for publishability rather than urban impact. They pursue datasets already available rather than those theoretically necessary. They adopt algorithms based on recency rather than appropriateness. This self-reinforcing cycle ultimately produces a field rich in technical demonstrations but poor in genuine urban problem-solving, where technical accessibility determines research trajectories rather than theoretical importance.

Having diagnosed the problem and its causes, we now turn to evidence that theoretical integration can fundamentally transform research outcomes. Our case studies reveal three fundamental transformations when theoretical grounding guides AI research, demonstrating practical pathways to break the technology-push cycle. First, theory expands the scope of solutions from narrow optimization to systemic intervention. Case A's food

waste routing initially minimized costs ( $\min \sum c_{ij}x_{ij}$ ), but it ignored food insecurity—a critical urban challenge. When Urban Metabolism and Environmental Justice theories were integrated, this narrow objective was reframed into a comprehensive food security framework. Specifically, data sources expanded from routes to vulnerability indices (2 → 4 categories), algorithms evolved from single-method to integrated frameworks (MOGA + GNN + LSTM), and metrics shifted from efficiency-only to multi-dimensional measures including equity and emissions<sup>102,103</sup>. Second, theory transforms technical gains into actionable planning tools. Case B initially pursued accuracy enhancement ( $R^2: 0.8 \rightarrow 0.95$ ) until Urban Climate Theory introduced physical constraints via PINN. When combined with Compact City and Sustainable Design theories, outcomes evolved from temperature predictions to vulnerability-aware planning guidelines with interpretable interventions—theory added not accuracy but actionability<sup>104,105</sup>. Third, the theory reverses the innovation direction from technology-push to need-pull. Case C's GRU-CNN seeking applications became a community resilience platform through Urban Resilience Theory. This transformation involved architecture expansion (adding ABM, Network Analysis, RL), data diversification (2 → 4 types), and critically, residents shifting from alert recipients to resilience co-creators<sup>106,107</sup>. These transformations share a common pattern: theoretical grounding fundamentally redefines problems rather than merely improving solutions. Across all cases, robustness scores exceeded 0.85, confirming that theoretical frameworks enhance rather than constrain performance. More importantly, theory-driven approaches generated value beyond technical metrics—Case A addressed food equity alongside efficiency, Case B produced planning prescriptions rather than predictions, and Case C enabled community participation instead of passive monitoring. This paradigmatic shift demonstrates that integrating urban theories into AI development produces outcomes simultaneously more comprehensive, actionable, and aligned with genuine urban needs than purely technical approaches<sup>88,108</sup>.

While our study provides comprehensive evidence of the theory-practice disconnect, several limitations warrant acknowledgment. First, our 1.16% finding captures only explicit theory citations; semantic density analysis reveals that 87.1% of high-density papers demonstrate domain knowledge invisible to citation-based detection, suggesting our measurement prioritizes transparency over capturing tacit expertise. Second, our coverage has geographical boundaries—the literature search may miss non-English publications, and the 46-theory knowledge base underrepresents Global South perspectives<sup>109</sup>. Third, the recommendation system operates conditionally, assuming users have determined AI approaches warrant exploration. Many urban problems involving community trust or political negotiation may be ill-suited to technical optimization; built-in safeguards ( $\sigma < 0.70, R < 0.85$ ) can signal such cases, but recommendations indicate how AI *could* be applied, not that it *should* be. Fourth, technical constraints remain: the system's keyword-based classification of research objectives may inadequately capture nuanced analytical goals, BERT classifier accuracy (80.85%) suggests potential misclassifications, and case studies represent conceptual demonstrations rather than real-world implementations<sup>110</sup>.

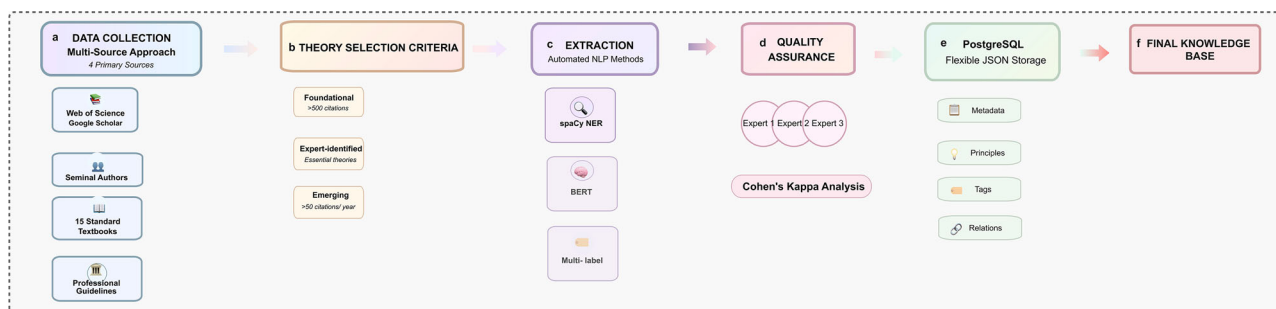
Despite these limitations, our research makes three significant contributions that advance both theoretical understanding and practical implementation. Empirically, we quantify barriers to theoretical integration (1.16% explicit rate, 47.3% technology-driven research) across

**Table 7 | From technical optimization to systemic solutions in urban food waste management**

Dimension	Original approach	Theory-driven solution
Problem Focus	Cost minimization	Equity and sustainability
Theory	None	Urban metabolism and environmental justice
Algorithm	Continuous approximation	MOEA/D + GNN + LSTM
Data	Routes and locations	Multi-domain: social, environmental, supply chain
Metrics	Cost and efficiency	Cost, equity, emissions, accessibility
Output	Optimized routes	Integrated food security framework

**Table 8 | From technology showcase to community resilience platform**

Dimension	Original approach	Theory-driven solution
Problem Focus	Demonstrate GRU-CNN capability	Build community disaster resilience
Theory	None	Urban Resilience + Disaster Risk Reduction + Socio-Ecological Systems
Algorithm	GRU-CNN only	GRU-CNN + ABM + Network Analysis + RL
Data	Time-series, hazard maps	Multi-domain: hazards, infrastructure, social communication
Metrics	Technical accuracy, alert speed	Resilience indices, evacuation efficiency, community participation
Output	Risk alerts to recipients	Resilience platform with active community co-creation



**Fig. 6 | Workflow for constructing the urban theory knowledge base.** **a** Data collection from four primary sources spanning 1960–2024. **b** Theory selection using composite criteria. **c** Automated extraction using NLP methods. **d** Three-expert

validation with inter-rater reliability assessment. **e** PostgreSQL implementation with flexible JSON storage. **f** Final knowledge base supporting semantic similarity matching for theory-to-practice recommendations.

1123 papers, establishing measurable baselines for tracking progress<sup>3,111</sup>. Methodologically, our multi-agent system demonstrates how LLMs can facilitate interdisciplinary knowledge integration, addressing expertise gaps affecting 88.3% of research teams<sup>112,113</sup>. Theoretically, we extract 127 computational principles from urban theories, revealing their inherent algorithmic compatibility and challenging assumptions about incompatibility between traditional wisdom and computational methods. Looking forward, future research should pursue three interconnected directions to build on these foundations. First, expand theoretical coverage by incorporating Global South planning theories, indigenous urban knowledge, and emerging post-pandemic frameworks<sup>8,17</sup> while integrating diverse research databases beyond traditional academic sources. Concretely, this could involve partnering with institutions like UN-Habitat to establish Global South theory databases and conducting systematic reviews in non-English journals. Second, validate real-world impact through comparative studies of theory-grounded versus purely technical AI applications, tracking implementation success, sustainability outcomes, and community acceptance<sup>114</sup>. Such validation studies could follow pilot projects over multi-year periods to assess long-term adoption and adaptation patterns. Third, foster institutional change by working with funding agencies, journals, and educational institutions to develop evaluation criteria and training programs that value theory-practice integration<sup>110,115</sup>. This might include establishing new review standards that explicitly assess theoretical grounding, creating interdisciplinary PhD programs, and incentivizing cross-sector collaborations. As cities confront unprecedented sustainability challenges, these efforts can ensure AI amplifies rather than replaces the accumulated wisdom of how cities actually work—and how they might work better for all their inhabitants.

**Methods**

Our methodological framework integrates diagnostic analysis with solution development. We first construct dual knowledge bases of urban theories and AI research papers, then measure theoretical integration and classify research motivations to quantify the theory-practice disconnect. We then develop a theory-algorithm-data mapping framework and a multi-agent recommendation system that bridges theoretical principles with AI implementations. Three case studies demonstrate practical applications across distinct research motivation patterns. We detail each component below.

**Data on traditional urban theories**

We construct a comprehensive knowledge base of urban planning and safety theories through a streamlined extraction process (Fig. 6), building on recent advances in theory formalization<sup>35,116</sup>. Theory selection integrates three complementary criteria (see Supplementary Material 5): foundational theories with over 500 citations, emerging theories with over 50 citations annually, and expert-identified essential theories, regardless of metrics<sup>117</sup>—ensuring comprehensive coverage across established and culturally diverse perspectives. Guided by these criteria, our corpus includes: (1) highly-cited

papers from Web of Science (WoS) and Scopus and Google Scholar using queries (“urban planning” OR “urban safety”) AND (“theory” OR “framework”), (2) works by seminal authors (Jacobs, Lynch, Newman, etc.) from 1960 to 2024, (3) 15 standard urban planning textbooks, and (4) professional guidelines from organizations like the American Planning Association (see Supplementary Table 6).

NLP implementation employs spaCy’s en\_core\_web\_lg model (v3.4) for Named Entity Recognition to identify theory names through capitalized noun phrases preceded by theory-indicating terms and author-attributed concepts. Core principle extraction combines BERT embeddings (bert-base-uncased) with TF-IDF scoring, ranking sentences by combined metrics, and applying dependency parsing for grammatical completeness. A multi-label tagging system<sup>118,119</sup> preserves theoretical complexity by maintaining associations across spatial, social, safety, and economic dimensions with relevance weights (0–1 scale). Complete taxonomy with semantic keyword clusters for each dimension is detailed in Supplementary Material 1, Table S1. The PostgreSQL-based knowledge base with flexible JSON storage enables semantic similarity matching between urban challenges and relevant theories<sup>120,121</sup>.

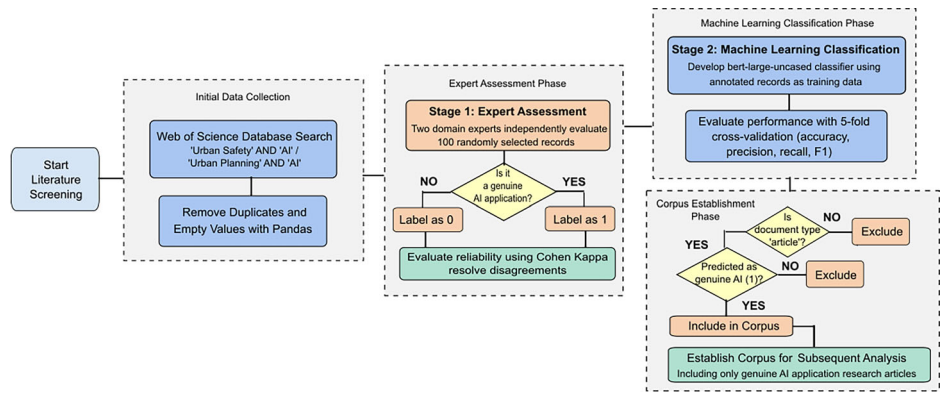
**Data on urban research literature**

To capture the current landscape of AI applications in urban planning and safety domains, we conduct a comprehensive literature search through the WoS and Scopus databases. Our search strategy employs two primary queries—“Urban Safety” AND “AI” and “Urban Planning” AND “AI”—without temporal restrictions to ensure comprehensive coverage of the field’s evolution. This unrestricted temporal coverage aligns with recent systematic reviews of urban AI applications<sup>16,122,123</sup>, enabling us to trace the full trajectory of AI adoption in urban planning and safety contexts. The initial search retrieved 2797 papers, which underwent systematic screening to establish the final corpus of 1123 validated AI applications (Fig. 7).

This keyword-based approach introduces systematic temporal bias that warrants acknowledgment. The term “AI” gained widespread use primarily after 2010, particularly following deep learning breakthroughs that catalyzed the adoption of AI terminology across disciplines<sup>43,46</sup>. Consequently, our corpus excludes earlier algorithmic applications in urban research—including regression models, optimization techniques, spatial statistics, and early neural networks—that were not explicitly labeled as “AI”<sup>44,45</sup>. Our analysis, therefore, captures the phenomenon of AI adoption as disciplinary framing rather than a comprehensive computational history, documenting when established methods became repackaged under AI terminology. This intentional scope limitation enables our core contribution: revealing how contemporary AI-labeled research exhibits a technology-first orientation regardless of algorithmic age.

A rigorous two-stage screening process combines human expertise with machine learning capabilities to evaluate the collected literature (Fig. 7). Domain experts first assess each paper’s relevance to genuine AI urban applications, after which a BERT-based classifier performs automated screening with NLP pipeline configurations detailed in Supplementary

**Fig. 7 | Literature Screening Process.** Literature screening workflow illustrating the progression from initial data collection through expert assessment and machine learning classification to final corpus establishment.



Material 2, Table S3. The BERT-large-uncased model was fine-tuned with learning rate 2e-5, batch size 16, training epochs 5, using AdamW optimizer with binary cross-entropy loss, achieving 80.85% accuracy (see Supplementary Material 8, Table S10) across 5-fold cross-validation (mean accuracy:  $0.8085 \pm 0.0467$ , precision:  $0.7825 \pm 0.0841$ , recall:  $0.7188 \pm 0.0470$ ). Building on established methods for automated article classification in systematic reviews<sup>24</sup>, our approach incorporates domain-specific adaptations<sup>124,125</sup>. Recent advances in machine learning for systematic reviews have demonstrated that such hybrid approaches can significantly reduce workload while maintaining high accuracy<sup>126,127</sup>.

Following the screening process, a novel dual LLM expert framework systematically analyzes the refined corpus (Fig. 8). Two independent LLMs work in parallel to extract AI algorithms and data sources from each paper, with human experts subsequently validating the outputs to ensure accuracy. Drawing inspiration from recent advances in generative information extraction<sup>128</sup> and LLM applications in computational social science<sup>129</sup>, this dual-model architecture addresses single-model biases and enhances extraction reliability-critical considerations highlighted in recent LLM security and privacy literature<sup>130</sup>. Semi-automated validation processes balance automation efficiency with accuracy requirements, creating a robust extraction pipeline that ensures replicability (see Supplementary Material 8, Table S9).

**Algorithm and data source grouping**

To organize the extracted algorithms and data sources into interpretable categories, we implement a four-stage transformation pipeline (Table 9, Fig. 1). First, we aggregate contextual information by concatenating descriptive text from all papers mentioning each item-for instance, “Random Forest” contexts from 84 papers produced approximately 500 tokens describing ensemble learning principles and urban applications. Building on this textual foundation, we employ sentence-transformers (all-mpnet-base-v2)<sup>131</sup> to encode all contexts into 768-dimensional embeddings. Critically, BERT<sup>132</sup> operates on aggregated text rather than directly on LLM JSON outputs-LLMs extract structured metadata while BERT encodes semantic meanings for clustering. To enable visualization and clustering, we then reduce embeddings to 2D coordinates via t-distributed Stochastic Neighbor Embedding<sup>27</sup> using standard parameters (perplexity = 30, learning rate = 200, iterations = 1000, random state = 42). Finally, DBSCAN<sup>133</sup> with  $\epsilon = 0.5$  and min\_samples = 5 identify initial clusters, which three domain experts-representing urban planning, computer science, and data science perspectives-iteratively refine through four consensus rounds based on methodological coherence and domain interpretability. This hybrid process produce four algorithm groups (G0–G3) and six data source groups (G0–G5, Table 1). External validation involve two independent researchers classifying stratified samples, with detailed agreement metrics and cluster quality measures documented in Supplementary Material 9 alongside complete technical specifications.

**Validation of theoretical integration measurement**

Our explicit citation-based measurement prioritizes transparency and replicability, yet may not capture domain knowledge operating without explicit citations<sup>89</sup>. To evaluate whether this approach underestimates implicit theoretical engagement, we implement a three-method validation protocol on a stratified subsample ( $n = 200$ ). The sample oversamples papers with explicit citations (35 papers, 17.5% vs. 1.16% in the full corpus) to enable robust statistical comparison while maintaining diversity across algorithm types, publication years, and team compositions.

We apply three complementary approaches. First, LLM-based classification using GPT-4-turbo (temperature = 0.2) analyzes introduction sections to test whether automated semantic analysis can detect theoretical engagement beyond explicit citations, with accuracy assessed against expert coding using Cohen’s kappa. Second, team composition analysis quantifies disciplinary diversity using Shannon’s entropy  $H_{norm} = -\sum_{i=1}^k p_i \ln(p_i) / \ln(k)$  and tests its correlation with theory-driven intensity scores (0–1 scale) assigned by an interdisciplinary expert panel ( $n = 3$ ). Third, semantic theoretical density (STD) measures implicit engagement through proximity to an 847-term theoretical lexicon derived from our 46-theory knowledge base. Using sentence-transformers (all-mpnet-base-v2)<sup>131</sup>, we compute:

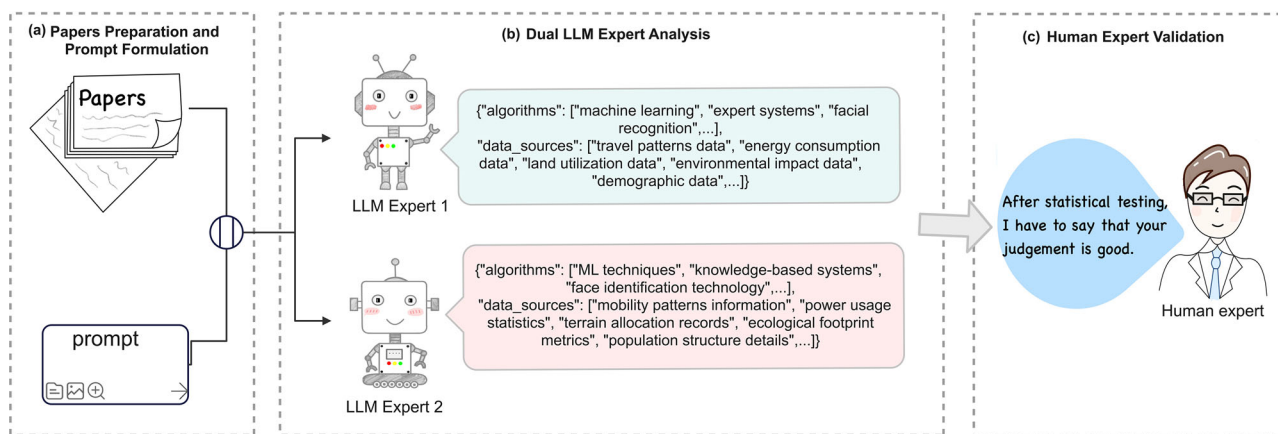
$$STD_i = \frac{1}{N} \sum_{j=1}^N \max_{k \in L} [\text{sim}(s_{ij}, t_k) \times w_k] \tag{1}$$

where  $s_{ij}$  are sentence embeddings,  $t_k$  are lexicon term embeddings,  $w_k$  are category weights (ranging from 1.0 for canonical theory constructs to 0.6 for causal mechanisms), and  $\text{sim}(\cdot)$  is cosine similarity. Lexicon construction, hierarchical weighting scheme, and validation protocol are detailed in Supplementary Material S10.

Measurement convergence is assessed through: (1) pairwise correlations between explicit citation (binary), semantic density (continuous), and expert assessment (continuous); (2) Steiger’s Z-test comparing correlation strengths; and (3) cross-team consistency examining rank ordering across team types (pure CS, pure Planning, interdisciplinary). This protocol addresses the challenge of balancing transparent measurement with sensitivity to context-dependent expertise<sup>134,135</sup>. All analyses are conducted in R 4.3.1 ( $\alpha = 0.05$ , two-tailed), with code available in our public repository.

**Research motivation classification**

To uncover the driving forces behind urban AI research, we employ automated pattern matching on abstracts, titles, and introductions<sup>136,137</sup> to classify papers into four categories: (1) problem-driven-studies initiated by specific real-world crises (e.g., “urban flooding occurred”); (2) method-driven-research motivated by limitations of current solutions (e.g., “existing methods are insufficiently accurate”); (3) technology-driven-studies exploring novel technological applications (e.g., “deep learning can be



**Fig. 8 | Dual LLM expert analysis framework for urban safety AI research.** The three-stage process encompasses: **a** paper preparation and prompt engineering, **b** parallel extraction of algorithms and data sources by two independent LLM experts, and **c** human expert validation and reconciliation of results.

**Table 9 | Seven-stage pipeline: from LLM extraction to validated grouping**

Stage	Process	Key outputs
<i>Phase I: Computational Processing</i>		
1	LLM Extraction	Dual models (GPT-4, Claude-3) extract items
2	Aggregation	Concatenate contexts per item
3	BERT Encoding	sentence-transformers embedding
4	t-SNE	Reduce to 2D (perplexity = 30, iter = 1000)
5	DBSCAN	Density clustering ( $\epsilon = 0.5$ , min_samples = 5)
<i>Phase II: Expert Refinement and Validation</i>		
6	Refinement	3 experts, 4 consensus rounds
7	Validation	Independent coding (stratified sample)

Complete specifications and validation results in Supplementary Material S9.

applied to flood prediction”); and (4) general needs-driven-research addressing broad urban requirements without specific triggers. This taxonomy extends beyond traditional technology-driven versus problem-driven dichotomies<sup>138</sup>, capturing nuanced distinctions between crisis-responsive and solution-improvement research.

Classification employs case-insensitive regex matching with contextual analysis (see Supplementary Material 3). Problem-driven patterns include crisis indicators (“occurred,” “emerged”) and temporal urgency markers; method-driven patterns capture performance critiques (“low accuracy,” “computationally expensive”) and comparative language; technology-driven patterns identify innovation phrases (“recent advances in,” “newly developed”) and technology-subject grammatical structures. Manual validation on 300 sampled papers achieved 85% inter-rater agreement (see Supplementary Material 3). For papers exhibiting multiple patterns, a semantic dominance hierarchy (Problem > Method > Technology > General) determines classification based on priority zone analysis of titles and opening sentences. Papers without clear matches are assigned to “Unclear/Mixed” (147 papers, 13.1%) rather than forced into categories. Complete protocols appear in Supplementary Material 3. This framework aligns with recent advances in automated systematic review classification<sup>139,140</sup> and has been validated against manual coding using established text classification methods<sup>141,142</sup>.

**Theory-algorithm-data mapping construction**

We develop a multi-method mapping framework that systematically connects urban theories, AI algorithms, and data sources by combining computational linguistics with knowledge engineering approaches<sup>143,144</sup>. Building on recent advances in domain-specific knowledge graph

construction<sup>145,146</sup>, our framework adapts these methods for urban planning contexts through three core processes. First, NLP techniques extract computational requirements from theoretical principles<sup>147,148</sup>; for instance, CPTED’s “maintain clear sightlines” translates to “visibility analysis” requirements. Second, algorithm capability profiles emerge from multiple sources, leveraging LLM-based knowledge extraction<sup>149</sup> with expert validation. Third, probabilistic co-occurrence models<sup>150</sup> identify data requirements by detecting algorithm-data source mention patterns within urban planning literature, thereby revealing implicit theory-implementation relationships<sup>151,152</sup>.

The Stanford CoreNLP pipeline processes theory texts using PTBTokenizer, statistical sentence splitter, maximum entropy POS tagger, rule-based lemmatizer, SR parser, and neural dependency parser. Algorithm capability assessment evaluates seven dimensions (see Supplementary Material 4, Table S4): spatial analysis, temporal analysis, pattern recognition, prediction, classification, optimization, and real-time processing (scores 0–1). Association rule mining employs the Apriori algorithm with a minimum support of 0.05, a minimum confidence of 0.7, a maximum itemset size of 4, a lift threshold of >1.2, and a window size of 5 sentences. Pointwise Mutual Information quantifies theory-algorithm co-occurrence:  $PMI(\text{theory}, \text{algorithm}) = \log[P(\text{theory}, \text{algorithm}) / (P(\text{theory}) \times P(\text{algorithm}))]$ . To ensure mapping quality, a three-stage validation protocol combines statistical measures, expert annotation, and logical consistency checks<sup>153,154</sup>, while acknowledging both technical and social assessment dimensions inherent to urban planning’s interdisciplinary nature<sup>155</sup>. Final compatibility scores combine PMI (30%), expert scores (50%), and consistency checks (20%) through weighted averaging based on expert consensus<sup>156</sup>, creating a robust bridge between theoretical principles and practical AI implementations.

**Table 10 | Scenario complexity scoring specifications**

Dimension	Measurement	Weight ( $w_i$ )	Normalization
$c_1$ : Domain complexity	Unique technical domains	0.20	$\min(n/5, 1)$
$c_2$ : Objective multiplicity	Competing objectives	0.18	$\min(n/4, 1)$
$c_3$ : Constraint severity	Hard constraints count	0.15	$\min(n/4, 1)$
$c_4$ : Stakeholder diversity	Shannon entropy	0.15	$H/\ln(6)$
$c_5$ : Temporal scope	Planning horizon (years)	0.12	$\min(T/10, 1)$
$c_6$ : Spatial boundaries	Categorical scale	0.12	See note <sup>a</sup>
$c_7$ : Data heterogeneity	Distinct data types	0.08	$\min(n/6, 1)$

<sup>a</sup>Block = 0.2, Neighborhood = 0.4, City = 0.6, Region = 0.8, Multi-region = 1.0. Weights calibrated via Delphi method (12 experts, 3 rounds, Kendall's  $W = 0.82$ ).

For empirical analysis of the resulting mappings, we employ two complementary statistical approaches: *instance-level distribution* analyzes the proportion of algorithm instances across categories (Fig. 4), providing insights into implementation patterns, while *paper-level distribution* quantifies the proportion of papers employing each algorithm or data type (Supplementary Material 11, Table S11), revealing adoption patterns across the research community. This dual-level analysis enables a comprehensive understanding of both technical implementation choices and research practice trends.

**Theory-driven multi-agent recommendation system**

To bridge the theory-practice gap identified in our analysis, we propose a multi-agent recommendation system comprising five specialized agents that collaborate via asynchronous coordination<sup>157–159</sup>. Each agent contributes domain-specific expertise through directed information flow, generating theory-grounded recommendations.

According to system scope and entry conditions, the system is designed for users who have preliminarily determined that AI-based approaches merit exploration. It does not assess whether AI is the most appropriate intervention—a determination requiring consideration of stakeholder preferences, institutional capacity, and the fundamentally social nature of many urban problems. Rather, the system transforms the question from “what can this algorithm do?” to “what theoretical principles should guide this intervention if AI is pursued?” Users encountering persistent low-similarity scores ( $\sigma < 0.70$ ), capability gaps, or robustness failures ( $R < 0.85$ ) should interpret these as signals warranting reconsideration of AI appropriateness.

Our design relies on three assumptions: (i) theoretical requirements exhibit sufficient independence for additive aggregation ( $|\rho| < 0.35$  for 94% of requirement pairs); (ii) expert knowledge can be reliably elicited ( $ICC > 0.80$ ); and (iii) scenario perturbations follow bounded uniform distributions reflecting real-world planning uncertainties.

Scenario Analyzer Agent transforms unstructured challenge descriptions into structured representations across seven dimensions: domain, objectives, constraints, stakeholders, temporal scope, spatial boundaries, and data characteristics. Critically, the agent also extracts research objective type (prediction, explanation, optimization, or classification), which shapes algorithm matching independently of theoretical alignment. Scenario complexity is quantified as:

$$\xi(S) = \sum_{i=1}^7 w_i \cdot c_i(S) \tag{2}$$

where  $w_i$  are importance weights and  $c_i(S) \in [0, 1]$  are normalized measures. Table 10 provides specifications; complete operationalization with worked examples across three case studies appears in Supplementary Material 12, Table S12.

Theory Retriever Agent employs BERT-based semantic matching (bert-base-uncased) to identify applicable theories<sup>131,132</sup>. Theory-

scenario alignment is quantified as:

$$\sigma(T_i, S) = \cos(\mathbf{e}_{T_i}, \mathbf{e}_S) \tag{3}$$

where  $\mathbf{e}_{T_i}, \mathbf{e}_S \in \mathbb{R}^{768}$  are embeddings from the BERT [CLS] token. Theories with  $\sigma > 0.70$  are relevant;  $\sigma > 0.85$  indicates strong alignment. For complex challenges ( $\xi > 0.7$ ), the agent identifies complementary theory combinations<sup>160,161</sup>.

Algorithm Matcher Agent formulates selection as constrained optimization<sup>162,163</sup>. Given candidates  $\mathcal{A} = \{A_1, \dots, A_m\}$  and requirements  $\mathcal{R} = \{r_1, \dots, r_n\}$ :

$$A^* = \arg \max_{A \in \mathcal{A}} \left\{ \sum_{i=1}^n w_i \cdot \text{cap}(A, r_i) - \lambda \cdot \text{cost}(A) \right\} \tag{4}$$

subject to  $\sum w_i = 1, w_i \geq 0$ , and  $\text{cap}(A, r_i) \geq 0.70$  for critical requirements. Here,  $\lambda = 0.15$  balances cost-performance trade-offs, and  $\text{cost}(A) \in [0, 1]$  combines computational time (60%) and memory (40%). The capability function:

$$\text{cap}(A, r_i) = 0.5 \cdot \text{PMI}_{\text{norm}}(A, r_i) + 0.3 \cdot \text{Expert}(A, r_i) + 0.2 \cdot \text{Consist}(A, r_i) \tag{5}$$

where  $\text{PMI}_{\text{norm}}$  is normalized pointwise mutual information, Expert scores derive from 8 specialists ( $ICC = 0.84$ ), and Consist enforces logical consistency with mandatory thresholds for specific semantic tags (Supplementary Material 4, Table S5)<sup>164,165</sup>.

Data Source Selector Agent evaluates sources using multiplicative aggregation<sup>166,167</sup>:

$$Q(D_i) = \prod_{j=1}^6 q_j(D_i)^{w_j} \tag{6}$$

where weights are: relevance ( $w_1 = 0.25$ ), quality ( $w_2 = 0.20$ ), temporal coverage ( $w_3 = 0.18$ ), accessibility ( $w_4 = 0.15$ ), reliability ( $w_5 = 0.12$ ), and compatibility ( $w_6 = 0.10$ ). Critically, **relevance** ( $q_1$ ) is computed dynamically via BERT semantic similarity between data source descriptions and scenario requirements—identical sources receive different scores across urban challenges (Supplementary Material 12, Table S13). Accessibility scores range from 1.0 (public datasets) to  $<0.5$  (restricted sources). The system recommends data categories with documented precedent in urban research rather than hypothetical sources, with final selection requiring human validation.

Integration Validator Agent implements two-stage validation<sup>168,169</sup>. Monte Carlo simulations ( $N = 1000$ ) assess stability:

$$R = \frac{1}{N} \sum_{k=1}^N \mathbb{1}[\text{perf}(S_k) > \tau] \tag{7}$$

where  $\tau = 0.70$  is the performance threshold and  $R_{\min} = 0.85$  the approval threshold. The performance function:

$$\text{perf}(S) = 0.40 \cdot \text{Theory}_{\text{align}} + 0.30 \cdot \min\{\text{Alg}_{\text{cap}}, \text{Data}_{\text{qual}}\} + 0.30 \cdot \text{Expert}_{\text{asses}} \quad (8)$$

Each iteration applies bounded perturbations ( $\pm 0.15$ ) to 2–3 randomly selected dimensions, representing the 90th percentile of variations in 200 real-world projects. Solutions exceeding  $R_{\min}$  advance to human validation through uncertainty sampling<sup>170</sup>. Agents collaborate via structured protocols: initial analysis (Scenario  $\rightarrow$  All), theory query (Theory  $\rightarrow$  Algorithm), data requirements (Algorithm  $\rightarrow$  Data) validation (All  $\rightarrow$  Validator), and conflict resolution through iterative refinement.

BERT was fine-tuned on 1143 papers for 5 epochs (learning rate  $3 \times 10^{-5}$ ). Expert assessments used 7-point Likert scales; disagreements ( $\text{ICC} < 0.70$ ) triggered facilitated discussion. Monte Carlo simulations ran in Python 3.10 with NumPy 1.24 (seed = 42). Code and data are available in our public repository.

### Case study development

To demonstrate the practical application of our theory-driven recommendation system, we construct three representative case studies based on the distinct research motivation patterns identified in our corpus analysis: problem-driven, method-driven, and technology-driven. These cases are synthesized from archetypal research patterns observed across our analyzed corpus ( $n = 1123$ ), capturing essential characteristics of each motivation type while maintaining generalizability<sup>2</sup>. Case archetype development follows a systematic protocol to ensure representativeness: identifying recurring patterns for each motivation type, synthesizing composite scenarios incorporating common problem domains, reflecting typical algorithmic approaches, maintaining realistic constraints, and avoiding direct replication of published work. For each case, we simulate the system’s recommendation process by inputting the original research framing and generating theory-algorithm-data suggestions based on our established mapping framework. The transformations are evaluated across multiple dimensions, including theoretical grounding, solution comprehensiveness, and practical value generation<sup>123</sup>. Rather than serving as empirical validations, these proof-of-concept demonstrations illustrate the system’s capability to identify relevant theoretical frameworks and transform technical solutions into theory-informed, multi-dimensional approaches that generate value beyond algorithmic performance metrics<sup>171</sup>. This approach aligns with recent methodological advances in design science research for urban informatics<sup>172</sup> and theory-driven AI system evaluation (see Supplementary Material 6, Table S7).

### Data availability

**Data Availability:** All data supporting the findings of this study are accessible through the publicly available LLM-based multi-agent recommendation system at: <https://frp-try.com:63105/>. This system enables researchers, urban planners, and policymakers to input specific urban challenges and receive integrated recommendations for relevant theories, appropriate AI methods, and suitable data sources. To address data availability concerns, the system displays accessibility ratings (0–1 scale) for each recommended data source, links to public portals where applicable, and usage frequency statistics derived from our corpus analysis. Reviewers can directly test the recommendation logic using pre-loaded scenarios (e.g., “urban heat vulnerability assessment”) to observe how theoretical requirements translate to accessible data recommendations. Additionally, the underlying corpus of 1123 validated papers was compiled from the publicly available Web of Science and Scopus databases using the search queries described in the Methods section. Please note that the system uses a locally issued HTTPS certificate; browsers may display a security warning on first access, which can be safely dismissed. A comprehensive video demonstration of the full pipeline is provided in Supplementary Material 13.

### Code availability

The source code for the multi-agent recommendation system is publicly available through our dedicated server at: <https://frp-try.com:63105/>. This includes the BERT fine-tuning pipeline, the theory-algorithm-data mapping framework, and the Monte Carlo validation scripts used to generate the results reported in this study. To ensure stable long-term availability, the system has been migrated from the original Streamlit prototype to a dedicated server with Qwen as the LLM backbone. All computational analyses were implemented in Python 3.10 with NumPy 1.24, and specific package versions and configuration parameters are documented in the “Methods” section and Supplementary Materials 2 and 8 to facilitate full reproducibility.

Received: 17 July 2025; Accepted: 8 March 2026;

Published online: 23 March 2026

### References

1. Cugurullo, F. et al. The rise of AI urbanism in post-smart cities: a critical commentary on urban artificial intelligence. *Urban Stud.* **61**, 1–18 (2023).
2. Son, T. H. et al. Algorithmic urban planning for smart and sustainable development: systematic review of the literature. *Sustain. Cities Soc.* **94**, 104562 (2023).
3. Yigitcanlar, T., Agdas, D. & Degirmenci, K. Artificial intelligence in local governments: perceptions of city managers on prospects, constraints and choices. *AI Soc.* **38**, 1335–1349 (2023).
4. Caprotti, F. et al. Why does urban artificial intelligence (AI) matter for urban studies? Developing research directions in urban AI research. *Urban Geogr.* **45**, 883–894 (2024).
5. Musa, M., Rahman, T., Deb, N. & Rahman, P. Harnessing artificial intelligence for sustainable urban development: advancing the three zeros method through innovation and infrastructure. *Sci. Rep.* **15**, 23673 (2025).
6. Palmi, O. & Cugurullo, F. Design culture for sustainable urban artificial intelligence: Bruno Latour and the search for a different AI urbanism. *Ethics Inf. Technol.* **26**, 1–12 (2024).
7. Jacobs, J. *The Death and Life of Great American Cities* (Random House, 1961).
8. Bibri, S. E., Alexandre, A., Sharifi, A. & Krogstie, J. Environmentally sustainable smart cities and their converging AI, IoT, and big data technologies and solutions: an integrated approach to an extensive literature review. *Energy Inform.* **6**, 9 (2023).
9. Batty, M. Artificial intelligence and smart cities. *Environ. Plan. B Urban Anal. City Sci.* **45**, 3–6 (2018).
10. Hollander, J., Hahn, S. & Reed, M. The ethical concerns of artificial intelligence in urban planning. *J. Am. Plan. Assoc.* **90**, 1–15 (2024).
11. Caprotti, F., Messier, L. & Wilson, J. Artificial intelligence adoption in urban planning governance: a systematic review of advancements in decision-making, and policy making. *Landsc. Urban Plan.* **259**, 105346 (2024).
12. Chen, J., Zhang, F., Fan, Z. & Liu, Y. Urban visual intelligence: studying cities with artificial intelligence and street-level imagery. *Ann. Am. Assoc. Geogr.* **114**, 876–897 (2024).
13. Cugurullo, F., Caprotti, F. & Cook, M. New stories of urban AI: exploring the artificial intelligence-city nexus beyond Frankenstein Urbanism. *Urban Geogr.* **45**, 1025–1048 (2024).
14. Laurière, M., Perrin, S., Geist, M. & Pietquin, O. Learning mean field games: a survey. *Nat. Mach. Intell.* **4**, 423–439 (2022).
15. Li, X., Wang, Q., Zhang, Y. & Chen, L. The fundamental issues and development trends of AI-driven transformations in urban transit and urban space. *Sustain. Cities Soc.* **101**, 105890 (2025).
16. Lartey, D. & Law, K. M. Artificial intelligence adoption in urban planning governance: a systematic review of advancements in decision-making, and policy making. *Landsc. Urban Plan.* **259**, 105346 (2025).

17. Yue, Y. et al. Shaping future sustainable cities with AI-powered urban informatics: toward human-AI symbiosis. *Comput. Urban Sci.* **5**, 31 (2025).
18. Xu, Y. & Cugurullo, F. When AIs become oracles: generative artificial intelligence, anticipatory urban governance, and the future of cities. *Cities* **145**, 104666 (2024).
19. Shin, M., Kim, J., van Ophesusden, B. & Griffiths, T. L. Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proc. Natl. Acad. Sci. USA* **120**, e2214840120 (2023).
20. Almulhim, A. & Cobbinah, P. Charting sustainable urban development through a systematic review of SDG11 research. *Nat. Cities* **1**, 117–131 (2024).
21. Titley, M., Butchart, S., Jones, V., Whittingham, M. & Willis, S. Global inequities and political borders challenge nature conservation under climate change. *Proc. Natl. Acad. Sci. USA* **118**, e2011204118 (2021).
22. Clarivate Analytics. Web of Science. <https://www.webofknowledge.com> (2024).
23. Elsevier. Scopus. <https://www.scopus.com> (2024).
24. Aum, S. & Choe, S. Srbert: automatic article classification model for systematic review using BERT. *Syst. Rev.* **10**, 285 (2021).
25. Rogers, A., Kovaleva, O. & Rumshisky, A. A primer in BERTology: what we know about how BERT works. *Trans. Assoc. Comput. Linguist.* **8**, 842–866 (2020).
26. Moradi, Z., Moradi, M. & Ziari, K. Comparative analysis of sustainable urban development: unraveling challenges and dimensions in different continents and utilizing AI with BERT model for articles classification. *Int. Rev. Spat. Plan. Sustain. Dev. D Plan. Assess.* **13**, 230–256 (2025).
27. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
28. Linderman, G. C. & Steinerberger, S. Clustering with t-SNE, provably. *SIAM J. Math. Data Sci.* **1**, 313–332 (2019).
29. Cai, T. T. & Ma, R. Theoretical foundations of t-SNE for visualizing high-dimensional clustered data. *J. Mach. Learn. Res.* **23**, 1–54 (2022).
30. Koumetio Tekouabou, S. C., Diop, E. B., Azmi, R., Jaligot, R. & Chenal, J. Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: potential, issues and challenges. *J. King Saud. Univ. Comput. Inf. Sci.* **34**, 5943–5967 (2022).
31. Pathak, A. R., Pandey, M. & Rautaray, S. Machine learning for spatial analyses in urban areas: a scoping review. *Appl. Soft Comput.* **108**, 107440 (2021).
32. Koumetio Tekouabou, C. S. et al. Identifying and classifying urban data sources for machine learning-based sustainable urban planning and decision support systems development. *Data* **7**, 170 (2022).
33. Djokić, V., Djordjević, A. & Milovanović, A. Big data and urban form: a systematic review. *J. Big Data* **12**, 17 (2025).
34. Tu, T., Zhang, E. & Long, Y. Profile and theoretical advances in urban big data studies: a systematic review of 57 representative journals (2013–2023). *Environment and Planning B: Urban Analytics and City Science*. Advance online publication. <https://doi.org/10.1177/23998083251346582> (2025).
35. Cook, M. & Karvonen, A. Urban planning and the knowledge politics of the smart city. *Urban Stud.* **61**, 370–382 (2024).
36. Sanchez, T. W. Planning on the verge of AI, or AI on the verge of planning. *Urban Sci.* **7**, 70 (2023).
37. Yigitcanlar, T., Agdas, D. & Degirmenci, K. Artificial intelligence in local governments: Perceptions of city managers on prospects, constraints and choices. *AI Soc.* **38**, 1135–1150 (2023).
38. Koumetio Tekouabou, S. C., Diop, E. B., Azmi, R., Jaligot, R. & Chenal, J. Artificial intelligence based methods for smart and sustainable urban planning: a systematic survey. *Arch. Comput. Methods Eng.* **30**, 1421–1438 (2023).
39. Wang, Z. & Ren, F. Developing a decision support system for sustainable urban planning using machine learning-based scenario modeling. *Sci. Rep.* **15**, 13210 (2025).
40. Xiang, X., Li, Q., Khan, S. & Khalaf, O. I. Urban water resource management for sustainable environment planning using artificial intelligence techniques. *Environ. Impact Assess. Rev.* **86**, 106515 (2021).
41. Xin, R., Ai, T., Ding, L., Zhu, R. & Meng, L. Impact of the COVID-19 pandemic on urban human mobility—a multiscale geospatial network analysis using New York bike-sharing data. *Cities* **126**, 103667 (2022).
42. Rogers, E. M. *Diffusion of Innovations* 5th edn (Free Press, 2003).
43. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
44. Batty, M. *Urban Modelling: Algorithms, Calibrations, Predictions*. No. 3 (Cambridge University Press, 1976).
45. Chapin, F. S. *Urban Land Use Planning* 2nd edn (eds Stuart, F. & Chapin Jr.) (University of Illinois Press, 1965).
46. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, vol. 25, 1097–1105 (NIPS, 2012).
47. Brown, T. et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, vol. 33, 1877–1901 (NeurIPS, 2020).
48. Alom, M. Z. et al. The history began from AlexNet: a comprehensive survey on deep learning approaches. Preprint at <https://doi.org/10.48550/arXiv.1803.01164> (2018).
49. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
50. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
51. Chen, T. & Guestrin, C. Xgboost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (ACM, 2016).
52. Jeffery, C. R. *Crime Prevention Through Environmental Design* (SAGE Publications, 1971).
53. Newman, O. *Defensible Space: Crime Prevention Through Urban Design* (Macmillan, 1972).
54. Chaturvedi, V. & de Vries, W. T. Machine learning algorithms for urban land use planning: a review. *Urban Sci.* **5**, 68 (2021).
55. Schirmer, P. M. & Axhausen, K. W. Machine learning for spatial analyses in urban areas: a scoping review. *Sustain. Cities Soc.* **85**, 104050 (2022).
56. Pery, C. A. The neighborhood unit: a scheme of arrangement for the family-life community. In *Regional Plan of New York and Its Environs*, vol. VII of *Neighborhood and Community Planning*, 2–140 (Regional Plan Association, 1929).
57. Lawhon, L. L. The neighborhood unit: physical design or physical determinism?. *J. Plan. Hist.* **8**, 111–132 (2009).
58. United States Department of Agriculture. Food waste FAQs (2019). <https://www.usda.gov/foodwaste/faqs>. In the United States, food waste is estimated at between 30–40 percent of the food supply, corresponding to approximately 133 billion pounds and \$161 billion worth of food in 2010.
59. Wolman, A. The metabolism of cities. *Sci. Am.* **213**, 179–190 (1965).
60. Kennedy, C., Cuddihy, J. & Engel-Yan, J. The changing metabolism of cities. *J. Ind. Ecol.* **11**, 43–59 (2007).
61. Mohai, P., Pellow, D. & Roberts, J. T. Environmental justice. *Annu. Rev. Environ. Resour.* **34**, 405–430 (2009).
62. Zhang, Q. & Li, H. Moea/d: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evolut. Comput.* **11**, 712–731 (2007).
63. Wu, Z. et al. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4–24 (2020).

64. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
65. Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L. & Lewis, B. A social vulnerability index for disaster management. *J. Homel. Secur. Emerg. Manag.* **8**, 3 (2011).
66. Liu, B., Guo, X. & Jiang, J. How urban morphology relates to the urban heat island effect: a multi-indicator study. *Sustainability* **15**, 10787 (2023).
67. Azizi, A. et al. A data-driven approach for urban heat island predictions: rethinking the evaluation metrics and data preprocessing. *Algorithms* **17**, 151 (2024).
68. Tanoori, G., Soltani, A. & Modiri, A. Machine learning for urban heat island (UHI) analysis: predicting land surface temperature (LST) in urban environments. *Urban Clim.* **56**, 101978 (2024).
69. Huang, C. et al. Analysis of the impact mechanisms and driving factors of urban spatial morphology on urban heat islands. *Sci. Rep.* **15**, 1–15 (2025).
70. Stewart, I. D. & Oke, T. R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **93**, 1879–1900 (2012).
71. Parker, D. E. Urban heat island effects on estimates of observed climate change. *Wiley Interdiscip. Rev. Clim. Change* **1**, 123–133 (2010).
72. Bibri, S. E. Compact city planning and development: emerging practices and strategies for achieving the goals of sustainability. *Dev. Built Environ.* **4**, 100021 (2020).
73. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
74. Cai, S., Wang, Z., Wang, S., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks for heat transfer problems. *J. Heat. Transf.* **143**, 060801 (2021).
75. Cuomo, S. et al. Scientific machine learning through physics-informed neural networks: where we are and what's next. *J. Sci. Comput.* **92**, 88 (2022).
76. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. Preprint at <https://doi.org/10.48550/arXiv.1609.02907> (2017).
77. Ai, T. & Yan, X. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS J. Photogramm. Remote Sens.* **150**, 259–273 (2019).
78. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
79. Uejio, C. K. et al. Intra-urban societal vulnerability to extreme heat: the role of heat exposure and the built environment, socioeconomics, and neighborhood stability. *Health Place* **17**, 498–507 (2011).
80. Harlan, S. L., Declet-Barreto, J. H., Stefanov, W. L. & Petitti, D. B. Neighborhood effects on heat deaths: social and environmental predictors of vulnerability in Maricopa County, Arizona. *Environ. Health Perspect.* **121**, 197–204 (2013).
81. Ribeiro, P. & Jardim Gonçalves, L. A. Urban resilience: a conceptual framework. *Sustain. Cities Soc.* **39**, 685–697 (2019).
82. Esposito, D. A ladder of urban resilience: an evolutionary framework for transformative governance of communities facing chronic crises. *Sustainability* **17**, 6010 (2025).
83. Felsenstein, D. & Grinberger, A. Y. Dynamic agent based simulation of welfare effects of urban disasters. *Comput. Environ. Urban Syst.* **59**, 129–141 (2017).
84. Schweikert, A., L'Her, G. & Deinert, M. Simple method for identifying interdependencies in service delivery in critical infrastructure networks. *Appl. Netw. Sci.* **6**, 1–23 (2021).
85. Zheng, Y. et al. Spatial planning of urban communities via deep reinforcement learning. *Nat. Comput. Sci.* **3**, 748–762 (2023).
86. Faraji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S. & Downey, K. An integrated 3d cnn-gru deep learning method for short-term prediction of PM2.5 concentration in urban environment. *Sci. Total Environ.* **834**, 155324 (2022).
87. Logan, T. M., Aven, T., Guikema, S. D. & Flage, R. Understanding cascading risks through real-world interdependent urban infrastructure. *Reliab. Eng. Syst. Saf.* **241**, 109652 (2023).
88. Cugurullo, F. et al. The rise of ai urbanism in post-smart cities: a critical commentary on urban artificial intelligence. *Urban Stud.* **61**, 197–216 (2024).
89. Li, L. & Zhao, N. Explicit and tacit knowledge have diverging urban growth patterns. *npj Urban Sustain.* **3**, 1–6 (2023).
90. van Lankveld, W. et al. Understanding disciplinary perspectives: a framework to develop skills for interdisciplinary research collaborations of medical experts and engineers. *BMC Med. Educ.* **24**, 1015 (2024).
91. Fenoglio, E. et al. Tacit knowledge elicitation process for Industry 4.0. *Discov. Artif. Intell.* **2**, 6 (2022).
92. Yang, J., Jiang, Z., Cheng, K. & Wu, L. Disciplinary barriers need communication: a behavioral and fnirs study under group decision-making paradigm shift based on cabin design. *Front. Neurosci.* **19**, 1594111 (2025).
93. Smith, P., Callagher, L. J., Hibbert, P., Krull, E. & Hosking, J. Developing interdisciplinary learning: Spanning disciplinary and organizational boundaries. *J. Manag. Educ.* **48**, 384–418 (2024).
94. McCance, K. R. & Blanchard, M. Measuring the interdisciplinarity and collaboration perceptions of U.S. scientists, engineers, and educators. *AERA Open* **10**, 23328584231218952 (2024).
95. Dietl, A.-K., Derksen, C., Keller, F. M. & Lippke, S. Interdisciplinary and interprofessional communication intervention: how psychological safety fosters communication and increases patient safety. *Front. Psychol.* **14**, 1128740 (2023).
96. Wang, S. et al. Artificial intelligence adoption in urban planning governance: a systematic review of advancements in decision-making, and policymaking. *Landsc. Urban Plan.* **259**, 105346 (2025).
97. Callaghan, M., Lamb, W. F. & Minx, J. C. Systematic global stocktake of over 50,000 urban climate change studies. *Nat. Cities* **2**, 1–12 (2025).
98. Malički, M., Jeroncic, A., Aalbersberg, I. J., Bouter, L. & Ter Riet, G. The present and future of peer review: ideas, interventions, and evidence. *Proc. Natl. Acad. Sci. USA* **121**, e2401232121 (2024).
99. Chen, B. et al. Contrasting inequality in human exposure to greenspace between cities of global north and global south. *Nat. Commun.* **13**, 4636 (2022).
100. Morewedge, C. K., Jost, C. E., Herzenstein, M. & Park, J. People see more of their biases in algorithms. *Proc. Natl. Acad. Sci. USA* **121**, e2317602121 (2024).
101. Bai, X., Nagendra, H., Shi, P. & Liu, H. Integration of urban science and urban climate adaptation research: opportunities to advance climate action. *npj Urban Sustain.* **3**, 32 (2023).
102. Shrivastava, M. et al. Urban pollution greatly enhances formation of natural aerosols over the amazon rainforest. *Nat. Commun.* **10**, 1046 (2019).
103. Kirkpatrick, J. et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. USA* **114**, 3521–3526 (2017).
104. Kashinath, K. et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philos. Trans. R. Soc. A* **379**, 20200093 (2021).
105. Bi, K. et al. Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **619**, 533–538 (2023).
106. Awad, E. et al. The moral machine experiment. *Nature* **563**, 59–64 (2018).
107. Brown, C. F. et al. Dynamic world, near real-time global 10 m land use land cover mapping. *Sci. Data* **9**, 251 (2022).
108. Rahwan, I. et al. Machine behaviour. *Nature* **568**, 477–486 (2019).

109. Arun, C. Ai and the global south: designing for other worlds **3**, 1–16 (2019).
110. Peng, Z.-R., Lu, K.-F., Liu, Y. & Zhai, W. The pathway of urban planning AI: from planning support to plan-making. *J. Plan. Educ. Res.* **44**, 2285–2302 (2024).
111. Palermo, P. C. & Ponzini, D. Whatever is happening to urban planning and urban design? musings on the current gap between theory and practice. *City, Territ. Archit.* **1**, 1–16 (2014).
112. Chen, W., Zhao, L., Kang, Q. & Di, F. Systematizing heterogeneous expert knowledge, scenarios and goals via a goal-reasoning artificial intelligence agent for democratic urban land use planning. *Cities* **101**, 102703 (2020).
113. Gao, C. et al. Large language models empowered agent-based modeling and simulation: a survey and perspectives. *Humanit. Soc. Sci. Commun.* **11**, 1–30 (2024).
114. Wang, J. et al. Large language models as urban residents: an LLM agent framework for personal mobility generation. In *Proc. 38th Int. Conf. Neural Information Processing Systems* Vol. 3957, 28 (NeurIPS, 2024).
115. Luusua, A., Ylipulli, J., Foth, M. & Aurigi, A. Urban ai: understanding the emerging role of artificial intelligence in smart cities. *AI Soc.* **37**, 1335–1344 (2022).
116. Pries, J. Spatial theory in planning practice? on the concepts of space that made urban design a planning solution for segregation in malmö, Sweden. *Antipode* **56**, 1024–1044 (2024).
117. Sanchez, T. W., Shumway, H., Gordner, T. & Lim, T. The prospects of artificial intelligence in urban planning. *Int. J. Urban Sci.* **27**, 179–194 (2023).
118. Ouma, Y. O. et al. Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach. *Eur. J. Remote Sens.* **56**, 2173659 (2023).
119. Yan, X. et al. A multimodal data fusion model for accurate and interpretable urban land use mapping with uncertainty analysis. *Int. J. Appl. Earth Observ. Geoinf.* **129**, 103805 (2024).
120. Yang, L. & Zhou, G. Dissecting the Analects: an NLP-based exploration of semantic similarities and differences across English translations. *Humanit. Soc. Sci. Commun.* **11**, 50 (2024).
121. Tyagi, N. & Bhushan, B. Demystifying the role of natural language processing (NLP) in smart city applications: background, motivation, recent advances, and future research directions. *Wireless Pers. Commun.* **130**, 857–908 (2023).
122. Abid, N. et al. Algorithmic urban planning for smart and sustainable development: systematic review of the literature. *Sustain. Cities Soc.* **94**, 104562 (2023).
123. Peng, Z.-R., Lu, K.-F., Liu, Y. & Zhai, W. The Pathway of Urban Planning AI: From Planning Support to Plan-Making. *J. Plan. Educ. Res.* **44**, 2263–2279 (2024).
124. Khadhraoui, M., Bellaaj, H., Ammar, M. B., Hamam, H. & Jmaiel, M. Survey of bert-base models for scientific text classification: Covid-19 case study. *Appl. Sci.* **12**, 2891 (2022).
125. Li, X. & Jia, L. English text topic classification using BERT-based model. *J. Comput. Methods Sci. Eng.* **25**, 669–684 (2025).
126. Bannach-Brown, A. et al. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. *Syst. Rev.* **8**, 23 (2019).
127. Tóth, B., Berek, L., Gulácsi, L., Péntek, M. & Zrubka, Z. Automation of systematic reviews of biomedical literature: a scoping review of studies indexed in PubMed. *Syst. Rev.* **13**, 174 (2024).
128. Xu, D. et al. Large language models for generative information extraction: a survey. *Front. Comput. Sci.* **18**, 186357 (2024).
129. Thapa, S. et al. Large language models (LLM) in computational social science: prospects, current state, and challenges. *Soc. Netw. Anal. Min.* **15**, 4 (2025).
130. Yao, Y. et al. A survey on large language model (LLM) security and privacy: the good, the bad, and the ugly. *High.-Confid. Comput.* **4**, 100211 (2024).
131. Reimers, N. & Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* 3982–3992 (2019).
132. Kumar, A., Singh, J. P., Kumar, N. P. et al. BERT applications in natural language processing: a review. *Artif. Intell. Rev.* **58**, 1–90 (2025).
133. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231 (AAAI Press, 1996).
134. Prager, E. M. et al. Improving transparency and scientific rigor in academic publishing. *Cancer Rep.* **2**, e1150 (2019).
135. Tobi, H. & Kampen, J. K. Research design: the methodology for interdisciplinary research framework. *Qual. Quant.* **52**, 1209–1225 (2017).
136. Gasparetto, A., Marcuzzo, M., Zangari, A. & Albarelli, A. A survey on text classification: from traditional to deep learning. *Information* **13**, 83 (2022).
137. Li, Q. et al. A survey on text classification: From traditional to deep learning. *ACM Trans. Intell. Syst. Technol.* **13**, 1–41 (2022).
138. Petit, N. & Teece, D. J. Innovating big tech firms and competition policy: favoring dynamic over static competition. *Ind. Corp. Change* **30**, 1168–1198 (2021).
139. van den Bulk, L. M. et al. Automatic classification of literature in systematic reviews on food safety using machine learning. *Curr. Res. Food Sci.* **5**, 84–95 (2022).
140. Lin, L., Zhou, D., Wang, J. & Wang, Y. A systematic review of big data driven education evaluation. *SAGE Open* **14**, 1–18 (2024).
141. Saputra, N. A., Riza, L. S., Setiawan, A. & Hamidah, I. A systematic review for classification and selection of deep learning methods. *Digit. Signal Process.* **146**, 104393 (2024).
142. Minaee, S. et al. Deep learning-based text classification: a comprehensive review. *ACM Comput. Surv.* **54**, 1–40 (2021).
143. Schneider, P. et al. A decade of knowledge graphs in natural language processing: a survey. In *Proc. 2nd Conf. Asia-Pacific Chapter Assoc. Comput. Linguistics and 12th Int. Joint Conf. Natural Language Processing*, vol. 1, Long Papers, 601–614 (2022).
144. Pan, J. et al. Large language models and knowledge graphs: opportunities and challenges. *Trans. Graph Data Knowl.* **1**, 2–1238 (2023).
145. Xuefeng, B. et al. Construction of a knowledge graph for framework material enabled by large language models and its application. *npj Comput. Mater.* **11**, 217 (2025).
146. Venugopal, V. & Olivetti, E. Matkg: an autonomously generated knowledge graph in material science. *Sci. Data* **11**, 217 (2024).
147. Mondal, I., Hou, Y. & Jochim, C. End-to-end construction of NLP knowledge graph. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP*, vol. 2021, 1885–1895 (2021).
148. Zhong, L., Wu, J., Li, Q., Peng, H. & Wu, X. A comprehensive survey on automatic knowledge graph construction. *ACM Comput. Surv.* **56**, 1–62 (2023).
149. Zhu, Y. et al. LLMs for knowledge graph construction and reasoning: recent capabilities and future opportunities. *World Wide Web* **27**, 58 (2024).
150. Zhou, X., Zhou, M., Huang, D. & Cui, L. A probabilistic model for co-occurrence analysis in bibliometrics. *J. Biomed. Inform.* **128**, 104047 (2022).
151. Yuan, C., Li, G., Kamarthi, S., Jin, X. & Moghaddam, M. Trends in intelligent manufacturing research: a keyword co-occurrence network based review. *J. Intell. Manuf.* **33**, 425–439 (2022).

152. Wang, Y. et al. Exploring academic influence of algorithms by co-occurrence network based on full-text of academic papers. *Aslib J. Inf. Manag.* **77**, 651–680 (2025).
153. Cook, D., Brydges, R., Ginsburg, S. & Hatala, R. Validation of educational assessments: a primer for simulation and beyond. *Adv. Simul.* **8**, 27 (2023).
154. Liang, X. & Zhang, Y. A validity framework for accountability: educational measurement and language testing. *Lang. Test. Asia* **12**, 3 (2022).
155. van Haastrecht, M. et al. Vast: a practical validation framework for e-assessment solutions. *Inf. Syst. e-Bus. Manag.* **21**, 1–32 (2023).
156. Zhang, S. et al. Development and validation of an instrument for assessing scientific literacy from junior to senior high school. *Discip. Interdiscip. Sci. Educ. Res.* **5**, 21 (2023).
157. Qian, K., Chen, Z., Jiao, H., Li, N. et al. AI agent as urban planner: steering stakeholder dynamics in urban planning via consensus-based multi-agent reinforcement learning. Preprint at <https://doi.org/10.48550/arXiv.2310.16772> (2023).
158. Zhou, Z., Lin, Y., Jin, D. & Li, Y. Large language model for participatory urban planning. Preprint at <https://doi.org/10.48550/arXiv.2402.17161> (2024).
159. Budenny, S. A., Voskresenskiy, A. V., Shichkin, A. V., Bekhtin, Y. et al. LLM agents for smart city management: enhancing decision support through multi-agent AI systems. *Smart Cities* **8**, 19 (2025).
160. Wang, J., Huang, J. X. & Sheng, J. An efficient long-text semantic retrieval approach via utilizing presentation learning on short-text. *Complex Intell. Syst.* **10**, 963–979 (2024).
161. Zhang, F., Khomiakov, M., Zhou, J., Noyman, A. & Duarte, F. Generative spatial artificial intelligence for sustainable smart cities: a pioneering large flow model for urban digital twin. *Sustain. Cities Soc.* **121**, 106043 (2025).
162. Bottou, L., Curtis, F. E. & Nocedal, J. Optimization methods for large-scale machine learning. *SIAM Rev.* **60**, 223–311 (2018).
163. Talbi, E.-G., Basseur, M., Nebro, A. J. & Alba, E. Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Mach. Learn.* **113**, 4055–4118 (2024).
164. Yu, B., Yin, H. & Zhu, Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proc. 27th Int. Joint Conf. Artificial Intelligence (IJCAI-18)*, 3634–3640 (2018).
165. Ma, W., Chu, Z., Chen, H. & Li, M. Spatio-temporal evolutionary graph neural network for traffic flow prediction in UAV-based urban traffic monitoring system. *Sci. Rep.* **15**, 1234 (2025).
166. Kahn, B. K., Strong, D. M. & Wang, R. Y. Information quality benchmarks: product and service performance. *Commun. ACM* **45**, 184–192 (2002).
167. Wang, R. Y. & Strong, D. M. Beyond accuracy: what data quality means to data consumers. *J. Manag. Inf. Syst.* **12**, 5–33 (1996).
168. Shashaani, S., Ng, S. H. & Eckman, D. Robust output analysis with Monte-Carlo methodology. Preprint at <https://doi.org/10.48550/arXiv.2207.13612> (2023).
169. El-Horbaty, Y. S. & Hanafy, E. M. A Monte Carlo permutation procedure for testing variance components using robust estimation methods. *Stat. Pap.* **65**, 335–356 (2024).
170. Kozlova, M. & Yeomans, J. S. Extending system dynamics modeling using simulation decomposition to improve the urban planning process. *Front. Sustain. Cities* **5**, 1129316 (2023).
171. Choi, H. S. & Zhang, W. Artificial intelligence as research methods in urban design. *J. Urban Des.* **29**, 182–203 (2024).
172. He, W. & Chen, M. Advancing urban life: a systematic review of emerging technologies and artificial intelligence in urban design and planning. *Buildings* **14**, 835 (2024).

## Acknowledgements

This research was supported by the Basic Research Program of Jiangsu (Grant No. BK20241815) and the Xi'an Jiaotong Liverpool University Research Development Fund (Grant No. RDF-23-02-004). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

J.T. conceived and designed the study, conducted the comprehensive literature review and data collection, developed the theoretical framework, designed and implemented the multi-agent recommendation system, performed all data analysis and technical validation, developed the case studies, and wrote the manuscript. S.W. provided supervision, guidance on urban theory integration, and contributed to manuscript revision. G.W. assisted with technical implementation and validation. Y.W. created partly visual illustrations and graphics. J.M. provided research supervision, theoretical guidance, and contributed to manuscript revision. All authors participated in manuscript review and revision, and have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42949-026-00377-2>.

**Correspondence** and requests for materials should be addressed to Shuihua Wang or John Moraros.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026