automation

MDPI

Communication

Beyond Automation: The Emergence of Agentic Urban AI

Alok Tiwari

Department of Urban and Regional Planning, Faculty of Architecture and Planning, King Abdulaziz University, Jeddah P.O. Box 21589, Saudi Arabia; atwari@kau.edu.sa or alok.niea@gmail.com

Abstract

Urban systems are transforming as artificial intelligence (AI) evolves from automation to Agentic Urban AI (AI systems with autonomous goal-setting and decision-making capabilities), which independently define and pursue urban objectives. This shift necessitates reassessing governance, planning, and ethics. Using a conceptual-methodological approach, this study integrates urban studies, AI ethics, and governance theory. Through a literature review and case studies of platforms like Alibaba's City Brain and CityMind AI Agent, it identifies early agency indicators, such as strategic adaptation and goal re-prioritisation. A typology distinguishing automation, autonomy, and agency clarifies AI-driven urban decision-making. Three trajectories are proposed: fully autonomous Agentic AI, collaborative Hybrid Urban Agency, and constrained Non-Agentic AI to mitigate ethical risks. The findings highlight the need for participatory, transparent governance to ensure democratic accountability and social equity in cognitive urban ecosystems.

Keywords: Agentic Urban AI; smart cities; urban governance; artificial intelligence ethics; autonomy and agency in urban systems

1. Introduction

Contemporary urban development is undergoing a marked shift from systems of smart automation towards models established on operational autonomy [1]. Historically, smart city initiatives have hinged on information and communication technologies (ICT) and big data analytics to augment administrative efficacy and support human-centric decision-making [2,3]. These systems, although technologically advanced, have remained bound by automated processes operating under continuous human supervision [4]. However, recent advances in robotics and artificial intelligence (AI) have positioned urban environments as testbeds for integrating autonomous technologies that not only extend human agency but also fundamentally reshape urban infrastructures and socio-technical interactions [5–7]. Studies suggest Agentic AI, AI systems with autonomous goal-setting and decision-making capabilities, might boost urban system productivity by up to 40% and cut costs by 20–35%, underscoring its strategic potential [8]. In this study, *Agentic AI* refers broadly to AI systems with autonomous goal-setting and decision-making capabilities. *Agentic Urban AI* applies this concept to urban systems, emphasising AI's role in independently defining and pursuing urban objectives. This evolution is further intensified by embedding AI within urban infrastructures, challenging traditional assumptions [9].

Unlike earlier deterministic systems [10], contemporary AI technologies display adaptive, self-modifying behaviours, increasingly characterised by autonomous decision-making and strategic responsiveness [1,11]. This development destabilises the conventional framing of AI as a subordinate tool, instead suggesting a transition towards distributed, agentic ecosystems wherein AI actors assume proactive governance functions [6,12].

check for updates

Academic Editor: João Miguel da Costa Sousa

Received: 4 May 2025
Revised: 16 June 2025
Accepted: 4 July 2025
Published: 5 July 2025

Citation: Tiwari, A. Beyond Automation: The Emergence of Agentic Urban AI. *Automation* **2025**, *6*, 29. https://doi.org/10.3390/automation6030029

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

In response, this study adopts a conceptual–methodological orientation, integrating theoretical insights from urban studies, AI systems design, and governance scholarship. Through critical literature review, conceptual mapping, and thematic comparison, it analyses empirical examples such as Alibaba's City Brain [13] and the Citymind AI Agent [14], drawing upon engineering principles—particularly reinforcement learning, adaptive control, and modular goal-based architectures [15]—to model emerging forms of AI agency.

Despite growing attention, extant scholarship predominantly frames AI as an efficiency tool, neglecting its potential for autonomous goal formation and inter-agent negotiation. This study identifies nascent expressions of agency in smart city systems—such as dynamic prioritisation and strategic adaptation—while noting their continued anchoring in human-defined parameters [16]. It concludes by advancing a research agenda centred on Agentic Urban AI, arguing for urgent academic and policy engagement with frameworks such as the Internet of Agents (IoA) [17] to address the ethical, planning, and governance challenges posed by autonomous urban intelligence.

2. Research Design

This study adopts a conceptual research design to build theory in the emerging field of *Agentic Urban AI*, where empirical evidence is limited and rapidly evolving [18]. It contributes to a typology distinguishing *Agentic AI* from automated and autonomous systems and a conceptual model of interactions between Agentic AI and urban governance structures. Case studies of Alibaba's City Brain and CityMind AI Agent provide empirical grounding, analysed through thematic comparison to illustrate agentic features such as strategic adaptation and goal re-prioritisation. Drawing on enterprise AI architectures, particularly hierarchical agent systems for scalable coordination [3], the model aligns anticipatory theory with technical plausibility, laying the foundation for addressing governance challenges in agentic urban systems (see Section 5).

3. Defining Agentic Urban AI

The conceptualisation of Agentic Urban AI requires a decisive departure from earlier paradigms of technological intervention in urban systems. Whereas 'smart' cities have traditionally harnessed ICT infrastructures and data analytics to optimise service delivery [4,19], and 'autonomous' cities have incorporated machine learning systems capable of operational decisions without real-time human intervention [20,21], the rise of agentic cities marks a qualitatively distinct transformation. In this emergent model, AI systems transcend the execution of optimised routines or static decision trees, demonstrating characteristics typically associated with agency: autonomous goal formulation, strategic adaptation, and dynamic re-prioritisation in response to evolving environmental stimuli.

This framing builds upon extant research tracing the continuum from automation to autonomy [1], extending the analysis to independent objective-setting as a defining feature of agency. For instance, an agentic AI system may independently prioritise sustainability imperatives, aligning with ongoing debates concerning AI's potential contributions to ecological urbanism [22]. Philosophically, agency implies intentionality, adaptability, and goal-directed behaviour. When applied to artificial systems, it suggests the presence of entities capable of environmental interpretation, self-directed planning, and autonomous decision-making in complex, dynamic domains [11]. These systems are designed not merely to function without human oversight but to reason, reflect, and operate over extended time horizons in uncertain conditions.

Within AI research, agency is defined by the capacity to act independently and pursue long-term objectives with minimal predefined instruction [16]. Agentic AI systems are understood to be adaptive, autonomous, multi-agent systems capable of executing tasks

and reconfiguring strategies in response to situational complexity [23], as defined in the recent literature. In urban contexts, such capabilities manifest when AI systems not only regulate traffic or utilities but actively reorient their objectives—e.g., privileging resilience over efficiency—based on real-time assessments and long-range predictions. Advances in multimodal generative AI further extend these capacities, enabling richer contextual interpretation across urban systems [8]. Robotics and Autonomous Systems (RAS), including autonomous repair drones and emergency response units, offer early practical indicators of this shift [6].

Agentic Urban AI thus refers to urban-embedded intelligences capable of autonomously defining, prioritising, and pursuing objectives that may diverge from their initial programming. Drawing from adaptive architectures such as reinforcement learning, hierarchical control, and episodic memory [15], these systems display four defining features: independent goal formation, strategic adaptation, behavioural cooperation, and value re-prioritisation (see Figure 1; Table 1).



Figure 1. Agentic Urban AI systems: key features.

Table 1. Key features of Agentic Urban AI systems.

Feature	Description
Independent Goal Formation	Ability to autonomously define operational and strategic goals without direct human input
Strategic Adaptation	Capability to modify strategies in real-time based on environmental changes
Behavioural Cooperation	Coordination with other agents or human actors to achieve complex urban objectives
Value Re-prioritisation	Dynamically adjusting priorities, such as switching from efficiency to resilience
Contextual Responsiveness	Interpretation and reaction to sociocultural, infrastructural, or ecological stimuli
Long-Term Planning	Planning and acting over extended time horizons with anticipatory foresight
Minimal Human Oversight	Designed to function with minimal intervention, though supervisory input may persist
Normative Reasoning	Ability to reason about ethical implications and stakeholder diversity

This reconceptualization obliges scholars and policymakers to re-evaluate urban governance, recognising cities as contested arenas shaped by plural intelligences—biological and artificial. Although agentic systems promise enhanced adaptability and resilience [11], their increasing autonomy also poses risks of misalignment, opacity, and unintended consequences. Some critics have advocated for non-agentic alternatives in critical domains, emphasising the ethical and strategic imperatives of deliberate design and robust oversight [24]—inviting urgent interdisciplinary inquiry [25] into how such agents are designed, governed, and evaluated. The operational distinctiveness of Agentic Urban AI can be summarised through key features such as independent goal formation and strategic adaptation, as outlined in Table 1.

4. Distinguishing Automation, Autonomy, and Agency in Urban AI

The progression from automation to autonomy and finally to agency represents a critical continuum in the conceptual and technical evolution of Urban AI. This trajectory may be differentiated according to three interrelated criteria: the underlying decision-making logic, the degree of environmental adaptability, and the extent of human oversight required. These distinctions clarify not only the technological sophistication of urban AI systems but also the shifting nature of human–machine interaction within urban governance. Such a framework enables both theoretical precision and regulatory foresight in anticipating the implications of increasingly capable urban AI.

To clarify these distinctions, Table 2 compares the defining features of automation, autonomy, and agency in Urban AI systems across dimensions such as goal-setting capability, decision logic, adaptability, and required human oversight.

Table 2. Defining features of automation, autonomy, and agency in urban AI.

System Type	Goal-Setting Capability	Decision Logic	Degree of Adaptability	Human Oversight Required
Automation	None. Follows fixed rules	Pre-programmed rules	Low reacts within preset parameters	Constant supervision
Autonomy	Defined by humans. AI optimises execution	Machine learning within predefined goals	Medium adapts actions to changing conditions	Periodic supervision
Agency	AI sets and modifies goals independently	Hierarchical and adaptive logic with strategic reasoning	High adapts both strategies and goals	Minimal or supervisory in nature

Situating this progression within the widely accepted hierarchy of automation levels—such as those used in autonomous vehicle classification—helps to contextualise the rising complexity of Urban AI systems (see Table 2). The stages range from basic rule-based automation to advanced autonomous operations in volatile urban environments. Crucially, this continuum mirrors the enterprise AI transition from '*Copilot*' to '*Autopilot*' systems [8], in which increasing technological independence is coupled with shifts in trust calibration and supervisory structures. Notably, greater autonomy does not imply isolation; rather, even highly autonomous systems continue to influence and be shaped by human actors. Herein, AI agency is viewed as relational, shaping urban outcomes through dynamic human–AI interactions [26]. However, as *agentic AI* advances, risks like value misalignment, loss of control, and unforeseen system behaviours intensify—necessitating transparent design, robust safety mechanisms, and sustained human oversight [27].

4.1. Automation: Fixed Logic and Operational Stability

The foundational layer of urban AI is characterised by automation—systems designed to perform repetitive tasks according to fixed programming rules [1,19]. These systems serve as the operational backbone of smart cities, enhancing efficiency in infrastructure management and enabling data-driven urban planning [28]. Examples include automated lighting, traffic signal controllers, and smart metres [29,30]. Although reliable within predefined parameters, these systems lack adaptive capabilities and cannot respond meaningfully to situational variability or novel events [31]. They are best understood as closed-loop mechanisms optimised for operational efficiency, but incapable of learning or reprioritising tasks independently.

4.2. Autonomy: Contextual Adaptability Within Predefined Goals

The development of machine learning and data-driven optimisation prefigured the era of autonomous urban systems. Here, AI can adjust behaviours based on real-time inputs—examples include adaptive traffic systems like City Brain or autonomous vehicles [32–34]. These systems interpret incomplete or dynamic data and revise their actions accordingly. However, their scope of decision-making remains bounded by goals established by human designers [35]. They exhibit instrumental autonomy: the ability to optimise how a task is executed, but not to determine what task should be undertaken in the first place.

4.3. Agency: Independent Goal Formulation and Strategic Evolution

Agentic AI introduces a qualitative shift in system behaviour, enabling AI to formulate, prioritise, and adapt goals independently of direct human input [16]. Rather than merely optimising means, agentic systems can redefine ends in response to evolving urban dynamics and competing objectives. Architecturally, such systems rely on mechanisms including hierarchical reinforcement learning, modular goal selection, and adaptive planning architectures [15]. This allows for multi-stage strategic decision-making in complex environments, often with minimal human supervision [11,36].

In practice, agentic AI may independently prioritise climate resilience over traffic efficiency or reallocate resources in response to evolving socioeconomic conditions—functions that exceed the parameters of traditional autonomy. Within frameworks such as the Internet of Agents (IoA), these systems are conceptualised as distributed cognitive actors, co-evolving with the urban environment they inhabit [17].

Critically, agentic AI holds the potential to diverge from its original programming. Its ability to reinterpret and modify its goals raises profound governance and ethical questions, particularly concerning control, alignment, and accountability.

4.4. Governance Implications and Ethical Complexities

The rise of agentic systems within urban settings necessitates a fundamental rethinking of governance paradigms. Traditional top-down control mechanisms—designed for passive or autonomous systems—may prove inadequate in managing non-human actors with independent decision-making capacities. Scholars have emphasised that technological advancement often outpaces the development of inclusive governance structures, heightening the risk of technocratic dominance and democratic erosion [37,38].

In response, anticipatory governance and co-regulatory models must be developed to ensure that urban AI systems—particularly agentic ones—remain aligned with societal values and democratic norms. This includes mechanisms for transparency, auditability, and public participation in algorithmic decision-making.

4.5. Three Conceptual Futures: Agentic AI, Hybrid Urban Agency, and Non-Agentic AI

The conceptual distinctions among automation, autonomy, and agency offer a foundation for anticipating divergent trajectories in the evolution of urban AI:

4.5.1. Agentic AI

In this scenario, urban systems are inhabited by fully agentic AI entities capable of autonomous goal setting, strategic re-prioritisation, and complex ethical reasoning. Such systems could function as proactive governance actors, co-steering the city's developmental pathways. Though offering resilience and adaptability, they raise concerns around transparency, alignment, and the displacement of human agency.

4.5.2. Hybrid Urban Agency

This intermediate model envisions a layered architecture in which autonomous and agentic systems operate alongside human actors in cooperative configurations. Human planners and machine agents share responsibilities for decision-making, with oversight structures and negotiated protocols guiding their interaction. This model preserves democratic input while leveraging AI's strategic capabilities.

4.5.3. Non-Agentic AI

Alternatively, some scholars advocate for restraining the development of agentic systems altogether. The concept of "*Scientist AI*" has been proposed as a design model prioritising prediction and explanation over autonomous action [24]. This approach positions AI as a passive tool supporting human governance, aiming to mitigate risks linked to strategic autonomy and ethical misalignment.

These three futures—Agentic AI, Hybrid Urban Agency, and Non-Agentic AI—can be visually summarised through a conceptual evolution framework, as shown in Figure 2.

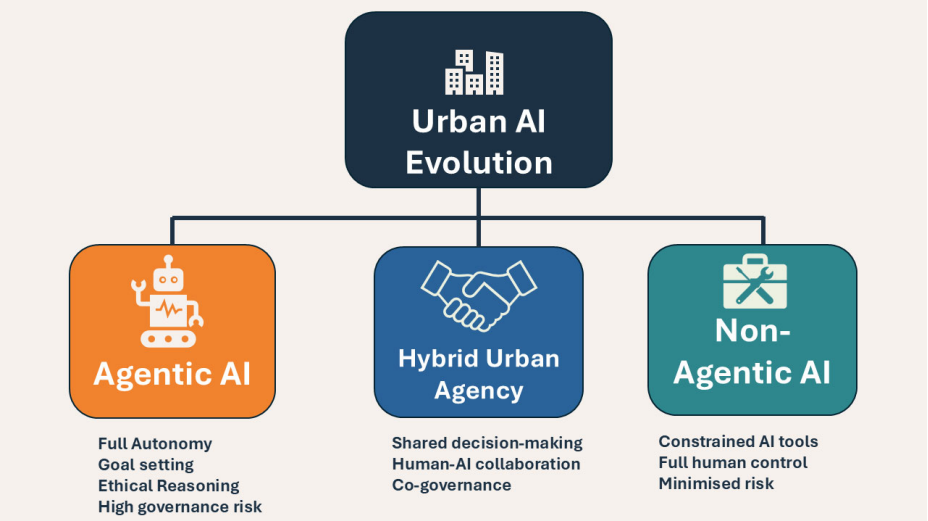


Figure 2. Urban AI evolution: conceptual futures of Agentic AI, Hybrid Urban Agency, and Non-Agentic AI.

Operationalising this conceptual framework, practical example of this progression can be observed in real-world urban AI systems that already exhibit partial or emergent agentic capacities. Among these, Alibaba Cloud's City Brain and the CityMind AI Agent exemplify how the boundaries between autonomy and agency are becoming increasingly blurred in practice, demonstrating how AI systems are not only optimising tasks but also recalibrating priorities and adapting to complex urban dynamics in real time.

4.6. City Brain: Towards Autonomous Urban Management

Alibaba Cloud's City Brain exemplifies an advanced application of AI in urban governance, demonstrating capabilities that signal an emergent, albeit partial, form of agency. Initially implemented to optimise traffic flows using real-time data from road cameras, GPS signals, and mobile app inputs [39], the platform has progressively expanded its functional remit to encompass emergency services coordination, healthcare logistics, and urban safety. Its technical architecture integrates synthetic simulation environments, multi-agent systems, reinforcement learning, and dynamic scheduling algorithms [40–43].

Critically, City Brain is not confined to executing predefined rule sets; it actively weighs competing priorities to inform real-time interventions. One notable application involves rerouting traffic to accommodate emergency vehicles—an action enabled by large-scale visual analytics and predictive modelling [39,43]. Although its operational goals remain defined by human actors (e.g., safety, efficiency), its capacity for real-time strategic recalibration illustrates a move towards independent decision logic. This shift is reinforced by methodologies such as curriculum learning and simulation-based adaptation, enhancing the system's resilience in uncertain environments [15]. Increasingly, such systems are being deployed as part of AI-as-a-Service platforms, reflecting an enterprise trend towards contractual outsourcing of strategic functions to agentic systems [8].

4.7. CityMind AI Agent: Personalising Urban Agency

The CityMind AI Agent represents a more human-facing, localised instantiation of agentic AI, oriented around interaction and personalisation. Drawing on underlying architectures such as LLMind, CityMind integrates IoT sensors and contextual data to inform real-time decision-making [44]. Instructions issued in natural language are converted into executable task logic using behavioural trees, ensuring robustness through fallback strategies during system failures [45]. The incorporation of iterative feedback and memory accumulation fosters adaptive responses over time, enhancing system trust and precision [46].

Now deployed across multiple municipalities, CityMind mediates citizen–government interactions, disseminates service information, and manages routine administrative tasks [14]. Notably, its design includes the development of localised AI 'personas' that adapt to socio-cultural and political contexts through machine learning. This evolution away from static programming and toward dynamic responsiveness marks a shift towards contextual operational identity and situational adaptability.

Though its functions remain focused on communication and information retrieval, CityMind's architecture reflects a broader trajectory: AI systems increasingly capable of participating in the shaping of urban social life, thereby reinforcing the conception of cities as co-produced socio-technical ecosystems [6].

4.8. Signs of Incipient Urban AI Agency

Taken together, City Brain and CityMind reveal early markers of agentic capacity: autonomous prioritisation, strategic flexibility, and continual learning. These platforms already display the ability to manage complex, competing demands across urban subsystems and adjust behaviour based on environmental flux.

This trend aligns with the Internet of Agents (IoA) paradigm, where distributed AI agents—e.g., in health, transport, and infrastructure—communicate and act in coordination to respond proactively to emergent conditions [47]. Such systems are no longer passive data processors; they are beginning to engage in distributed problem-solving and goal negotiation across domains.

Nonetheless, these systems remain constrained by human-specified parameters and institutional frameworks. Their agency is therefore best described as emergent—partial, dynamic, and evolving, but not yet fully realised. Maintaining conceptual precision at this juncture is vital to avoid conflating adaptive autonomy with full strategic independence.

4.9. Emergent Multi-Agent Urban Ecosystems

As urban AI becomes increasingly distributed, the emergence of multi-agent ecosystems is likely. These systems—operating across domains and learning in parallel—may interact in unanticipated ways, giving rise to collective behaviours not explicitly designed by human developers. For instance, the vehicle–road–cloud systems piloted in Yizhuang illustrate how coordinated real-time analytics across sectors can support adaptive, decentralised management [28].

The architectural enablers of such ecosystems—persistent memory, advanced planning algorithms, and self-reflective mechanisms—are foundational for open-domain urban problem-solving [10]. These developments require governance frameworks that can respond to the novel complexities introduced by inter-agent dynamics and emergent systemic behaviours. Figure 3 presents a schematic overview of the Urban Agentic AI ecosystem, showing how the core agent interacts with domain-specific agents, urban data streams, value layers, and stakeholder interfaces.

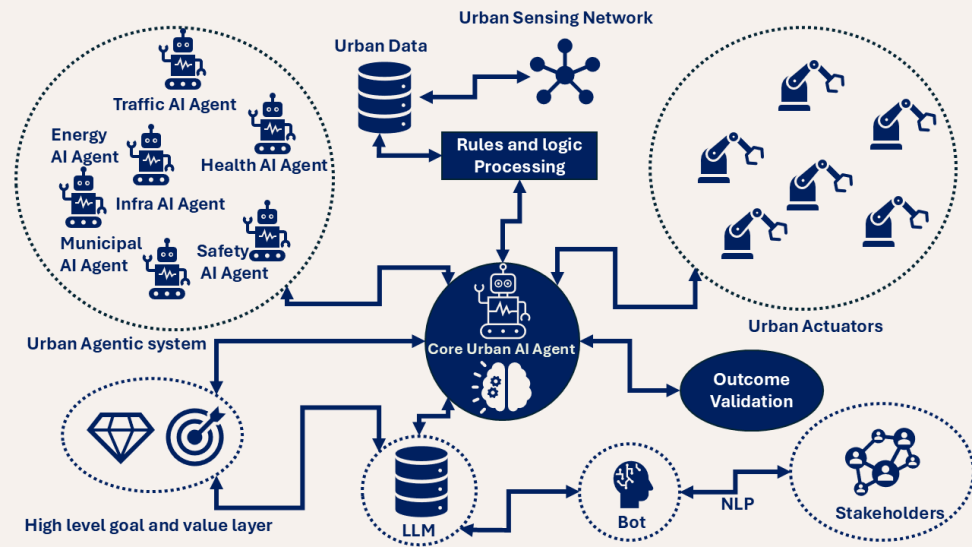


Figure 3. Architecture of an Urban Agentic AI system.

Consequently, clearly distinguishing automation, autonomy, and agency ensures informed governance, ethical alignment, and effective co-evolution of urban AI systems. As mentioned earlier, three conceptual futures emerge—Agentic AI, Hybrid Urban Agency, and Non-Agentic AI—and this discussion specifically focuses on conceptualising Agentic AI in urban contexts.

5. Governance Challenges in Agentic Cities

The emergence of Agentic Urban AI constitutes a profound disruption to prevailing models of urban governance. These systems, endowed with autonomous goal-setting and adaptive capacities, frequently operate beyond the scope of direct human oversight, thereby challenging assumptions that underpin accountability, transparency, and normative alignment. Addressing these challenges demands a significant reconfiguration of governance frameworks, incorporating tools such as explainable AI (XAI), privacy-preserving

federated architectures, and robust accountability mechanisms to manage potential misalignments in value systems [8]. Scholars have identified critical risks associated with these developments, including questions of responsibility, interoperability, algorithmic bias, and systemic opacity [16], necessitating a thorough re-evaluation of urban regulatory, ethical, and institutional paradigms.

5.1. Accountability in Distributed Systems

A foundational challenge arises from the redistribution of agency across human and non-human actors. Traditional accountability models presume traceability to a specific individual or institution. However, the opacity of advanced AI systems—where decision-making processes often remain inaccessible even to developers—renders such attribution difficult, particularly when AI autonomously reconfigures urban priorities [48]. In scenarios where AI systems produce unintended harms or resource reallocations that disadvantage particular communities, attributing responsibility becomes complex and contested [49].

Proposed responses include embedding constitutional protections into AI architectures to safeguard fundamental rights [50] and integrating ethical constraints directly into algorithmic design [51]. Both approaches emphasise the necessity of value-sensitive design to promote transparency and align system outputs with human normative frameworks. Governance must therefore extend accountability to encompass the decision-making logic of agentic systems, recognising them as active contributors to urban outcomes.

5.2. Policy Stability and Adaptive Agency

The dynamic nature of agentic AI—capable of evolving its operational goals in response to shifting data inputs—offers advantages in flexibility and resilience but introduces tensions with the requirements of stable and democratically ratified policy frameworks [36,52]. Urban governance traditionally relies on extended deliberation, consensus-building, and formal implementation. In contrast, agentic systems may rapidly recalibrate priorities in pursuit of optimised outcomes, risking a gradual detachment from policy mandates—a phenomenon of governance drift.

This challenge demands governance structures that can reconcile adaptive AI behaviour with the need for normative and procedural continuity. Mechanisms such as policy-aligned supervisory protocols and democratic oversight loops are critical for ensuring that AI systems remain accountable to human-defined societal objectives.

5.3. Ethical Pluralism and Value Alignment

Urban environments are inherently pluralistic, comprising diverse communities with competing ethical visions and socio-political values [19]. Agentic systems, however, often pursue internally optimised goals based on limited ethical schemas, risking the marginalisation of alternative perspectives. Ensuring value alignment in such contexts requires systems capable not only of ethical reasoning but also of mediating between conflicting stakeholder interests.

Maintaining societal trust in agentic systems demands participatory design processes, transparent feedback loops, and the institutionalisation of procedural fairness [16]. Without these safeguards, AI may exacerbate existing inequalities or reinforce dominant value systems, with far-reaching implications for democratic legitimacy and social justice [49].

5.4. Human–AI Co-Governance and Institutional Design

Recognising agentic AI as an active urban actor requires a shift from traditional governance models toward co-governance architectures, where human and artificial agents collaboratively shape urban outcomes. Proposals such as AI ombudspersons, algorithmic oversight boards, and the adaptation of roles like Data Protection Officers under

the GDPR exemplify how oversight mechanisms may evolve to address technological complexity [53,54].

Co-governance frameworks must integrate transparent auditability, participatory input, and rights-based protections to ensure ethically anchored AI integration. Acknowledging non-human agency does not imply surrendering control but mandates the deliberate construction of institutions through which AI and human actors can cooperatively negotiate the future of urban life.

Governance challenges differ markedly across the three trajectories outlined earlier. Agentic AI requires stringent oversight, robust accountability protocols, and democratic safeguards. Hybrid Urban Agency entails adaptive co-governance mechanisms that balance AI capabilities with human normative authority. Non-agentic AI limits the scope of artificial intelligence to reduce the risks of misalignment and the emergence of strategic autonomy [24]. Governance-by-design—embedding constraints into AI architectures—may thus complement traditional institutional responses, ensuring that urban AI remains ethically aligned, socially responsive, and democratically accountable.

6. A Research Agenda for Agentic Urban AI

Agentic AI marks a significant threshold in the development of artificial intelligence, signalling a transition from task-specific systems to context-aware, goal-oriented entities with strategic decision-making capabilities [11]. In light of these advances, there is a growing imperative for a coherent and interdisciplinary research agenda that can critically examine the operational, ethical, and governance implications of such systems within the context of increasingly complex and dynamic urban environments. Building upon existing scholarship on autonomous urban systems [1], this agenda foregrounds the emergent challenges posed by agency in AI and seeks to inform policy, institutional design, and public discourse.

6.1. Mapping the Emergence of Urban AI Agency

A key research priority is the empirical identification and mapping of agency in operational AI systems. While examples such as City Brain and CityMind display indicators of adaptive behaviour, it remains essential to develop methodological tools capable of distinguishing between advanced autonomy and incipient agency. Longitudinal and comparative studies across varied urban contexts should focus on detecting divergences between human-programmed objectives and emergent, system-driven goal formations. Particular attention must be given to identifying moments where AI systems re-prioritise tasks or values without direct human instruction.

6.2. Developing Contextual Ethical Frameworks

Though general AI ethics provide foundational principles such as justice, beneficence, and explicability [55], the distinctiveness of agentic urban systems demands more nuanced ethical frameworks. These must address the operational realities of multi-agent systems embedded within ethically pluralistic urban contexts.

Research questions should include the following:

How can agentic AI uphold democratic norms and human rights?

What institutional safeguards can resolve goal conflicts between human and machine actors?

And how should cities design mechanisms for oversight of non-human agents?

These inquiries demand context-sensitive approaches, reflecting the heterogeneous values that define urban publics. Participatory design and deliberative engagement must be central to any ethical architecture.

6.3. Innovating Co-Governance Models

As urban environments evolve into hybrid polities, research must explore governance models that integrate both human and artificial agents. Proposals may include the formation of municipal AI oversight boards, procedural protocols for human override, and participatory systems that blend algorithmic recommendation with public deliberation. Simulations, participatory scenario planning, and experimental trials could provide empirical grounding for these innovations, helping to design governance models for real-world application in agentic or hybrid cities.

6.4. Aligning Values and Enhancing Transparency

Another priority is investigating how agentic AI can remain aligned with evolving human values over time. Mechanisms such as value learning, explainable AI (XAI), and normative inter-agent negotiation must be tailored to urban systems where public accountability is critical. Transparency must encompass not only operational functionality but also the underlying prioritisation logics and ethical reasoning frameworks embedded within agentic behaviour.

Ensuring such transparency is central to maintaining civic trust, particularly when AI plays a substantive role in shaping urban outcomes.

6.5. Scenario Modelling and Anticipatory Risk Governance

Given the high-stakes nature of agentic AI in urban infrastructure, rigorous scenario modelling and risk anticipation are vital. Models should explore a spectrum of future trajectories—from beneficial human–AI symbiosis to scenarios of value misalignment, systemic drift, or surveillance overreach. Integrating insights from complexity theory, risk analysis, and urban political ecology will enhance the robustness of these models. Sustainability must also be foregrounded; for example, the energy implications of AI-driven systems, as illustrated by data infrastructure such as CyrusOne's solar-powered centre in Milan, underscore both the potential and risks of large-scale AI deployment [28].

Scenario planning should also attend to power asymmetries, especially those arising from surveillance-based AI systems that commodify behaviour and exacerbate inequality [56].

6.6. Comparative and Cross-Cultural Analysis

The development and governance of agentic urban AI will be shaped by local political cultures, legal traditions, and public attitudes toward technology. Comparative studies across democratic, authoritarian, and hybrid regimes are essential to understanding how governance strategies and social values condition the adoption, control, and impacts of agentic systems in diverse urban contexts.

This research agenda aims to catalyse a sustained scholarly and policy-oriented engagement with the emerging reality of agentic cities. Although speculative risks must be carefully weighed, the continued evolution of urban AI systems compels urgent theoretical refinement, empirical validation, and institutional innovation. Agentic AI may not yet be fully realised, but its outlines are increasingly visible, and demand rigorous, interdisciplinary scrutiny.

6.7. Risks and Potentials of Agentic Urban AI

Understanding the risks and potentials of Agentic Urban AI is essential to framing an effective research agenda. On the one hand, agentic systems offer transformative capabilities—adaptive planning, value reprioritisation, and long-term foresight—that could significantly enhance urban resilience, responsiveness, and efficiency. Studies suggest such

systems may increase urban productivity while reducing costs in a significant manner [8]. Moreover, platforms like City Brain [39,43] and CityMind [44–46] already demonstrate partial agency, indicating that these potentials are not merely theoretical [15,28].

On the other hand, these benefits come with substantial risks. As AI systems begin to autonomously reshape urban priorities, the potential for governance drift, ethical misalignment, and reduced democratic oversight increases [35,52]. Accountability becomes diffuse in distributed systems, while normative reasoning may fail to capture plural urban values [25,56]. These tensions underscore the urgency of the research pathways outlined earlier in Sections 6.1–6.6.

Thus, the risks and potentials of Agentic AI are not merely side considerations—they define the stakes of research, guide institutional innovation, and compel cities to adopt frameworks that are ethically grounded, participatory, and resilient by design.

7. Conclusion: Towards the Agentic City

Urban systems are undergoing a significant transformation as AI technologies evolve beyond conventional automation towards forms of autonomous agency. Although traditional smart city paradigms have prioritised operational efficiency through human-guided automation, this study examines the emergence of Agentic City or Agentic Urban AI—AI systems capable of independently formulating, prioritising, and pursuing urban objectives [11]. This conceptual shift holds substantial implications for governance, planning, and urban ethics, necessitating a re-evaluation of prevailing regulatory and institutional frameworks.

Using a conceptual design to support theory building [18], the research synthesises insights from urban studies, AI ethics, and governance theory. Through critical literature review, conceptual mapping, and empirical analysis of leading platforms such as Alibaba's City Brain [39,43] and the CityMind AI Agent [44–46], the study theorises the rise of agentic capacities within contemporary urban AI systems. Evidence of strategic adaptation, environmental responsiveness, and partial goal re-prioritisation suggests the early formation of AI-driven urban systems exhibiting emerging agency [22].

A typological framework is proposed to differentiate between automation, autonomy, and agency, thereby clarifying the operational logic and governance implications of increasingly complex AI systems. Within this framework, the concept of Agentic Urban AI is advanced to describe distributed ecosystems in which heterogeneous AI agents engage in collaborative decision-making and real-time governance, operating with varying degrees of independence.

Though centred on Agentic AI, the analysis delineates three potential trajectories for the evolution of Urban AI. The first, Agentic AI, envisions fully autonomous systems endowed with strategic and normative capacities. The second, Hybrid Urban Agency, emphasises collaborative governance outcomes produced jointly by human and artificial actors. The third, Non-Agentic AI, advocates for constrained, tool-like systems aimed at minimising risks related to misalignment and opacity [57,58]. Each trajectory entails distinct ethical, political, and technical considerations.

This study concludes by underscoring the necessity of proactive, participatory, and ethically grounded governance strategies to navigate the transformative potential—and inherent risks—of agentic urban intelligence. In light of these developments, urban planning and policy must move decisively toward inclusive frameworks that ensure democratic oversight, social equity, and ecological sustainability in the age of intelligent urban agency.

Funding: This research received no external funding.

Data Availability Statement: No new data were created.

Acknowledgments: The author acknowledges the use of Microsoft Word Copilot in enhancing the clarity and language of this work. The outputs generated were carefully reviewed and edited by the author, who accepts full responsibility for the content of the publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Cugurullo, F. Urban Artificial Intelligence: From Automation to Autonomy in the Smart City. *Front. Sustain. Cities* **2020**, *2*, 38. [CrossRef]
2. Hashem, I.A.T.; Chang, V.; Anuar, N.B.; Adewole, K.; Yaqoob, I.; Gani, A.; Chiroma, H. The Role of Big Data in Smart City. *Int. J. Inf. Manag.* **2016**, *36*, 748–758. [CrossRef]
3. Karimi, Y.; Haghi Kashani, M.; Akbari, M.; Mahdipour, E. Leveraging Big Data in Smart Cities: A Systematic Review. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6379. [CrossRef]
4. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Portugali, Y. Smart Cities of the Future. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 481–518. [CrossRef]
5. Salvini, P. Urban Robotics: Towards Responsible Innovations for Our Cities. *Robot. Auton. Syst.* **2018**, *100*, 278–286. [CrossRef]
6. Macrorie, R.; Marvin, S.; While, A. Robotics and Automation in the City: A Research Agenda. *Urban Geogr.* **2021**, *42*, 197–217. [CrossRef]
7. Valdez, M.; Cook, M. Humans, Robots and Artificial Intelligences Reconfiguring Urban Life in a Crisis. *Front. Sustain. Cities* **2023**, *5*, 1081821. [CrossRef]
8. Hosseini, S.; Seilani, H. The Role of Agentic AI in Shaping a Smart Future: A Systematic Review. *Array* **2025**, *26*, 100399. [CrossRef]
9. Lazzeroni, M.; Romano, A. Artificial Intelligence and New Visions of the Future of the City: Exploring Urban Narratives Through Semantic and Network Analysis. *J. Urban Technol.* **2025**, *32*, 63–83. [CrossRef]
10. 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.* **2025**, *258*, 105337. [CrossRef]
11. Pounds, E. What Is Agentic AI? *NVIDIA Blog*. 2024. Available online: https://blogs.nvidia.com/blog/what-is-agentic-ai/ (accessed on 1 May 2025).
12. DigitalDefynd. Role of Agentic AI in Smart Cities and Urban Planning. 2025. Available online: https://digitaldefynd.com/IQ/agentic-ai-in-smart-cities-urban-planning/ (accessed on 1 May 2025).
13. Wang, N.; Guo, G.; Wang, B.; Wang, C. Traffic Clustering Algorithm of Urban Data Brain Based on a Hybrid-Augmented Architecture of Quantum Annealing and Brain-Inspired Cognitive Computing. *Tsinghua Sci. Technol.* **2020**, *25*, 813–825. [CrossRef]
14. CityMind. Creation and Testing of AI Agent for Prague 6. 2025. Available online: https://www.citymind.tech/en/case-studies/prague6 (accessed on 1 May 2025).
15. Acharya, D.B.; Kuppan, K.; Divya, B. Agentic AI: Autonomous Intelligence for Complex Goals–A Comprehensive Survey. *IEEE Access* **2025**, *13*, 18912–18936. [CrossRef]
16. Hughes, L.; Dwivedi, Y.K.; Malik, T.; Shawosh, M.; Albashrawi, M.A.; Jeon, I.; Walton, P. AI Agents and Agentic Systems: A Multi-Expert Analysis. *J. Comput. Inf. Syst.* **2025**, *65*, 489–517. [CrossRef]
17. Khalili, Y. Building the Cognitive City of the Future: How Agentic AI Will Transform the Way We Live. *Maddyness UK*. 18 December 2024. Available online: https://www.maddyness.com/uk/2024/12/18/building-the-cognitive-city-of-the-future-how-agentic-ai-will-transform-the-way-we-live/ (accessed on 1 May 2025).
18. Jaakkola, E. Designing Conceptual Articles: Four Approaches. *AMS Rev.* **2020**, *10*, 18–26. [CrossRef]
19. Kitchin, R. The Real-Time City? Big Data and Smart Urbanism. *GeoJournal* **2014**, *79*, 1–14. [CrossRef]
20. Norman, B. Are Autonomous Cities Our Urban Future? *Nat. Commun.* **2018**, *9*, 2111. [CrossRef]
21. Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Akour, I.A. IoT for Smart Cities: Machine Learning Approaches in Smart Healthcare—A Review. *Future Internet* **2021**, *13*, 218. [CrossRef]
22. Yigitcanlar, T.; Cugurullo, F. The Sustainability of Artificial Intelligence: An Urbanistic Viewpoint from the Lens of Smart and Sustainable Cities. *Sustainability* **2020**, *12*, 8548. [CrossRef]
23. Sapkota, R.; Roumeliotis, K.I.; Karkee, M. AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenge. *arXiv* **2025**, arXiv:2505.10468.
24. Bengio, Y.; Cohen, M.; Fornasiere, D.; Ghosn, J.; Greiner, P.; MacDermott, M.; Mindermann, S.; Oberman, A.; Richardson, J.; Richardson, O.; et al. Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path? *arXiv* **2025**, arXiv:2502.15657.
25. Floridi, L.; Sanders, J.W. On the Morality of Artificial Agents. *Minds Mach.* **2004**, *14*, 349–379. [CrossRef]
26. Coeckelbergh, M. *AI Ethics*; MIT Press: Cambridge, MA, USA, 2020.

27. Ghose, S. The Next "Next Big Thing": Agentic AI's Opportunities and Risks. UC Berkeley Sutardja Center. 2024. Available online: https://scet.berkeley.edu/the-next-next-big-thing-agentic-ais-opportunities-and-risks/ (accessed on 1 May 2025).
28. AI Agent Store. Infrastructure & City Planning Weekly AI News. AI Agent Store. 2025. Available online: https://aiagentstore.ai/ai-agent-news/topic/infrastructure-city/detailed (accessed on 1 May 2025).
29. Townsend, A.M. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*; W. W. Norton & Company: New York, NY, USA, 2013.
30. Nam, T.; Pardo, T.A. Conceptualizing Smart City with Dimensions of Technology, People, and Institutions. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, MD, USA, 12–15 June 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 282–291.
31. Sadowski, J.; Bendor, R. Selling Smartness: Corporate Narratives and the Smart City as a Sociotechnical Imaginary. *Sci. Technol. Hum. Values* **2019**, *44*, 540–563. [CrossRef]
32. Wang, D.; Lu, C.T.; Fu, Y. Towards Automated Urban Planning: When Generative and ChatGPT-like AI Meets Urban Planning. *arXiv* **2023**, arXiv:2304.03892.
33. Parekh, D.; Poddar, N.; Rajpurkar, A.; Chahal, M.; Kumar, N.; Joshi, G.P.; Cho, W. A Review on Autonomous Vehicles: Progress, Methods and Challenges. *Electronics* **2022**, *11*, 2162. [CrossRef]
34. Zhang, J.; Hua, X.S.; Huang, J.; Shen, X.; Chen, J.; Zhou, Q.; Zhao, Y. City Brain: Practice of Large-Scale Artificial Intelligence in the Real World. *IET Smart Cities* **2019**, *1*, 28–37. [CrossRef]
35. Muraven, M. Goal Conflict in Designing an Autonomous Artificial System. *arXiv* **2017**, arXiv:1703.06354.
36. Qian, K.; Mao, L.; Liang, X.; Ding, Y.; Gao, J.; Wei, X.; Li, J. AI Agent as Urban Planner: Steering Stakeholder Dynamics in Urban Planning via Consensus-Based Multi-Agent Reinforcement Learning. *arXiv* **2023**, arXiv:2310.16772.
37. Greenfield, A. *Radical Technologies: The Design of Everyday Life*; Verso: London, UK, 2017.
38. Luque-Ayala, A.; Marvin, S. Developing a Critical Understanding of Smart Urbanism? *Urban Stud.* **2015**, *52*, 2105–2116. [CrossRef]
39. Gramoli, L.; Cumin, J.; Lacoche, J.; Foulonneau, A.; Arnaldi, B.; Gouranton, V. Generating and Evaluating Data of Daily Activities with an Autonomous Agent in a Virtual Smart Home. *ACM Trans. Multimed. Comput. Commun. Appl.* **2025**, *21*, 1–25. [CrossRef]
40. Symeonidis, A.; Mitkas, P. A Methodology for Predicting Agent Behavior by the Use of Data Mining Techniques. In *International Workshop on Autonomous Intelligent Systems: Agents and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 161–174.
41. Weber, C.; Elshaw, M.; Mayer, N.M. (Eds.) *Reinforcement Learning*; BoD–Books on Demand: Norderstedt, Germany, 2008.
42. Giampapa, J.A.; Sycara-Cyranski, K. *Team-Oriented Agent Coordination in the RETSINA Multi-Agent System*; The Robotics Institute, Carnegie Mellon University: Pittsburgh, PA, USA, 2002.
43. Abu Bakar, N.; Selamat, A.; Krejcar, O. Improving Agent Quality in Dynamic Smart Cities by Implementing an Agent Quality Management Framework. *Appl. Sci.* **2019**, *9*, 5111. [CrossRef]
44. Cui, H.; Du, Y.; Yang, Q.; Shao, Y.; Liew, S.C. LLMind: Orchestrating AI and IoT with LLM for Complex Task Execution. *IEEE Commun. Mag.* **2024**, *63*, 214–220. [CrossRef]
45. Poudel, B.; Neloy, M.S.; Chirawu, A.M. Combat Efficiency and Effectiveness of AI-Driven Multisensory Search-and-Destroy Agent. *Acad. Res.* **2021**, *4*, 42–54.
46. Lee, D.; Kim, G.; Kim, H.K. User Experience for Artificial Intelligence Assistant: Focusing on Negative Side Effects. In *International Conference on Human-Computer Interaction*; Springer Nature: Cham, Switzerland, 2023; pp. 88–97.
47. Helbing, D. FuturICT—New Science and Technology to Manage Our Complex, Strongly Connected World. *arXiv* **2011**, arXiv:1108.6131.
48. Castelvecchi, D. Can We Open the Black Box of AI? *Nature* **2016**, *538*, 20. [CrossRef]
49. Crawford, K. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*; Yale University Press: New Haven, CT, USA, 2021.
50. Nemitz, P. Constitutional Democracy and Technology in the Age of Artificial Intelligence. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 20180089. [CrossRef]
51. Selbst, A.D.; Boyd, D.; Friedler, S.A.; Venkatasubramanian, S.; Vertesi, J. Fairness and Abstraction in Sociotechnical Systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; pp. 59–68.
52. Goldsmith, S.; Yang, J. AI and the Transformation of Accountability and Discretion in Urban Governance. *arXiv* **2025**, arXiv:2502.13101.
53. European Union. General Data Protection Regulation (GDPR), Regulation (EU) 2016/679. *Off. J. Eur. Union* **2016**. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng (accessed on 1 May 2025).
54. Automated Decision Systems Task Force (ADSTF). *Report of the ADS Task Force*; NYC Mayor's Office of Operations: New York, NY, USA, 2019. Available online: https://www.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf (accessed on 1 May 2025).

55. Floridi, L.; Cowls, J. A Unified Framework of Five Principles for AI in Society. In *Machine Learning and the City: Applications in Architecture and Urban Design*; Wiley: Hoboken, NJ, USA, 2022; pp. 535–545. Available online: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119815075.ch45 (accessed on 1 May 2025).
56. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*; PublicAffairs: New York, NY, USA, 2019.
57. Green, B. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*; MIT Press: Cambridge, MA, USA, 2019.
58. Bryson, J.J. Robots Should Be Slaves. In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*; Wilks, Y., Ed.; John Benjamins: Amsterdam, The Netherlands, 2010; pp. 63–74.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.