# Urban planning in the era of large language models

**Yu Zheng[1,2], Fengli Xu[1], Yuming Lin[3*], Paolo Santi[2,4], Carlo Ratti[2], Qi R. Wang[5*], and Yong Li[1*]**

[1]Department of Electronic Engineering, Tsinghua University, Beijing, China
[2]Senseable City Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]School of Architecture, Tsinghua University, Beijing, China
[4]Istituto di Informatica e Telematica del CNR, Pisa, Italy
[5]Department of Civil and Environmental Engineering, Northeastern University, Boston, Massachusetts, USA
[*]Corresponding authors: linyuming9@foxmail.com, q.wang@northeastern.edu, liyong07@tsinghua.edu.cn

## ABSTRACT

City plans are the product of integrating human creativity with emerging technologies, which continuously evolve and reshape urban morphology and environments. Here, we argue that large language models (LLMs) hold large untapped potential in addressing the growing complexities of urban planning and enabling a more holistic, innovative, and responsive approach to city design. By harnessing their advanced generation and simulation capabilities, LLMs can contribute as an intelligent assistant for human planners in synthesizing conceptual ideas, generating urban designs, and evaluating the outcomes of planning efforts.

## Introduction

Urban planning plays a critical role in shaping the quality of life for city residents, influencing every aspect of the city, from land use and facility placement to transportation and travel[1–4]. With billions of people living in or migrating to cities, developed urban regions require renovation, while newly developing urban areas need to be carefully planned to accommodate growth[5]. Historically, urban planning methodologies have continually evolved to meet the ever-expanding needs of urban environments[6–8]. Today, as new trends emerge in urban planning, planners are confronting critical challenges. First, urban planning is increasingly complex within interdisciplinary contexts, extending beyond simple spatial layouts to encompass comprehensive factors of social equity[9,10], environmental resilience[11], urban sustainability[12], and so forth. Addressing this growing complexity demands a deep understanding of diverse knowledge across multiple domains. Second, new concepts, such as the 15-minute city[13–15], which promotes efficient layout such that basic urban services are accessible within walking or cycling distance, are emerging. Adapting to new concepts, reducing their negative effects such as potential segregation risk in 15-minute city[13], and further translating these theoretical ideas into actionable urban plans present a significant challenge that requires strong reasoning capabilities to align tangible geospatial layout with concrete descriptions of planning objectives. Third, cities are giant systems involving a large number of entities and complicated interactions[16–20], which makes the evaluation of urban planning particularly challenging[21–24]. To address these challenges, future approaches in urban planning are anticipated to empower designers through the integration of cutting-edge technologies featuring advanced abilities in understanding, reasoning and simulating urban dynamics.

Artificial Intelligence (AI) has made a substantial impact to the processes of creation and design across a wide range of areas and disciplines. In particular, since the release of ChatGPT in late 2022, Large Language Models (LLMs) have demonstrated noteworthy generative abilities[25–28]. Specifically, different from pre-existing AI models that narrowly focus on specific cognitive tasks and application domains, LLMs use natural language as a unifying "code" to represent knowledge in a wide variety of domains[29,30]. Through large-scale self-supervised pretraining on diverse domain data, supervised fine-tuning on instruction pairs, and post-training based on reinforcement learning, LLMs develop superior reasoning capabilities[31]. With notable examples including the recently released DeepSeek-R1 model[32] and OpenAI o-series models[33,34], LLMs are now able to proficiently engage in chatting[26,35], writing codes[36], and even deriving mathematical expressions[37,38]. Beyond natural language, visual large models (VLMs)[39,40], such as large diffusion models[41,42] that align different modalities by learning from vast image-caption datasets, can generate high-quality artworks including images[41], 3D scenes[43,44], and even videos[45]. Moreover, LLM Agents integrate external memory, tool usages, and planning into LLMs, enhancing their ability to accomplish complex tasks in an interactive environment[46–49], such as robotic manipulation[50] and social simulation[51,52].

Despite LLMs' emergent design abilities across various fields, what remains largely unexplored is the design of the environments in which we live—our cities. Here, we advocate for integrating LLM into the future collaborative workflows between planners and AI toward more intelligent city design.

## Contemporary urban planning

### Two traditional approaches

Since Howard's Garden City[53] proposal in the 1890s, urban planning has been a critical discipline continuously evolving to meet the challenges posed by rapid urbanization. Before 1940s, the planning of cities was regarded primarily as architectural design at a larger scale[54]. As a consequence, planners used to take a physical and aesthetic view, with the planning theories focusing more on the spatial arrangement of urban structures and functions[55–59], exemplified by the New Towns Act in the United Kingdom[60] and Urban Renewal Program in the United States[61]–both largely adhering to these concepts. However, this "physical art" approach overlooked the social dimensions of city life and did not fully account for human activities, failing to account for critical issues such as high crime rates and urban decay[62].

After the second World War, urban planning shifted from an artistic practice to a more scientific discipline[54,62]. This new approach regards cities as large, complex systems comprising various interconnected sub-systems, and employs well-curated models to account for the diverse social aspects of cities[7]. Rather than creating static designs, modern urban planning emphasizes the dynamic processes within cities, particularly the daily activities of urban residents[63–65]. Influential approaches include Lindblom's incremental planning[66], Faludi's planning theory[67], and Healey's communicative planning[68]. This "analytic science" approach has also inspired contemporary ideas such as the concept of 15-minute city[13,14], which advocates for dense, compact land use that fosters more equitable, convenient, and inclusive urban living.

Despite their substantial impact on modern urban planning, the two approaches discussed above are insufficient to address the complexities of cities in the current era. First, the planning process remains primarily planner-centered, involving complex concepts and regulations often inaccessible to the general public. This limits effective public participation of multiple stakeholders, an increasingly crucial practice in today's urban planning[69,70]. Second, the evaluation of urban planning in these two approaches is typically coarse-grained, subjective, and qualitative. However, contemporary urban planning increasingly demands fine-grained, quantitative, and objective feedback to inform scientific decision-making[22,23].

### Recent advances using pre-LLM AI models

With the unprecedented availability of urban geospatial data, a data-driven approach has emerged to enhance traditional urban planning methods with pre-LLM AI models[71–73]. These AI models feature two main advantages. First, they have the capability to discover the underlying rules from these rich datasets. These rules guide more accurate prediction of urban spatio-temporal patterns[74], enabling the generation of city designs that closely resemble real-world structures both visually and statistically[75]. Generative Adversarial Networks (GANs)[76] and Variational Auto-Encoders (VAEs)[77] are two of these models that are commonly used to synthesize urban elements, including street networks[75,78], functional zoning[79–81], and building footprints[82,83]. Second, these models assist in making strategic decisions to optimize specific metrics that reflect the efficiency and quality of urban life. Reinforcement learning[84] is typically employed to achieve effective urban planning, such as 15-minute community layout[85,86], road planning[87], and metro network expansion[88,89].

While these AI models have enhanced urban planning in important ways, their limited scope hinders their ability to fully capture the complexities of modern cities. Specifically, urban planning requires a comprehensive understanding of diverse knowledge. Since pre-LLM small models are trained on limited task-specific datasets, they struggle to address the growing interdisciplinary nature of urban planning. As a consequence, these models often focus on a narrow range of aspects applicable in restricted scenarios, with limited generalizability across diverse urban planning tasks. In light of these shortcomings, LLMs, with rich embedded knowledge, hold the potential to overcome the limitations of smaller models in urban planning.

## Opportunities brought by LLMs

The past two years have witnessed advancements in LLMs, transforming the workflows of human designers with their generative abilities. The following developments provide valuable opportunities to enhance urban planning practices:

- **LLMs possess and leverage vast arrays of transdisciplinary knowledge.** Effective urban planning demands a comprehensive understanding of various domains, including geography, economics, sociology, and so forth. The versatile nature of LLMs is crucial for addressing such interdisciplinary complexity inherent in urban planning. Specifically, LLMs have demonstrated emergent abilities in answering complicated questions across different subjects and domains. For example, benchmarking results have shown that LLMs demonstrate deep knowledge in math[90], medicine[91], law[92], and finance[93]. In particular, researchers have observed a scaling law where performance substantially surpasses random guessing once LLMs cross a certain threshold of model parameters and training data[94]. Therefore, by increasing model size and pre-training on larger urban planning datasets, LLMs have the potential to process and leverage rich domain knowledge to account for the multiple aspects of urban planning.

- **LLMs are capable of more elaborate reasoning based on conceptual instructions.** Translating abstract concepts into concrete, satisfactory design is a challenging task. Through reinforcement learning adopted by the recently released DeepSeek-R1 model[32] and OpenAI o-series models[33,34], LLMs can achieve more elaborate reasoning to decompose complex tasks into smaller steps. The reasoning capabilities can extend beyond a single language modality and encompass high-dimensional modalities. In particular, VLMs have demonstrated good performance in analyzing multi-modal contents[95–97] and generating highly realistic images[41] and videos[98] based on conceptual linguistic instructions, potentially offering benefits for designing cities that align with diverse planning concepts. Particularly, significant advancements are observed in understanding urban-related visual images such as satellite images[95] and street-view images[96,97] with VLMs.
- **LLMs can enhance predictive evaluation and public participation in urban planning.** LLM agents could perform role-playing with heterogeneous and personalized profiles defined by residents and planners, enabling predictive simulation of residents' daily activities. Researchers have successfully utilized LLM agents for simulations across physical[50], social[51,52,99], and cyber domains[100], indicating large potential for urban simulation that allows for more accurate evaluations of urban planning. Also, community engagement plays an increasingly crucial role in contemporary urban planning, yet traditional planning approaches often fail to make public participation accessible. LLMs can help address this discrepancy by providing a user-friendly protocol to enhance public participation: with capabilities in conducting human-like dialog using natural language, LLMs can facilitate an intuitive and conversational interface to engage the public. In this way, residents, together with planners, can directly interact with LLMs, refining urban plans through multiple rounds of discussions.

These advantages of LLMs have led to a series of research projects that developed specialized urban LLMs[101–110] that feature a deepened understanding of urban environments, showcasing the new opportunities brought to urban planning by LLMs.

## Towards LLM-driven urban planning

To bridge the gap, we introduce an LLM-driven planning framework that serves as an intelligent assistant to support human planners, integrating AI's computational power with human expertise and domain insights to enhance planning efficiency and decision-making. This framework consists of three interconnected components to enhance creativity and decision-making, as illustrated in Figure 1. First, LLMs support the conceptualization of urban planning by assisting planners in developing prototypical ideas and crafting detailed descriptive planning texts for the site to be designed. Through analyzing textual input, such as planning needs, requirements, and guidelines, LLMs can help identify regions in need of renovation and offer strategic suggestions for optimizing urban layout, incorporating diverse knowledge from related fields. Second, VLMs facilitate the generation of urban designs by transforming planners' input prompts—reflecting planning concepts and constraints—into detailed visual output, such as layouts and cityscapes, all conditioned on user-defined draft plans. Third, an LLM agent is employed to evaluate the planning effect, which incorporates generated urban plans, demographics and other conditions to simulate complex urban dynamics, including human mobility. This approach has the potential to provide a quantitative and accurate assessment of how residents interact with the city, offering actionable feedback for planners. The provided framework creates a blueprint for how LLMs can assist planners, where LLMs, VLMs, and LLM agents enable a systematic and collaborative workflow of conceptualization, generation, and evaluation in urban planning, addressing critical limitations of traditional approaches.

### Conceptualization

Synthesizing conceptual ideas is the first step in urban design, as it sets the foundational tone and overarching vision for the planned city. Conceptualization involves the abstract and preliminary arrangement of urban functionalities and forms, often depicted through condensed textual descriptions, overall morphologies, and critical elements such as corridors and hubs tailored for specific purposes. A principled approach is essential to achieve effective conceptual design, requiring a comprehensive consideration of the complex urban context, including geospatial conditions, master plans, and guidelines, as well as key factors like environmental sustainability, resilience, social equity, and economic prosperity. Traditionally, human designers undertake this intricate task manually, engaging in discussions, consultations, and negotiations with multiple stakeholders to refine these concepts. The outcome is typically a set of textual documents that clearly articulate the motivations and logic behind the prototypical planning concepts. However, the conceptualization process is highly time-consuming, particularly when relying solely on human effort.

Figure 2 illustrates how the framework enhances the conceptualization process by integrating LLMs to support human planners and boost their productivity. An LLM, exposed to vast textual contents from various urban planning-related fields, such as environmental science, sociology, and economics, is employed within the framework. Through large-scale pretraining, the

LLM gains extensive knowledge across these fields, allowing it to generate informed responses in urban planning discussions, complementing the expertise of experienced human planners. Then, the framework facilitates the conceptualization process through a conversational iteration between human planners and the LLM. In each round of conversation, human planners begin by conducting prompt engineering, transforming planning needs, requirements, guidelines, and other relevant materials into informative prompts. The LLM then responds with conceptual ideas, offering suggestions on urban forms and functionalities. Leveraging its extensive spatio-temporal knowledge of the real world[111,112], the LLM is able to propose layouts for conceptual elements such as centers and corridors, while its advanced reasoning abilities allow it to navigate complex contexts, including input needs, guidelines, and prior conversations. Post-training and prompting strategies such as GraphRAG[113] and chain-of-thought (COT)[114] combined with reinforcement learning (RL)[115] further enhance the LLM's reasoning capabilities, making it more adept at achieving effective planning conceptualization. Based on the LLM's responses, human planners can refine the input and initiate additional rounds of conversation to revise and improve the conceptualization results. Meanwhile, since LLMs may exhibit biases—particularly toward overrepresented regions or domains in their training data[116]—the involvement of human planners is essential to identify and correct potential biases in the generated planning outcomes, ensuring that the results remain contextually appropriate and equitable across diverse urban environments. Ultimately, this collaborative process produces a detailed and satisfactory textual description of urban planning concepts, with all prototypical spatial arrangements documented in output texts that can be further visualized as draft plans. In brief, LLMs have the capability to effectively support human planners during the conceptualization phase, acting as responsive, human-like consultants with transdisciplinary knowledge and reasoning abilities, aiding in the synthesis of ideas based on rich domain knowledge and contextual information.

To offer an intuitive illustration on LLMs' ability in handling diverse and complex urban planning concepts and planning-related text, we tested their performance in qualification exams for professional human planners[117], which cover a comprehensive range of urban planning-related disciplines, including transportation, economics, geography, sociology, ecology and environment. We evaluated the accuracy of questions regarding *basic theories* and *practical know-how* of urban planning, using the Qwen2 LLM[118] of different sizes. The results show that the largest LLM with 70 billion parameters has outperformed the top 10% of human planners in answering challenging questions related to planning concepts, implying the potential in synthesizing conceptual ideas in the initial stages of urban planning with LLMs.

## Generation

Generating specific urban layouts lies at the core of urban planning, as it shapes the spatial organization of cities and influences subsequent urban activities. The generation process addresses diverse objectives, including land use, road networks, facility locations, public transportation systems, and so forth. Unlike conceptual ideas expressed through natural language, urban layouts are described with concrete geospatial elements and locations, which require precise representation, often in imagery or more accurate vector formats. This process demands a principled approach capable of handling contents across multiple modalities while capturing the complex interrelationships between them. Moreover, urban layouts cannot be generated arbitrarily; they must adhere to various constraints such as geography, social dynamics, land ownership and so forth. Thus, controllable generation must account for these constraints, such that the final layout aligns seamlessly with the initial urban context as well as customized planning concepts.

In the LLM-driven urban planning framework, VLMs assist human planners in the urban layout generation process, as illustrated in Figure 3a. To fully harness the multi-modal content generation capabilities of VLMs, we propose first constructing a large-scale conditional text-to-image generation dataset tailored towards urban design, which includes the basemap, layouts, and corresponding planning description texts collected from existing cities. Using this dataset, we can fine-tune modern VLMs specifically for urban layout tasks. For example, standard VLMs like CLIP[39] can be integrated with generative models and finetuned on the urban design task. Meanwhile, the finetuning process can be efficiently achieved through low-rank adaption (LoRA)[119]. Finetuned VLMs then map planning descriptions of textual prompts by human planners to visual urban designs, including land use zoning layout (Figure 3b), building footprint (Figure 3c), and 3D urban scene (Figure 3d)[120]. It is worth noting that controllable generation is essential for assisting planners and designers, where techniques such as ControlNet[121] and Dreambooth[122] can be employed to generate urban layouts that comply with the geospatial constraints of the built environment.

Not all urban design can be computationally generated, and we anticipate VLMs to serve as critical planning support tools such as GIS once did. For example, planners can iteratively refine the prompts based on the generated designs for further improvement–wherein discussions with policymakers and other stakeholders play a critical role in aligning urban plans with broader public interests. VLMs' multi-modal and controllable generation capabilities have the potential to enable an agile and flexible urban design process, freeing human planners from labor-intensive layout tasks and allowing them to focus more on refining conceptual ideas, developing innovative solutions, and coordinating different stakeholders.

## Evaluation

Evaluating urban planning is crucial for measuring its impact, delivering actionable feedback, and guiding future improvements. Traditionally, this evaluation has been subjective, depending on personal judgment such as expert opinions and stakeholder interviews. Quantitative metrics–such as economic indicators and environmental quality–have often been limited in scope or used only in specific contexts, making it difficult to compare outcomes objectively or provide consistent, data-driven feedback. In addition, the evaluation is typically conducted after the design and construction, thus leaving insufficient room for improvements. To address this, a more predictive approach is needed to facilitate evaluation in advance, typically involving multi-dimensional simulations of urban life within city digital twins[22,23,123–126]. Particularly, human activities vary widely depending on demographic characteristics and are deeply interconnected, further complicating accurate simulations.

Figure 4 illustrates the LLM-driven evaluation process of urban planning using LLM agents, which could allow more accurate bottom-up simulation of daily experiences and behaviors of urban residents, generating outputs that provide detailed, quantitative insights into how different planning scenarios may affect city life, including mobility patterns, facility usage, and so forth. These agents, equipped with advanced decision-making abilities, could simulate everyday activities in the city environment, surpassing traditional agent-based simulation tools. First, LLM agents can be personalized with diverse demographic profiles, such as gender, occupation, and age, each representing unique roles with heterogeneous needs and activities, to fully leverage their role-playing capabilities[51] for comprehensive evaluation. Second, LLM agents come equipped with tools that enhance the effectiveness of simulations. Within urban context, geospatial tools such as navigation and Google Places API can be utilized to determine routes between locations, and simulate real-world mobility patterns. Third, LLMs possess memory, allowing them to store historical behaviors, observations, and interactions with other agents. To accelerate and scale up urban simulation based on LLM agents, it is essential to adopt prompt optimization strategies[127] to address the computational and communication challenges. This memory enables reflection and adaptive learning, improving the simulation of complex urban dynamics over time. As a result, LLM-driven evaluation in urban planning could provide more accurate and quantitative feedback on how residents will interact with the city, effectively coupling simulation and decision-making for continuous iterations.

It is important to recognize that human behavior is inherently uncertain and cannot be fully captured by deterministic agent-based simulations[128]. To address this limitation, it is essential to develop evaluation benchmarks that explicitly incorporate uncertainty and contextual variability. For instance, recent studies have drawn on behavioral economics theories to evaluate LLM-based decision-making under uncertain conditions, accounting for diverse socio-demographic factors[129]. In parallel, stochastic modeling approaches[130] can be integrated with LLM agents to better reflect the randomness and variability in human behavior—such as the probabilistic nature of pedestrian movements and decision-making processes[131].

To demonstrate the effectiveness of the above evaluation approach, we employed LLM agents to simulate facility visitation of two communities in New York and Chicago, and compareed the simulation results with real mobility data[2]. The frequently visited locations by LLM agents closely mirrored the groundtruth hot spots observed in empirical human mobility data. Meanwhile, quantitative metrics such as the 15-minute usage[13] of different types of urban facilities, which measures whether these services are accessible within 15-minute walking or cycling by simulation, were also consistent with the groundtruth for communities in both New York and Chicago. These results indicate that LLM agent simulations could predict how residents will use the planned city, offering useful feedback on the quality of urban planning.

## Limitations

Despite the promises, transforming the above blueprint into practical tools within the urban planning workflow faces several technical challenges. First, to build reliable LLMs, millions or billions of training samples are often required, as evidenced by the scaling laws in LLMs where emergent intelligence is only realized when training data surpasses a critical threshold[132]. While remote sensing data, such as satellite images and crowdsourced platforms, has lowered barriers to accessing urban functionality details[133,134]—such as land use types, road networks, and points of interest (POI)—there remains an urgent need for high-quality urban design data. This includes spatial layouts and corresponding descriptive texts, which are typically owned by governments and design firms under strict access restrictions. We encourage the urban planning community to increase data availability by creating open, collaborative platforms for sharing and exchanging urban design data.

Second, besides data accessibility, the substantial computational resources required to train LLMs and VLMs–often necessitating large number of advanced GPUs–also remain prohibitively expensive for most researchers and practitioners in urban planning. Therefore, the development of computationally efficient variants of LLMs is essential to reduce training costs. Particularly, it is a promising direction to build smaller, specialized LLMs tailored to urban planning, as specialized LLMs usually require far fewer parameters than general-purpose counterparts[135], which can substantially lower the barrier to LLM adoption and make it more widely available to practitioners. It is also worth noting that computational challenge exists in LLM agent-driven evaluation of urban planning, especially when simulating large urban populations at the individual level[22]. To address this, simulation can be conducted for a carefully sampled subset of the population, the results of which can then be statistically extrapolate to the whole population[136]. Meanwhile, smaller, specialized, or distilled LLMs can be employed to reduce computational cost for agent simulation rather than using massive general-purpose LLMs.

Third, to evaluate the effect of LLMs in urban planning, it is crucial to perform accurate simulations of human activities.

However, human activities exhibit large uncertainty and divergent patterns in different timescales from daily routines to multi-year trends. Estimating these patterns is particularly complicated when considering longer time spans, which typically displays larger uncertainty as it involves factors like population change that fundamentally influence the urban dynamics[137–139]. Therefore, it poses additional challenges to the usage of LLM agents for evaluating urban plans, both in terms of modeling the intricate decision-making processes of urban residents and managing the significant computational complexity of large-scale, high-fidelity simulations.

Last, real-world application of LLMs in urban planning needs to account for various biases, in both the training data and algorithms, as well as model outputs. Particularly, several specialized LLMs have been proposed to address urban related tasks, such as mobility prediction[101], traffic signal control[102], and spatial navigation[103]. These LLMs rely on large-scale urban data during pretraining and finetuning, which typically contain significant bias. For example, large cities tend to generate much more data, which can make LLMs biased towards these population centers, resulting in suboptimal performance when applied to smaller, less representative cities. In particular, recent research[116] has shown that LLMs are geographically biased, especially for locations with lower socioeconomic conditions. Meanwhile, social biases related to demographics such as gender and racial bias have also been observed in LLM-generated content[140–142], posing significant challenges for their deployment in real-world scenarios. Therefore, understanding and mitigating the bias in LLMs is essential to enhance their availability to a wider range of practitioners and residents, and to ensure that these models can be applied equitably across diverse urban environments. In this regard, careful curation of training data and the implementation of algorithmic fairness techniques such as specialized fine-tuning[142] will be essential to mitigate the negative impacts of bias in urban applications. Moreover, future work should explore participatory design approaches that actively involve local stakeholders, ensuring that LLM-driven tools are not only technically robust but also socially inclusive and contextually grounded in diverse urban environments.

To address these limitations, an LLM-driven urban planning framework can be validated through extensive case studies and ablation analyses across diverse contexts, ranging from large metropolitan areas to smaller, less-represented cities, reducing the impact of data quality and biases on model performance and ensuring the equitable and effective deployment of LLMs in urban planning. Additionally, as benchmarking LLMs gains momentum in other fields[143,144], urban planning needs to develop its own benchmark to objectively evaluate LLM performance.

## Final remarks

The practical implementation of LLM-driven urban planning is not a simple technical issue, but involves challenges beyond technical feasibility and is also shaped by policy and societal considerations[145]. For instance, the integration of LLM-driven tools may require updates to existing planning policies and regulatory frameworks to accommodate new methodologies. In the meantime, the computational resources needed for implementing LLMs can limit accessibility–particularly for small municipalities or organizations with limited technical infrastructure. Additionally, the acceptance of LLM-driven urban planning depends on trust from multiple stakeholders including policymakers, urban designers, and the public. Planners may remain skeptical of AI-generated urban plans, raising concerns about their interpretability, transparency, and alignment with human-centered urban values. A collaborative approach, where human planners and LLMs work in synergy, can help leverage AI's computational strengths while ensuring that urban plans align with human values and societal needs. Future research should explore human-in-the-loop methods to ensure the transparent and trustworthy generation of urban plans by LLM, and provide actionable strategies to guide and regulate their usage in urban planning practices.

Planners translate human creativity into tangible urban designs that shape vibrant cities, and LLMs can provide planners with more effective tools to enhance creativity and boost productivity. Throughout history, the adoption of advanced technologies has redefined how urban spaces are utilized and how cities are designed. Our Perspective advocates for a collaborative workflow in which planners push the boundary of urban planning with LLMs connecting existing data, tools, and resources. We believe that the effective integration of LLMs promises to substantially benefit urban planning, paving the way for more efficient, inclusive, and sustainable cities that better serve the needs of their residents, and shaping a future where the potential of human ingenuity is fully realized in the urban environments we create.

## References


**1.** Un-Habitat. *World Cities Report 2020: The value of sustainable urbanization* (UN, 2020).

**2.** Su, H., Zheng, Y., Ding, J., Jin, D. & Li, Y. Large-scale urban facility location selection with knowledge-informed reinforcement learning. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, 553–556 (2024).

**3.** Xu, Y., Çolak, S., Kara, E. C., Moura, S. J. & González, M. C. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nat. Energy* **3**, 484–493 (2018).

**4.** Xu, Y., Olmos, L. E., Abbar, S. & González, M. C. Deconstructing laws of accessibility and facility distribution in cities. *Sci. advances* **6**, eabb4112 (2020).

**5.** Bettencourt, L. M. Urban growth and the emergent statistics of cities. *Sci. advances* **6**, eaat8812 (2020).

**6.** Batty, M. Big data, smart cities and city planning. *Dialogues human geography* **3**, 274–279 (2013).

**7.** Batty, M. *The new science of cities* (MIT press, 2013).

**8.** 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).

**9.** Nilforoshan, H. *et al.* Human mobility networks reveal increased segregation in large cities. *Nature* **624**, 586–592 (2023).

**10.** Xu, F. *et al.* Using human mobility data to quantify experienced urban inequalities. *Nat. Hum. Behav.* 1–11 (2025).

**11.** Testi, I. *et al.* Big mobility data reveals hyperlocal air pollution exposure disparities in the bronx, new york. *Nat. Cities* **1**, 512–521 (2024).

**12.** Collins, P. Y. *et al.* Making cities mental health friendly for adolescents and young adults. *Nature* **627**, 137–148 (2024).

**13.** Abbiasov, T. *et al.* The 15-minute city quantified using human mobility data. *Nat. Hum. Behav.* **8**, 445–455 (2024).

**14.** Moreno, C., Allam, Z., Chabaud, D., Gall, C. & Pratlong, F. Introducing the "15-minute city": Sustainability, resilience and place identity in future post-pandemic cities. *Smart cities* **4**, 93–111 (2021).

**15.** Allam, Z., Bibri, S. E., Chabaud, D. & Moreno, C. The '15-minute city'concept can shape a net-zero urban future. *Humanit. Soc. Sci. Commun.* **9**, 1–5 (2022).

**16.** Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–913 (2010).

**17.** Bettencourt, L. M. *Introduction to urban science: evidence and theory of cities as complex systems* (MIT Press, 2021).

**18.** Bettencourt, L. M. The kind of problem a city is: new perspectives on the nature of cities from complex systems theory. *Decod. City: How Big Data can Chang. Urban. Offenhuber, D., Ratti, C., Eds* 168–179 (2013).

**19.** Schläpfer, M. *et al.* The universal visitation law of human mobility. *Nature* **593**, 522–527 (2021).

**20.** Bongiorno, C. *et al.* Vector-based pedestrian navigation in cities. *Nat. computational science* **1**, 678–685 (2021).

**21.** Caldarelli, G. *et al.* The role of complexity for digital twins of cities. *Nat. Comput. Sci.* **3**, 374–381 (2023).

**22.** Batty, M. Digital twins in city planning. *Nat. Comput. Sci.* **4**, 192–199 (2024).

**23.** Bettencourt, L. M. Recent achievements and conceptual challenges for urban digital twins. *Nat. Comput. Sci.* **4**, 150–153 (2024).

**24.** Mohammadi, N. & Taylor, J. E. Thinking fast and slow in disaster decision-making with smart city digital twins. *Nat. Comput. Sci.* **1**, 771–773 (2021).

**25.** Brown, T. *et al.* Language models are few-shot learners. *Adv. neural information processing systems* **33**, 1877–1901 (2020).

**26.** Touvron, H. *et al.* Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

**27.** Liu, A. *et al.* Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437* (2024).

**28.** Pan, J. Large language model for molecular chemistry. *Nat. Comput. Sci.* **3**, 5–5 (2023).

**29.** Singhal, K. *et al.* Large language models encode clinical knowledge. *Nature* **620**, 172–180 (2023).

**30.** Luo, X. *et al.* Large language models surpass human experts in predicting neuroscience results. *Nat. human behaviour* 1–11 (2024).

**31.** Hagendorff, T., Fabi, S. & Kosinski, M. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in chatgpt. *Nat. Comput. Sci.* **3**, 833–838 (2023).

**32.** Guo, D. *et al.* Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948* (2025).

**33.** Hurst, A. *et al.* Gpt-4o system card. *arXiv preprint arXiv:2410.21276* (2024).

**34.** Jaech, A. *et al.* Openai o1 system card. *arXiv preprint arXiv:2412.16720* (2024).

**35.** Achiam, J. *et al.* Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

**36.** Chen, M. *et al.* Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).

**37.** Trinh, T. H., Wu, Y., Le, Q. V., He, H. & Luong, T. Solving olympiad geometry without human demonstrations. *Nature* **625**, 476–482 (2024).

**38.** Romera-Paredes, B. *et al.* Mathematical discoveries from program search with large language models. *Nature* **625**, 468–475 (2024).

**39.** Radford, A. *et al.* Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763 (PMLR, 2021).

**40.** Wang, W. *et al.* Visionllm: Large language model is also an open-ended decoder for vision-centric tasks. *Adv. Neural Inf. Process. Syst.* **36** (2024).

**41.** Esser, P. *et al.* Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning* (2024).

**42.** Song, Y., Dhariwal, P., Chen, M. & Sutskever, I. Consistency models. In *International Conference on Machine Learning*, 32211–32252 (PMLR, 2023).

**43.** Huang, S. *et al.* Diffusion-based generation, optimization, and planning in 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16750–16761 (2023).

**44.** Ye, J. *et al.* Dreamreward: Text-to-3d generation with human preference. In *European Conference on Computer Vision*, 259–276 (Springer, 2024).

**45.** Ho, J. *et al.* Video diffusion models. *Adv. Neural Inf. Process. Syst.* **35**, 8633–8646 (2022).

**46.** Shanahan, M., McDonell, K. & Reynolds, L. Role play with large language models. *Nature* **623**, 493–498 (2023).

**47.** Gao, C. *et al.* Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanit. Soc. Sci. Commun.* **11**, 1–24 (2024).

**48.** Shang, Y. *et al.* Agentsquare: Automatic llm agent search in modular design space. In *The Fourteenth International Conference on Learning Representations* (2025).

**49.** Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language models. *Nature* **624**, 570–578 (2023).

**50.** Joublin, F. *et al.* Copal: corrective planning of robot actions with large language models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 8664–8670 (IEEE, 2024).

**51.** Park, J. S. *et al.* Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, 1–22 (2023).

**52.** Li, N., Gao, C., Li, M., Li, Y. & Liao, Q. Econagent: Large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 15523–15536 (2024).

**53.** Howard, E., Osborn, F. J. & Mumford, L. *Garden cities of to-morrow* (Routledge, 2013).

**54.** Taylor, N. Anglo-american town planning theory since 1945: three significant developments but no paradigm shifts. *Plan. Perspectives* **14**, 327–345 (1999).

**55.** Burgess, E. W. The growth of the city: an introduction to a research project. In *The city reader*, 212–220 (Routledge, 2015).

**56.** Geddes, P. *Cities in evolution: an introduction to the town planning movement and to the study of civics* (London, Williams, 1915).

**57.** Caves, R. W. *Encyclopedia of the City* (Routledge, 2004).

**58.** Sitte, C. *The art of building cities: city building according to its artistic fundamentals* (Ravenio Books, 1945).

**59.** Lynch, K. *The image of the city* (MIT press, 1964).

**60.** Heraud, B. J. The new towns and london's housing problem. *Urban Stud.* **3**, 8–21 (1966).

**61.** Teaford, J. C. Urban renewal and its aftermath. *Hous. policy debate* **11**, 443–465 (2000).

**62.** Fuller, M. & Moore, R. *An Analysis of Jane Jacobs's The Death and Life of Great American Cities* (Macat Library, 2017).

**63.** Davidoff, P. Advocacy and pluralism in planning. *J. Am. Inst. planners* **31**, 331–338 (1965).

**64.** Friedmann, J. *Planning in the public domain: From knowledge to action* (Princeton University Press, 1987).

**65.** Harvey, D. *Social justice and the city*, vol. 1 (University of Georgia press, 2010).

**66.** Lindblom, C. E. Still muddling, not yet through. *Public administration review* **39**, 517–526 (1979).

**67.** Faludi, A. *A reader in planning theory*, vol. 5 (Elsevier, 2013).

**68.** Healey, P. Planning through debate: The communicative turn in planning theory. *Town planning review* **63**, 143 (1992).

**69.** Lane, M. B. Public participation in planning: an intellectual history. *Aust. geographer* **36**, 283–299 (2005).

**70.** Du, J., Ye, X., Jankowski, P., Sanchez, T. W. & Mai, G. Artificial intelligence enabled participatory planning: a review. *Int. J. Urban Sci.* **28**, 183–210 (2024).

**71.** Sanchez, T. W. Planning on the verge of ai, or ai on the verge of planning. *Urban Sci.* **7**, 70 (2023).

**72.** Dong, L. *et al.* Defining a city—delineating urban areas using cell-phone data. *Nat. Cities* **1**, 117–125 (2024).

**73.** Zheng, Y. *et al.* A survey of machine learning for urban decision making: Applications in planning, transportation, and healthcare. *ACM Comput. Surv.* **57**, 1–41 (2024).

**74.** Yuan, Y., Ding, J., Feng, J., Jin, D. & Li, Y. Unist: a prompt-empowered universal model for urban spatio-temporal prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4095–4106 (2024).

**75.** Fang, Z., Jin, Y. & Yang, T. Incorporating planning intelligence into deep learning: A planning support tool for street network design. *J. urban technology* **29**, 99–114 (2022).

**76.** Goodfellow, I. *et al.* Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).

**77.** Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings* (2014). http://arxiv.org/abs/1312.6114v10.

**78.** Fang, Z. *et al.* A framework for human-computer interactive street network design based on a multi-stage deep learning approach. *Comput. Environ. Urban Syst.* **96**, 101853 (2022).

**79.** Wang, D. *et al.* Deep human-guided conditional variational generative modeling for automated urban planning. In *2021 IEEE international conference on data mining (ICDM)*, 679–688 (IEEE, 2021).

**80.** Wang, D., Fu, Y., Wang, P., Huang, B. & Lu, C.-T. Reimagining city configuration: Automated urban planning via adversarial learning. In *Proceedings of the 28th international conference on advances in geographic information systems*, 497–506 (2020).

**81.** Wang, D. *et al.* Human-instructed deep hierarchical generative learning for automated urban planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 4660–4667 (2023).

**82.** Qin, Y., Zhao, N., Sheng, B. & Lau, R. W. Text2city: One-stage text-driven urban layout regeneration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 4578–4586 (2024).

**83.** He, L. & Aliaga, D. Globalmapper: Arbitrary-shaped urban layout generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 454–464 (2023).

**84.** Sutton, R. S. Reinforcement learning: An introduction. *A Bradf. Book* (2018).

**85.** Zheng, Y. *et al.* Spatial planning of urban communities via deep reinforcement learning. *Nat. Comput. Sci.* **3**, 748–762 (2023).

**86.** Su, H. *et al.* Reinforcement learning with adaptive reward modeling for expensive-to-evaluate systems. In *International Conference on Machine Learning* (PMLR, 2025).

**87.** Zheng, Y., Su, H., Ding, J., Jin, D. & Li, Y. Road planning for slums via deep reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5695–5706 (2023).

**88.** Wei, Y., Mao, M., Zhao, X., Zou, J. & An, P. City metro network expansion with reinforcement learning. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2646–2656 (2020).

**89.** Su, H., Zheng, Y., Ding, J., Jin, D. & Li, Y. Metrognn: Metro network expansion with reinforcement learning. In *Companion Proceedings of the ACM on Web Conference 2024*, 650–653 (2024).

**90.** Imani, S., Du, L. & Shrivastava, H. Mathprompter: Mathematical reasoning using large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, 37–42 (2023).

**91.** Thirunavukarasu, A. J. *et al.* Large language models in medicine. *Nat. medicine* **29**, 1930–1940 (2023).

**92.** Xiao, C., Hu, X., Liu, Z., Tu, C. & Sun, M. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open* **2**, 79–84 (2021).

**93.** Huang, A. H., Wang, H. & Yang, Y. Finbert: A large language model for extracting information from financial text. *Contemp. Account. Res.* **40**, 806–841 (2023).

**94.** Wei, J. *et al.* Emergent abilities of large language models. *Transactions on Mach. Learn. Res.* (2022).

**95.** Le, N.-T., Thai, N.-T. & Bui, C. V. Advancing urban development through vision-language models: Applications and challenges of satellite imagery analysis. In *2024 9th International Conference on Applying New Technology in Green Buildings (ATiGB)*, 184–188 (IEEE, 2024).

**96.** Beneduce, C., Lepri, B. & Luca, M. Urban safety perception through the lens of large multimodal models: A persona-based approach. *arXiv preprint arXiv:2503.00610* (2025).

**97.** Liang, X., Xie, J., Zhao, T., Stouffs, R. & Biljecki, F. Openfacades: An open framework for architectural caption and attribute data enrichment via street view imagery. *arXiv preprint arXiv:2504.02866* (2025).

**98.** Hong, W., Ding, M., Zheng, W., Liu, X. & Tang, J. Cogvideo: Large-scale pretraining for text-to-video generation via transformers. In *The Eleventh International Conference on Learning Representations* (2022).

**99.** (FAIR)†, M. F. A. R. D. T. *et al.* Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science* **378**, 1067–1074 (2022).

**100.** Lan, X., Gao, C., Jin, D. & Li, Y. Stance detection with collaborative role-infused llm-based agents. In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 18, 891–903 (2024).

**101.** Li, Z. *et al.* Urbangpt: Spatio-temporal large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5351–5362 (2024).

**102.** Lai, S., Xu, Z., Zhang, W., Liu, H. & Xiong, H. Large language models as traffic signal control agents: Capacity and opportunity. *arXiv preprint arXiv:2312.16044* (2023).

**103.** Feng, J. *et al.* Citygpt: Empowering urban spatial cognition of large language models. *arXiv preprint arXiv:2406.13948* (2024).

**104.** Fu, J., Han, H., Su, X. & Fan, C. Towards human-ai collaborative urban science research enabled by pre-trained large language models. *Urban Informatics* **3**, 8 (2024).

**105.** Zhang, Y., Wei, C., He, Z. & Yu, W. Geogpt: An assistant for understanding and processing geospatial tasks. *Int. J. Appl. Earth Obs. Geoinformation* **131**, 103976 (2024).

**106.** Zhang, W. *et al.* Urban foundation models: A survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6633–6643 (2024).

**107.** Dubey, R., Hardy, M. D., Griffiths, T. L. & Bhui, R. Ai-generated visuals of car-free us cities help improve support for sustainable policies. *Nat. Sustain.* **7**, 399–403 (2024).

**108.** Wang, D., Lu, C.-T. & Fu, Y. Towards automated urban planning: When generative and chatgpt-like ai meets urban planning. *arXiv preprint arXiv:2304.03892* (2023).

**109.** Wang, S. *et al.* Gpt, large language models (llms) and generative artificial intelligence (gai) models in geospatial science: a systematic review. *Int. J. Digit. Earth* **17**, 2353122 (2024).

**110.** Bieri, V., Zamboni, M., Blumer, N. S., Chen, Q. & Engelmann, F. Opencity3d: What do vision-language models know about urban environments? In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5147–5155 (IEEE, 2025).

**111.** Gurnee, W. & Tegmark, M. Language models represent space and time. In *The Twelfth International Conference on Learning Representations* (2024).

**112.** Manvi, R. *et al.* Geollm: Extracting geospatial knowledge from large language models. In *The Twelfth International Conference on Learning Representations* (2024).

**113.** Edge, D. *et al.* From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130* (2024).

**114.** Wei, J. *et al.* Chain-of-thought prompting elicits reasoning in large language models. *Adv. neural information processing systems* **35**, 24824–24837 (2022).

**115.** Zhong, T. *et al.* Evaluation of openai o1: Opportunities and challenges of agi. *arXiv preprint arXiv:2409.18486* (2024).

**116.** Manvi, R., Khanna, S., Burke, M., Lobell, D. & Ermon, S. Large language models are geographically biased. In *Proceedings of the 41st International Conference on Machine Learning*, 34654–34669 (2024).

**117.** Zheng, Y. *et al.* Urbanplanbench: A comprehensive urban planning benchmark for evaluating large language models. *arXiv preprint arXiv:2504.21027* (2025).

**118.** Yang, A. *et al.* Qwen2 technical report. *arXiv preprint arXiv:2407.10671* (2024).

**119.** Hu, E. J. *et al.* Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022* (OpenReview.net, 2022).

**120.** Shang, Y. *et al.* Urbanworld: An urban world model for 3d city generation. *arXiv preprint arXiv:2407.11965* (2024).

**121.** Zhang, L., Rao, A. & Agrawala, M. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3836–3847 (2023).

**122.** Ruiz, N. *et al.* Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 22500–22510 (2023).

**123.** Tzachor, A., Sabri, S., Richards, C. E., Rajabifard, A. & Acuto, M. Potential and limitations of digital twins to achieve the sustainable development goals. *Nat. Sustain.* **5**, 822–829 (2022).

**124.** Yuan, Y., Ding, J., Wang, H., Jin, D. & Li, Y. Activity trajectory generation via modeling spatiotemporal dynamics. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4752–4762 (2022).

**125.** Yuan, Y., Wang, H., Ding, J., Jin, D. & Li, Y. Learning to simulate daily activities via modeling dynamic human needs. In *Proceedings of the ACM Web Conference 2023*, 906–916 (2023).

**126.** Yuan, Y., Ding, J., Wang, H. & Jin, D. Generating daily activities with need dynamics. *ACM Transactions on Intell. Syst. Technol.* **15**, 1–28 (2024).

**127.** Yan, Y. *et al.* Opencity: A scalable platform to simulate urban activities with massive llm agents. *arXiv preprint arXiv:2410.21286* (2024).

**128.** Yuan, Y., Ding, J., Jin, D. & Li, Y. Learning the complexity of urban mobility with deep generative network. *PNAS nexus* **4**, pgaf081 (2025).

**129.** Jia, J. J., Yuan, Z., Pan, J., McNamara, P. & Chen, D. Decision-making behavior evaluation framework for llms under uncertain context. *Adv. Neural Inf. Process. Syst.* **37**, 113360–113382 (2024).

**130.** Feng, L., Li, B., Podobnik, B., Preis, T. & Stanley, H. E. Linking agent-based models and stochastic models of financial markets. *Proc. Natl. Acad. Sci.* **109**, 8388–8393 (2012).

**131.** Antonini, G., Bierlaire, M. & Weber, M. Discrete choice models of pedestrian walking behavior. *Transp. Res. Part B: Methodol.* **40**, 667–687 (2006).

**132.** Kaplan, J. *et al.* Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361* (2020).

**133.** Jean, N. *et al.* Combining satellite imagery and machine learning to predict poverty. *Science* **353**, 790–794 (2016).

**134.** Xue, J. *et al.* Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nat. Mach. Intell.* **4**, 246–257 (2022).

**135.** Fu, Y., Peng, H., Ou, L., Sabharwal, A. & Khot, T. Specializing smaller language models towards multi-step reasoning. In *International Conference on Machine Learning*, 10421–10430 (PMLR, 2023).

**136.** Narain, R., Golas, A., Curtis, S. & Lin, M. C. Aggregate dynamics for dense crowd simulation. *ACM Trans. Graph.* **28**, 122, DOI: 10.1145/1618452.1618468 (2009).

**137.** Xu, Y. *et al.* Urban dynamics through the lens of human mobility. *Nat. computational science* **3**, 611–620 (2023).

**138.** Xu, F., Li, Y., Jin, D., Lu, J. & Song, C. Emergence of urban growth patterns from human mobility behavior. *Nat. Comput. Sci.* **1**, 791–800 (2021).

**139.** Barreras, F. & Watts, D. J. The exciting potential and daunting challenge of using gps human-mobility data for epidemic modeling. *Nat. Comput. Sci.* 1–14 (2024).

**140.** Ling, L., Rabbi, F., Wang, S. & Yang, J. Bias unveiled: Investigating social bias in llm-generated code. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 27491–27499 (2025).

**141.** An, J., Huang, D., Lin, C. & Tai, M. Measuring gender and racial biases in large language models. *arXiv preprint arXiv:2403.15281* (2024).

**142.** Hu, T. *et al.* Generative language models exhibit social identity biases. *Nat. Comput. Sci.* **5**, 65–75 (2025).

**143.** Kirk, H. R. *et al.* The prism alignment dataset: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track* (2024).

**144.** Wang, B. *et al.* Decodingtrust: A comprehensive assessment of trustworthiness in gpt models. In *NeurIPS* (2023).

**145.** Sanchez, T. W., Brenman, M. & Ye, X. The ethical concerns of artificial intelligence in urban planning. *J. Am. Plan. Assoc.* 1–14 (2024).

## Acknowledgements

Y.Li. was supported by The National Key Research and Development Program of China under grant 2024YFC3307603.

## Author contributions statement

Q.R.W. and Y.Li. conceived the main theme of the research idea. Y.Z. wrote the first draft of the manuscript. Y.Lin, P.S., Q.R.W. and Y.Li provided crucial feedback. All authors reviewed, edited, and approved the final version of the manuscript.

## Ethics declaration

This paper does not involve any new experimental research on human subjects or animals, nor does it include the collection or analysis of new, identifiable private data. Therefore, no specific ethics committee approval was required for this study.

## Competing interests

The authors declare no competing interests.

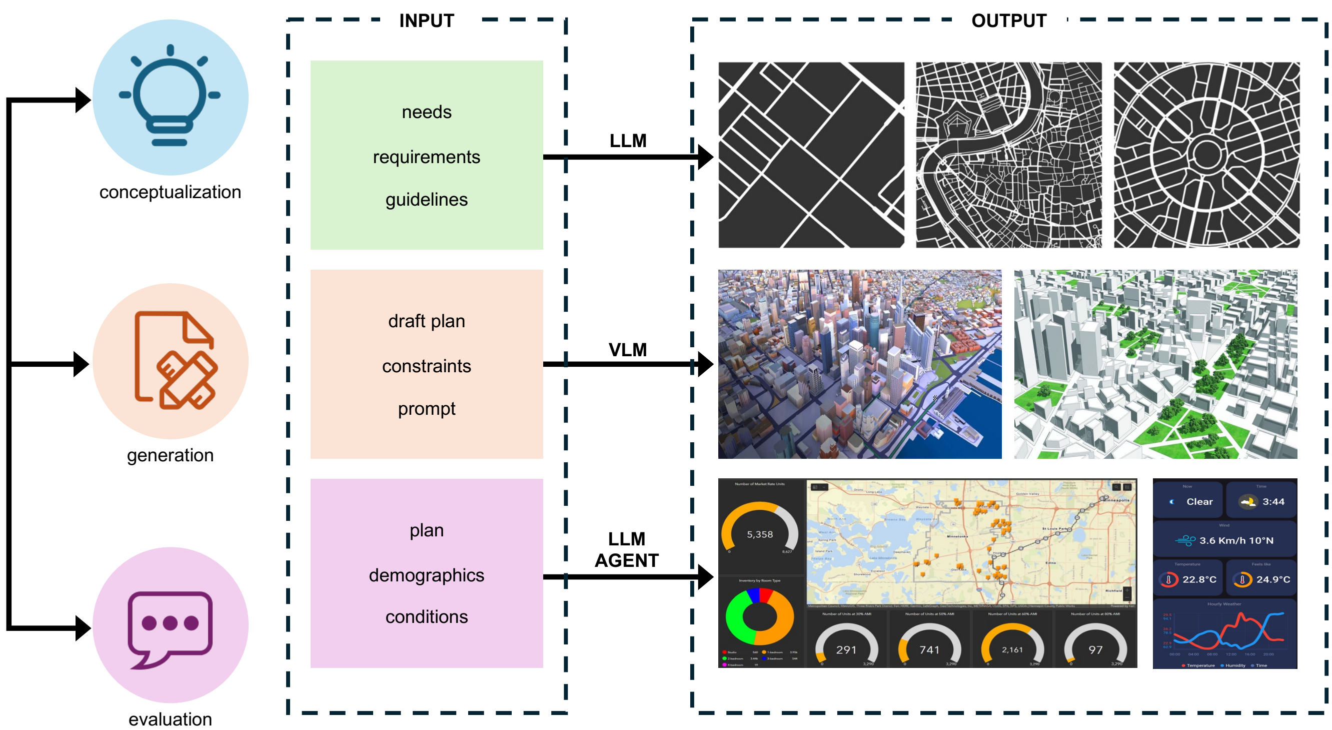


**Figure 1. The proposed LLM-driven urban planning framework.** The framework consists of three stages, including conceptualization, generation, and evaluation, driven by LLM, VLM, and LLM agent. In the conceptualization stage, LLMs produces conceptual ideas expressed with textual descriptions characterizing urban forms and functionalities. In the generation stage, VLMs generates specific and detailed urban visual designs such as layouts and cityscapes. In the evaluation stage, LLM agents simulate residents' activities and output quantitative assessments of the planning effect.

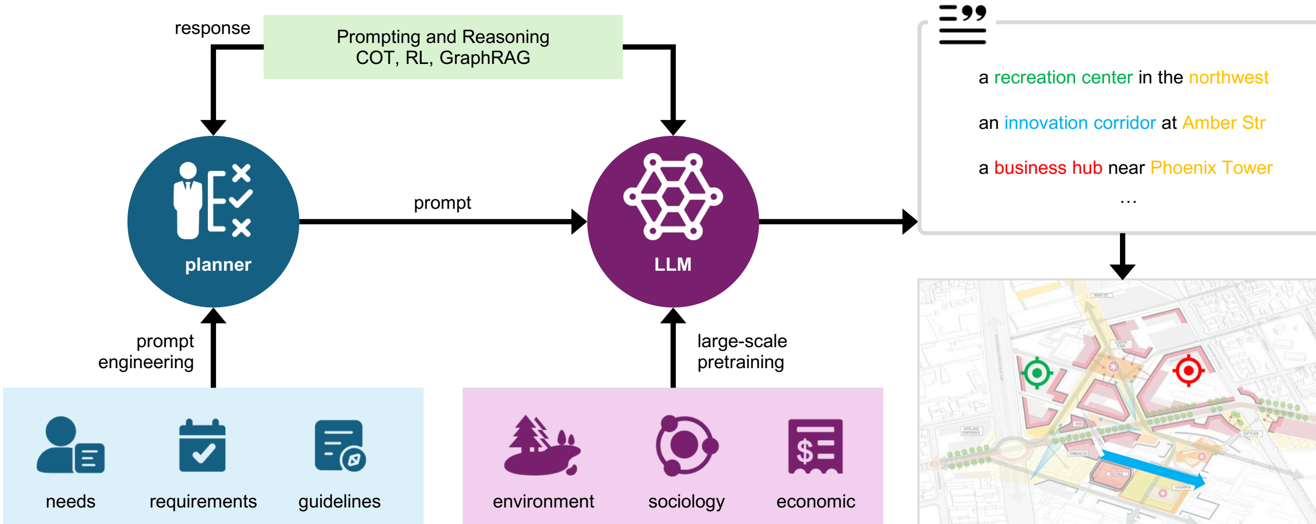


**Figure 2. Conceptualization in LLM-driven urban planning.** Planners conduct prompt engineering to devise informative prompts stating the needs, requirements, and guidelines of urban planning. LLMs pretrained on diverse domains of data are adopted to respond to planners' prompt, with enhanced prompting and reasoning strategies like GraphRAG and COT. Conceptual ideas expressed with textual descriptions are output to assist human planners in accomplishing draft plans.

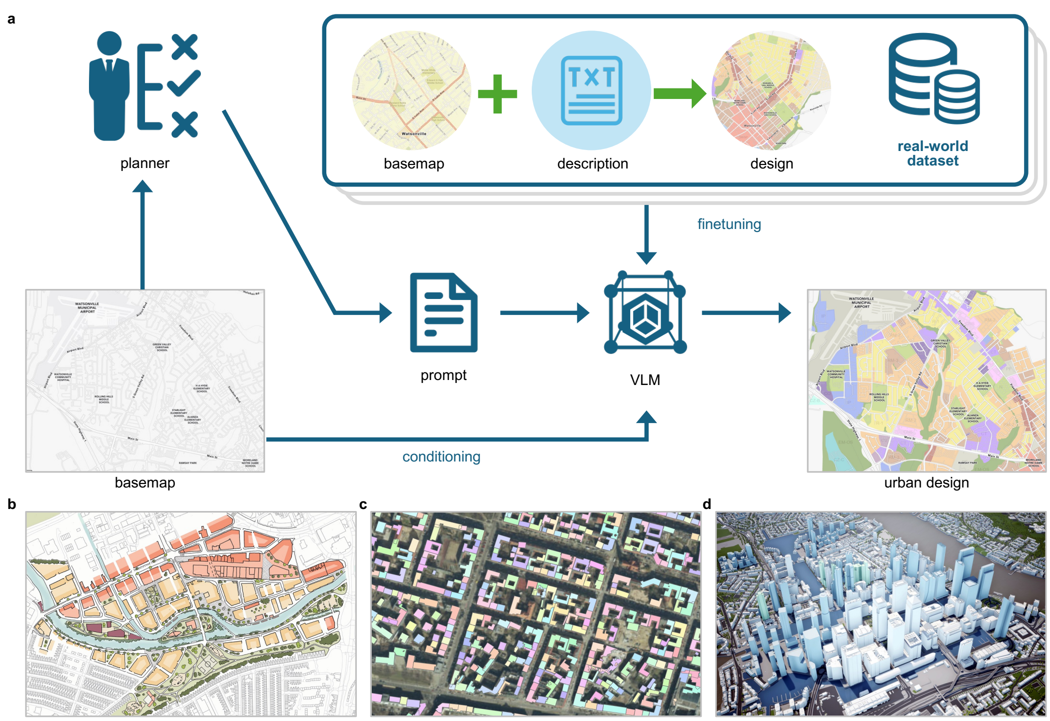


**Figure 3. Plan generation in LLM-driven urban planning. a.** Finetuning VLMs to develop urban design skills using large-scale datasets collected from real-world planning solutions containing imagery design and textual descriptions. Planners craft prompts describing their requirements and conceptual ideas, based on which finetuned VLMs generate urban design, conditioning on the base map to meet geospatial restrictions. **b-d.** Generated results from initial layout and planners' prompts by VLMs in different scenarios, including **b.** land use zoning layout, **c.** building footprint, and **d.** 3D urban scene.

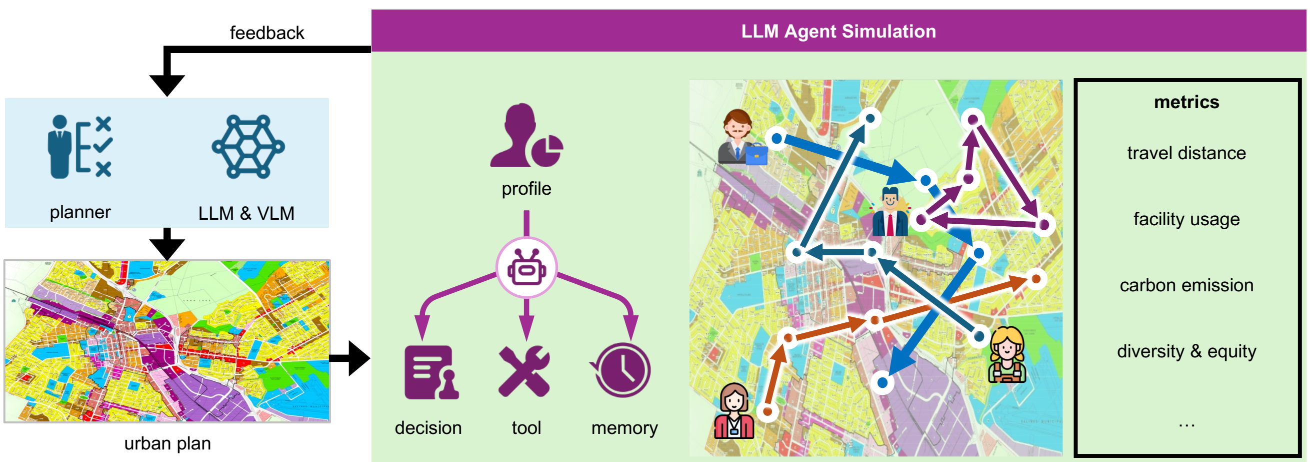


**Figure 4. Evaluation in LLM-driven urban planning.** LLM agent is adopted to simulate daily activities and mobility of community residents, given the generated urban plan. Agents make decisions on their activity/mobility according to the input personalized profile and their historical memory. Quantitative metrics such as travel distance and facility usage can be calculated based on the simulated results, providing tangible feedback to planners and LLM/VLM for further iteration.