automation

MDPI

Perspective

From Optimisation to Closed-Loop Urban Automation: A Conceptual Framework for Spatial Intelligence and Physical AI in AI–IoT-Enabled Smart Cities

Alok Tiwari * and Yasser Qaffas

Department of Urban and Regional Planning, Faculty of Architecture and Planning, King Abdulaziz University, Jeddah 80210, Saudi Arabia; yqaffas@kau.edu.sa

* Correspondence: atwari@kau.edu.sa

Abstract

Artificial intelligence (AI) and the Internet of Things (IoT) are converging to create densely sensed, connected, and increasingly automated urban environments. However, research on AI–IoT integration remains dominated by optimisation-centric perspectives that underspecify spatial reasoning, physical actuation, and institutional accountability. This perspective addresses that gap by developing closed-loop urban automation as an analytical lens for understanding how AI–IoT systems move from urban sensing to consequential intervention. Drawing on a narrative, theory-driven synthesis of literature on Urban AI, IoT, Physical AI, spatial intelligence, digital twins, robotics, automation, and governance, the paper differentiates the framework from AIoT, cyber-physical systems, embodied AI, and optimisation-centric smart-city models. It identifies the distinctive conditions of urban automation, proposes six examinable propositions, and outlines implications for planning support, intelligent control, human oversight, and democratic legitimacy.

Keywords: smart cities; artificial intelligence; Internet of Things; closed-loop urban automation; IoT sensor networks; physical AI; spatial intelligence; sensor–actuator systems; urban digital twins; AI governance

1. Introduction

The integration of artificial intelligence (AI) with the Internet of Things (IoT) is reconfiguring cities as connected and increasingly automated environments. Large-scale sensor networks, pervasive connectivity, and edge-to-cloud analytics generate continuous urban data streams that inform decisions across transportation, energy, environmental monitoring, public safety, and infrastructure management [1,2]. These capabilities enable cities to function as increasingly physically intelligent environments [3,4]. Yet, the conceptual basis of the transition from analysis to intervention remains under-specified [5,6]. As AI moves into streets, buildings, infrastructures, and control systems [7,8], the central issue is no longer only how accurately models predict or optimise, but how computational systems act under spatial, physical, and institutional constraints [4,5]. This paper argues that Physical AI requires an expanded conception of intelligence grounded in real-world interaction, spatial reasoning, and accountable automation [8].

The stakes are practical as well as conceptual. Municipal governments are already procuring systems that adjust traffic signals, issue emergency warnings, reroute mobility flows, modulate energy demand, and regulate access to public space with limited human

Check for updates

Academic Editor: Isam Shahrour

Received: 3 July 2026
Revised: 30 July 2026
Accepted: 1 August 2026
Published: 3 August 2026

Copyright: © 2026 by the authors. 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.

intervention. Failures in autonomous mobility, including the 2023 suspension of Cruise's driverless operations in San Francisco, show that the consequences extend beyond model accuracy to public safety, liability, regulatory trust, and infrastructure investment. Planners, city officials, infrastructure managers, and regulators, therefore, need a clear basis for specifying, supervising, auditing, and contesting automated interventions.

Across adaptive traffic control, autonomous mobility, robotics, emergency response, and responsive infrastructure, AI is acquiring direct spatial, physical, and ethical consequences. Once intelligence is enacted in urban space, it becomes accountable to latency, safety, physical constraints, social relations, and potentially irreversible effects of action [9]. The analytical question, therefore, shifts from how AI optimises the city to how it interprets urban conditions, intervenes safely, and remains institutionally governable in dynamic and contested settings.

In this paper, closed-loop urban automation refers to a spatially distributed AI–IoT arrangement in which sensed urban conditions inform decisions or control signals, interventions alter physical or institutional conditions, and the resulting effects return as feedback. Spatial intelligence denotes the capacity to interpret relations across locations and scales, anticipate how effects may propagate, and constrain action according to context. These concepts shift attention from a data-processing pipeline to an accountable architecture of urban intervention.

Existing research provides only part of this account. Control-theoretic and cyber-physical-systems scholarship offers mature concepts of feedback, stability, and hierarchical control, but commonly abstracts urban spatial structure and institutional mediation. Smart-city and urban-informatics research represents space and data in greater detail, yet generally treats AI outputs as predictions or recommendations rather than consequential interventions. The missing bridge is an analytical framework in which non-stationary, multi-agent disturbance, spatially distributed actuation, and institutionally mediated feedback are constitutive rather than peripheral. The framework developed here also engages the longer lineage of cybernetic urbanism and identifies three features of the present configuration: real-time sensor–actuator coupling at city scale, simulation as a supervisory layer, and institutional mediation within the control architecture.

Accordingly, the paper pursues three objectives. First, it clarifies the ontological status of Physical AI, spatial intelligence, and closed-loop urban automation. Second, it differentiates the proposed lens from AIoT, cyber-physical systems, urban digital twins, embodied AI, and optimisation-centric smart-city models. Third, it translates the conceptual synthesis into propositions that can be applied, contested, and empirically tested.

Methodologically, the paper employs a theory-driven conceptual synthesis [10], supported by structured and purposive engagement with literature on Urban AI, IoT-enabled urban systems, Physical AI, spatial intelligence, digital twins, robotics, intelligent control, automation, and AI governance. The next section details the literature-selection strategy and the conceptual procedure through which the framework and its propositions are developed.

2. Conceptual Approach and Literature Selection

This perspective uses a theory-driven conceptual synthesis [10,11] to integrate work across disciplines, clarify the core concepts, trace the development of Urban AI and AI–IoT integration, and construct the proposed framework. The review is narrative and purposive rather than systematic; its aim is theory building through conceptual integration [10], not exhaustive coverage, meta-analysis, or corpus-based review. Formal protocols such as PRISMA are, therefore, not applied, although the selection logic is reported below.

The research design proceeds in five connected steps: (i) identification of the theoretical problem, namely the optimisation-centric under-specification of feedback, space, and actuation in Urban AI research; (ii) purposive engagement with literature across six term families; (iii) conceptual clarification of the framework's constructs and their ontological status; (iv) differentiation of the proposed lens from adjacent frameworks; and (v) derivation of the framework, its architecture, and its propositions. This sequence makes the progression from the research problem and literature synthesis to conceptual differentiation and framework development explicit.

2.1. Literature Selection

To situate the framework within the temporal development of the field, the literature search followed a staged selection strategy. Foundational smart-city and code/space scholarship provided the conceptual baseline [3,12,13]. Critical and political work on Urban AI and platform urbanism was then incorporated to frame governance transformations, shifts in urban knowledge production, and the institutional implications of automated urban decision-making [4,6,14–16]. Recent literature from 2023 to 2026 was selected to capture technical advances in geospatial AI, urban informatics, foundation models, embodied and Physical AI, IoT-enabled urban sensing, simulation-enabled systems, edge intelligence, sensor–actuator systems, and intelligent control [1,2,17–20].

Database searches were conducted in Web of Science, Scopus, IEEE Xplore, ACM Digital Library, and ScienceDirect for the period January 2016 to January 2026. The searches used Boolean combinations of six term families: urban AI and smart cities; AI and IoT integration, edge intelligence, and sensor networks; automation, intelligent control, cyber-physical systems, and closed-loop systems; spatial intelligence and spatiotemporal modelling; embodied or Physical AI and sensor–actuator systems; and governance, planning support, and algorithmic accountability. Their purpose was to map the discourse and identify conceptually relevant sources rather than to construct a systematic-review corpus.

Sources were included when they (a) addressed the coupling of sensing, inference, control, or actuation in urban settings; (b) offered theoretical accounts of embodiment, spatial reasoning, automation, or governance relevant to cities; or (c) documented deployed AI–IoT systems with observable feedback structures. Sources were excluded when they addressed purely algorithmic performance without urban or physical instantiation, or when they duplicated conceptual ground already covered by stronger or more recent sources. No formal quality-scoring protocol was applied, consistent with the narrative design.

Because discourses on Physical AI, urban foundation models, and closed-loop automation are advancing partly through industry research laboratories and standard-setting platforms, the perspective also incorporated a bounded set of vendor and institutional publications as contextual and illustrative material. These included IBM explainers, AWS and NVIDIA technical blogs, and UN-Habitat toolkits. They were used only to illustrate how these technologies are framed and deployed in practice, not to substantiate conceptual claims of the paper, which rest primarily on the peer-reviewed literature.

2.2. The Status of the Core Concepts

Because the framework mobilises several concepts whose theoretical status could otherwise remain ambiguous, their ontological status is fixed at the outset. Physical AI is treated as a class of system architecture; spatial intelligence as a cognitive capacity attributable to artificial systems; and closed-loop urban automation as the analytical lens that couples the two with governance. These are not competing theories, but conceptually distinct elements performing different functions within the same framework. Table 1 summarises these distinctions and separates each concept from its nearest neighbours.

Table 1. Ontological status of the framework's core concepts and their distinction from adjacent concepts.

Concept	Ontological Status	Role in the Framework	Distinguished from
Physical AI	A class of system architecture: sensor–actuator-coupled systems operating in real time under physical constraint [21,22]	The operative substrate through which the loop perceives and acts in physical urban space	Embodied AI (a broader research programme including cognitive claims about embodiment [23,24]); robotics (one platform class among several through which Physical AI is realised)
Spatial intelligence	A cognitive capacity; a theoretical construct attributable to artificial systems [25–27]	The reasoning substrate that binds sensing, inference, control, and actuation into spatially coherent, multi-scalar action	GeoAI (a toolset of geospatial methods [28]); spatiotemporal modelling (a family of techniques that spatial intelligence draws upon but exceeds, since it includes anticipation of consequences and constraint-aware action)
Closed-loop urban automation	An analytical lens: an integrative framework, not a technology	The frame that couples Physical AI and spatial intelligence with institutional governance into a single, accountable control architecture	Cyber-physical systems (which abstract institutional feedback); smart-city optimisation (which leaves the loop open at actuation); AIoT (which specifies connectivity, not accountable closure)

2.3. What the Proposed Lens Adds to Existing Frameworks

A reasonable objection is that AI–IoT convergence is already theorised as AIoT, urban cyber-physical systems, digital twins, embodied AI, or optimisation-centric smart-city architectures, and that the proposed lens risks relabelling existing ideas. Table 2 addresses this objection by comparing the frameworks along five dimensions: the treatment of urban space; where, and whether, the loop closes; the treatment of non-stationarity and multi-agent disturbance; the role of institutional feedback; and the limitation addressed by the proposed lens.

Table 2. Differentiation of the proposed lens from adjacent frameworks.

Framework	Treatment of Urban Space	Where the Loop Closes	Non-Stationarity/Multi-Agent Disturbance	Institutional Feedback	Limitation Addressed by the Proposed Lens
AIoT	Deployment context for networked devices	Typically open: analytics inform human operators	Treated as data drift	Largely absent	Adds spatial reasoning, actuation, and institutional feedback as first-order elements
Cyber-physical systems	Abstracted into plant dynamics	Closes technically at the actuator	Assumed away or bound by design	Modelled, if at all, as delay	Restores spatial structure and institutional mediation to the loop
Urban digital twins	Richly represented, but virtually	Usually does not close; outputs are representations	Simulated rather than confronted	Enters informally through planning practice	Positions twins as the supervisory layer of a larger loop rather than an end point

Table 2. *Cont.*

Framework	Treatment of Urban Space	Where the Loop Closes	Non-Stationarity/Multi-Agent Disturbance	Institutional Feedback	Limitation Addressed by the Proposed Lens
Embodied AI/urban robotics	Navigable local environment	Closes at the individual agent	Handled locally, agent by agent	Peripheral	Scales from agent to city; adds multi-scalar propagation and governance
Optimisation-centric smart city	A dataset to be optimised	Open: predictions, dashboards, recommendations	A retraining problem	External regulation after the fact	Reframes prediction as one stage in an accountable control loop
Proposed: spatially intelligent closed-loop urban automation	An active medium of sensing, computation, control, and intervention	Closes across sensing–inference–control–actuation–feedback, including the institutional path	Constitutive assumptions: non-stationary plant, multi-agent disturbances	Constitutive: set-points and oversight sit inside the loop	—

The comparison yields two principal observations. First, no existing framework simultaneously treats urban space as an active medium, closes the loop through actuation, and incorporates institutional feedback as a constitutive element. Second, the proposed lens does not replace adjacent frameworks but repositions them within a broader architecture; digital twins become the supervisory layer of a larger loop, embodied AI supplies its agent-level substrate, and CPS contributes its control vocabulary. Later sections extend this vocabulary by identifying four ways in which urban automation departs from classical control assumptions.

Apart from above, Prompts used for creating Figures 1–3 is available as Supplementary Material File S1.

3. From Digital AI to Physical AI

For much of its recent development, artificial intelligence has operated on representations of the physical world rather than direct interaction with it [23,24], even when its applications were intended to shape physical environments. Deep learning, large language models, and predictive analytics have been optimised for text, images, networks, and datasets, producing highly capable but largely disembodied systems. In this tradition, intelligence is defined mainly as inference from data, with limited attention to action under physical constraints [9].

This orientation has privileged symbolic manipulation and statistical pattern recognition, allowing AI to operate at scale while remaining partly removed from the frictions and irreversibilities that define physical reality. Although this conception may be sufficient for closed computational domains, it becomes inadequate when AI systems intervene directly within urban infrastructures, mobility systems, public spaces, and automated control environments. Growing scholarly attention to embodied interaction, real-time sensing, control, and actuation, therefore, indicates a shift from optimisation-dominant understandings of AI towards more physically grounded conceptions of intelligence.

A concrete urban contrast makes the distinction clear. A conventional traffic model analyses historical detector data and outputs predicted congestion or recommended signal timings; any street-level consequence depends on separate human implementation. An adaptive signal-control system coupled to roadside sensors and connected intersections,

by contrast, perceives current conditions and directly changes signal operation. Its errors are enacted physically rather than merely reported. The distinction, therefore, concerns whether model output becomes action and the safety and accountability obligations that follow. Similar contrasts apply to flood management (a forecast report versus automatically triggered alerts or barriers) and energy management (a demand forecast versus automated demand-response control).

As defined in Table 1, Physical AI comprises sensor–actuator-coupled systems operating in real time under physical constraint [21,22]. In cities, networked devices provide perception, edge and cloud systems support inference and control, and connected vehicles, robots, signals, and responsive infrastructures provide channels for action. Physical coupling makes uncertainty, safety, and institutional responsibility part of system performance rather than external considerations.

Advances in differentiable physics engines and vision–language–action models may allow machines to generalise across environments rather than depend on narrowly specified programmes. Their urban significance lies in the need to understand spatial relations and anticipate consequences before acting in open-ended, dynamic settings. Spatial reasoning, therefore, is an operational requirement of Physical AI, not simply an enhancement to representation.

Persistent challenges, including data scarcity, safety-critical constraints, brittleness in real-world deployment, and the gap to human-level dexterity, continue to restrict the scalability and generalisation of embodied systems. These challenges have accelerated interest in high-fidelity synthetic data, simulation environments, and large physics-aware foundation models as mechanisms for expanding learning capacity under physical constraints [29]. Their growing use reflects a central challenge for Physical AI: intelligence developed through abstract computation must be adapted to environments characterised by uncertainty, friction, feedback, and irreversible consequences.

Urban AI is an operative component of systems that intervene in and reshape urban conditions. Because these interventions occur in socially governed environments, technical capability must be matched by accountability for spatial consequences.

Figure 1 presents an analytical typology rather than a fixed historical or technological progression. It distinguishes four configurations by sensor–actuator embodiment, the scope of spatial reasoning, the role of simulation, and the depth of institutional mediation. Existing deployments may combine characteristics of several configurations. The typology synthesises literature on embodied systems [21–24]; multi-scalar spatial reasoning [25–27,30]; and the governance of automated urban decision-making [4,15,16].

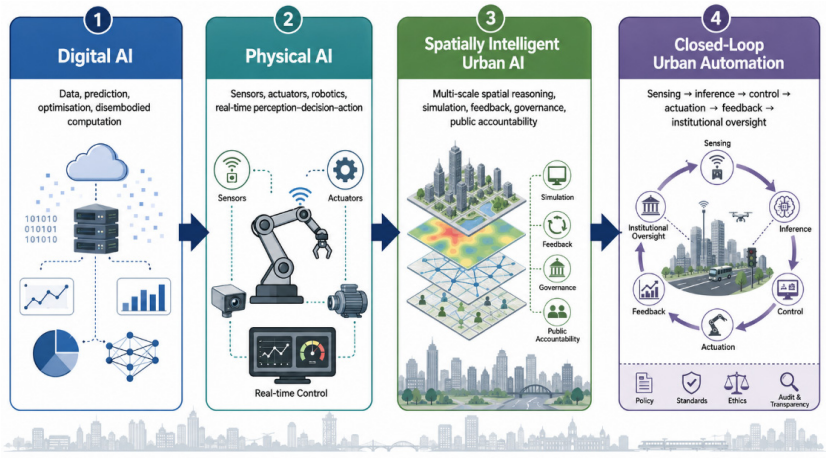


Figure 1. From Digital AI to Spatially Intelligent Closed-Loop Urban Automation (created using ChatGPT 5.5).

4. Spatial Intelligence as a Core Capability

Spatial intelligence gives operational substance to the framework enabling artificial systems to relate locations, scales, forms, flows, and likely consequences [5,25–27]. Its relevance is clearest when an intervention at one point alters conditions elsewhere.

Its multi-scalar character is important in cities, where a local intervention can propagate through transport networks, infrastructure systems, governance arrangements, and socio-spatial relations [22,27]. A signal plan that relieves one arterial may displace congestion onto residential streets; a district-scale flood may miss street-level drainage failure; and an energy demand-response event may interact with heat-wave vulnerability in ways that remain invisible at feeder level. Infrastructure-monitoring similarly requires building-scale observations to be reconciled with network-scale logistics and safety.

A second requirement is to connect relatively static forms, such as buildings and infrastructure, with dynamic flows of people, vehicles, energy, water, and information [30]. This relational view allows an automated system to assess not only what is present at a location, but how an intervention may alter movement, access, exposure, or network performance elsewhere. Without such spatial capacities, AI remains vulnerable to flattened abstractions and grid-based simplifications that obscure the volumetric, relational, and dynamic complexity of urban environments [22,31,32].

Cities function as non-stationary and continually evolving systems, reshaped by development pressures, climate change, and shifting social practices [33,34]. They are also multi-agent environments formed through interactions among humans, vehicles, infrastructures, automated systems, and institutions. Models of these environments, therefore, must account for distributed agency, feedback effects, and spatial interdependence [2,27]. Urban spaces are politically mediated by rules, rights, and ethical constraints [35] and materially bounded by finite land, energy, and resource capacities [36–38].

These characteristics make local optimisation an unreliable proxy for system-wide performance. An intervention can be technically efficient at its point of application yet socially misaligned once its effect encounter inequality, infrastructural interdependence, regulatory complexity, and competing claims over public space [2,9].

Generative world models add an anticipatory dimension by simulating how environments may evolve while preserving geometric and physical consistency. By combining images, text, sensor data, and spatial representations, they create testing grounds in which potential actions can be assessed before implementation [9]. Their value to urban automation lies in exposing causal dependencies, cascading effects, and safety risks under changing conditions, rather than merely producing more detailed representations.

5. AI–IoT Integration as Closed-Loop Urban Automation

Early formulations of Urban AI were strongly oriented towards optimisation. Research agendas and applied systems focused on traffic-flow prediction, energy-demand forecasting, land-use classification, and crime-hotspot mapping, delivering tangible gains in efficiency and managerial control. However, these approaches often rested on a narrow epistemological foundation in which the city was treated primarily as a dataset to be optimised. Urban complexity was reduced to tractable variables, with limited attention to how social practices, spatial relations, and physical constraints unfold across time and scale. Consequently, calculability was often privileged over contextual understanding.

Optimisation also operates as a value-laden ordering logic that can privilege efficiency while marginalising power, contestation, and lived spatial experience [9]. The convergence of Physical AI and spatial intelligence moves Urban AI towards what recent scholarship describes as AI urbanism, in which analysis is more directly coupled to intervention.

Optimisation remains important, but it is increasingly conditioned by richer spatial models, simulation, contextual reasoning, and public oversight.

AI–IoT integration supplies the operational bridge for this transition by connecting distributed urban sensing to computational decision-making and channels of intervention [1,22]. The analytical significance of that connection lies in the distance between inference and physical consequence [9].

Within this recursive architecture, simulation and generative world models provide an anticipatory and supervisory layer. They allow interventions to be tested, cascading effects to be examined, and control strategies to be revised before implementation in physical space [19,22,39]. This extends reactive optimisation towards scenario-based and consequence-aware decision-making.

Spatially grounded AI systems can reveal nonlinear interactions among transportation networks, land-use patterns, environmental conditions, and public space that have long eluded conventional urban models grounded in linear assumptions. Recent studies using interpretable machine learning show, for example, that built-environment features and urban vitality interact in highly nonlinear ways that traditional models fail to capture [40]. Explainable AI approaches also uncover complex and multi-scalar relationships between urban morphology and thermal dynamics [41]. These applications demonstrate the analytical value of reasoning across spatial scales and temporal horizons, particularly where urban dynamics depart from static or equilibrium-based assumptions.

Advances in geospatial AI further expand the range and resolution of urban analysis, from satellite-based site suitability and ecological assessments to fine-grained modelling of urban heat islands and pollution dynamics [18,30,42]. Within the proposed framework, however, predictive capacity becomes consequential only when its relationship to intervention, feedback, and governance is made explicit.

The AI–IoT Automation Stack: Sensing, Edge Intelligence, Control, Actuation, and Feedback

At the sensing layer, networks of fixed and mobile IoT devices, roadside units, environmental and air-quality sensors, CCTV, LiDAR-equipped vehicles, and crowdsourced 360-degree capture continuously instrument the urban fabric, narrowing the gap between the city and its computational representation [1,43].

Figure 2 translates the conceptual lens into control-theoretic terms, locating urban subsystems, sensing, hierarchical control, actuation, disturbances, policy set-points, and human oversight within one architecture.

Figure 2 can be read through six control elements. The plant comprises the urban subsystems being influenced, such as traffic, energy, drainage, or public space. Sensors measure changing conditions; the controller compares those measurements with policy, regulatory, or service set-points; and actuators apply the resulting control input through signals, alerts, routing systems, or adaptive infrastructure. Disturbances include weather, demand shocks, incidents, and human behaviour, while feedback records the effects of prior interventions.

As Figure 2 makes explicit, urban deployments depart from classical closed-loop assumptions in at least four respects: the plant is non-stationary, disturbances d arise from multiple interacting agents, actuation is spatially distributed across heterogeneous infrastructures, and part of the feedback path is institutionally mediated and, therefore, subject to delays and contestation. These four departures distinguish urban automation from the classical settings in which closed-loop control was developed, and they generate the propositions formalised later.

Latency-sensitive perception, filtering, inference, and local control increasingly occur at the edge or fog layer, close to the data source [22,31]. Cloud systems support higher-

order model training, large-scale simulation, digital-twin synchronisation, and system-level calibration. Actuation is distributed across vehicles, robots, adaptive signals, environmental controls, and responsive infrastructure. This allocation of functions reflects differing requirements for latency, computational capacity, safety, and jurisdiction.

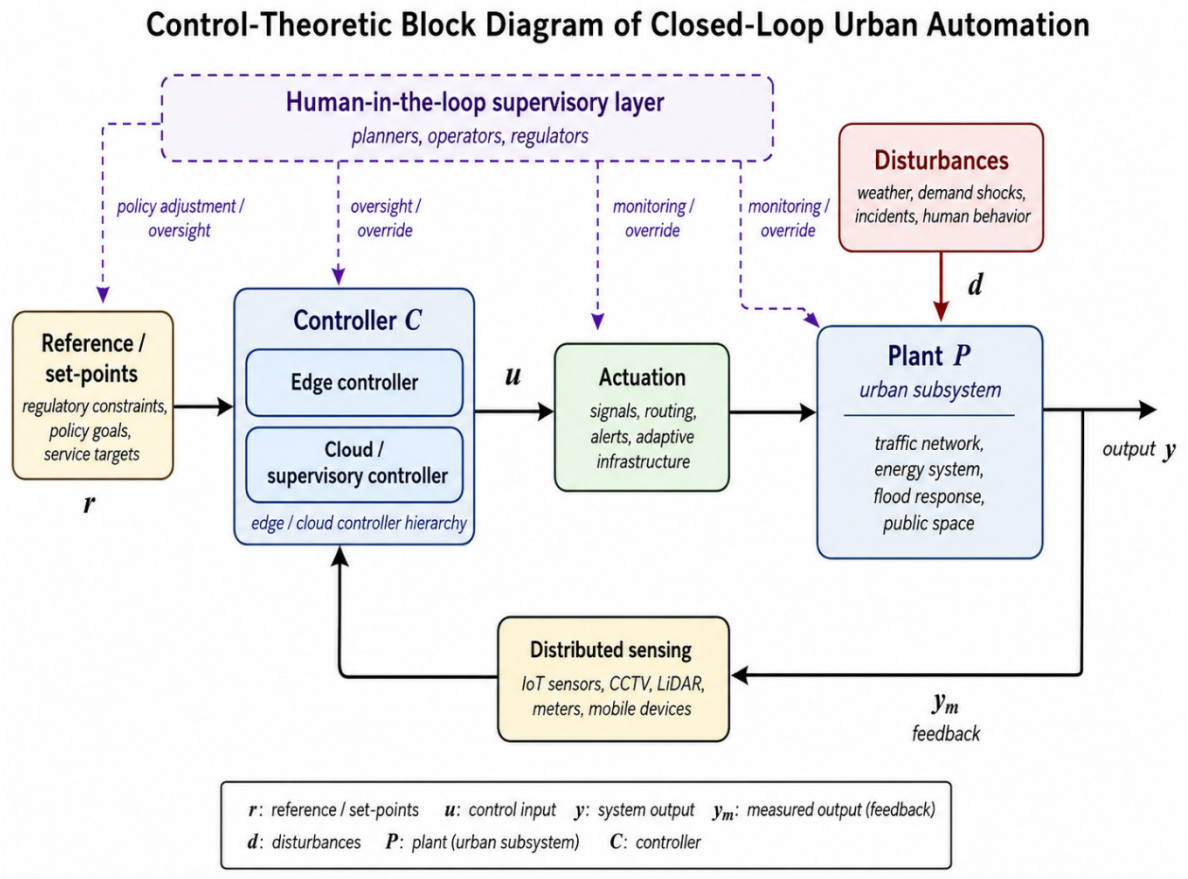


Figure 2. Closed-loop architecture of AI–IoT urban automation (Created using ChatGPT 5.5).

Two cross-cutting concerns follow directly from this architecture. Continuous, fine-grained urban data require privacy-preserving approaches, including federated learning, that can reduce the centralisation of raw data [44–46]. Dense interconnection also makes security and interoperability first-order design issues because vulnerabilities can cascade across coupled infrastructures [47]. The central architectural challenge, therefore, is coordination across technical layers without losing legal, ethical, or spatial constraints.

Table 3 compares two ideal-typical configurations of this architecture. Autonomous mobility combines direct actuation with immediate physical consequences and, therefore, requires real-time safety controls. Urban digital twins usually operate at the supervisory level; their errors emerge indirectly through planning, policy, and investment decisions and require institutional review, auditability, and transparent interpretation. The distinction concerns the pathway and timing through which model error becomes material.

Figure 3 positions selected Urban AI applications according to their degree of embodiment and reliance on anticipatory simulation. Autonomous urban mobility sits high on embodiment because it is directly sensor–actuator-coupled in shared public space, and increasingly high on anticipatory simulation because world models and scenario testing now mediate its control stack. Urban digital twins sit high on simulation but low-to-moderate on embodiment because their outputs are typically enacted through human planning decisions rather than direct actuation. Robotics and surveillance sit high on embodiment but lower on anticipatory simulation, operating largely through real-time monitoring and

local navigation. Climate adaptation and emergency response sit high on simulation and forecasting with limited direct actuation, since warnings are enacted institutionally by human agencies. The diagonal gradient represents a consequence–oversight relationship: as the material consequences of automated action increase, so does the need for stronger human oversight, auditability, liability arrangements, and public accountability.

Table 3. Ideal-Typical Configurations of Spatially Intelligent Closed-Loop Urban Automation.

Dimension	Autonomous Urban Mobility (Embodied Physical AI)	Urban Digital Twins (Simulation-Enabled Spatial Intelligence)
Primary Mode of Operation	Real-time sensing–decision–control–actuation loop	Simulation-based modelling, scenario testing, and supervisory decision support
Embodiment Level	High: directly coupled to sensor–actuator systems operating in physical space	Low to Moderate: primarily computational; may inform actuators indirectly
Automation Logic	Direct closed-loop automation in dynamic public environments	Indirect or supervisory automation through modelling, prediction, and decision support
Spatial Reasoning Scope	Lane-level to network-scale; dynamic multi-agent navigation	Multi-scalar system modelling across infrastructure, flows, environment, and risk
Temporal Orientation	Immediate, real-time adaptation	Anticipatory, counterfactual, and scenario-based reasoning
Material Consequence of Error	Direct and immediate, including safety, collision risk, liability, and public trust	Indirect but consequential, mediated through planning, policy, infrastructure, and investment decisions
Governance Entanglement	Strongly implicated in system architecture, including liability, regulatory compliance, infrastructure redesign, and operational safety	Institutional and policy-facing, including planning support, decision evaluation, transparency, and public accountability
Relationship to Optimisation	Optimisation remains present but is embedded within real-time spatial decision-making, safety constraints, control logic, and embodied action	Optimisation is mediated through simulation, counterfactual comparison, scenario-based spatial reasoning, and policy evaluation
Feedback Role	Immediate feedback from physical movement, sensor readings, infrastructure response, and human behaviour	Delayed or model-mediated feedback through updated data, monitoring, validation, and planning revision
Human Oversight	Requires human-in-the-loop or human-on-the-loop control, safety assurance, and regulatory intervention mechanisms	Requires institutional review, participatory interpretation, auditability, and transparent use in decision-making
Normative Exposure	High, operates in contested public space with socio-spatial negotiation	Moderate to High: shapes long-term policy and infrastructural decisions through indirect but consequential planning pathways

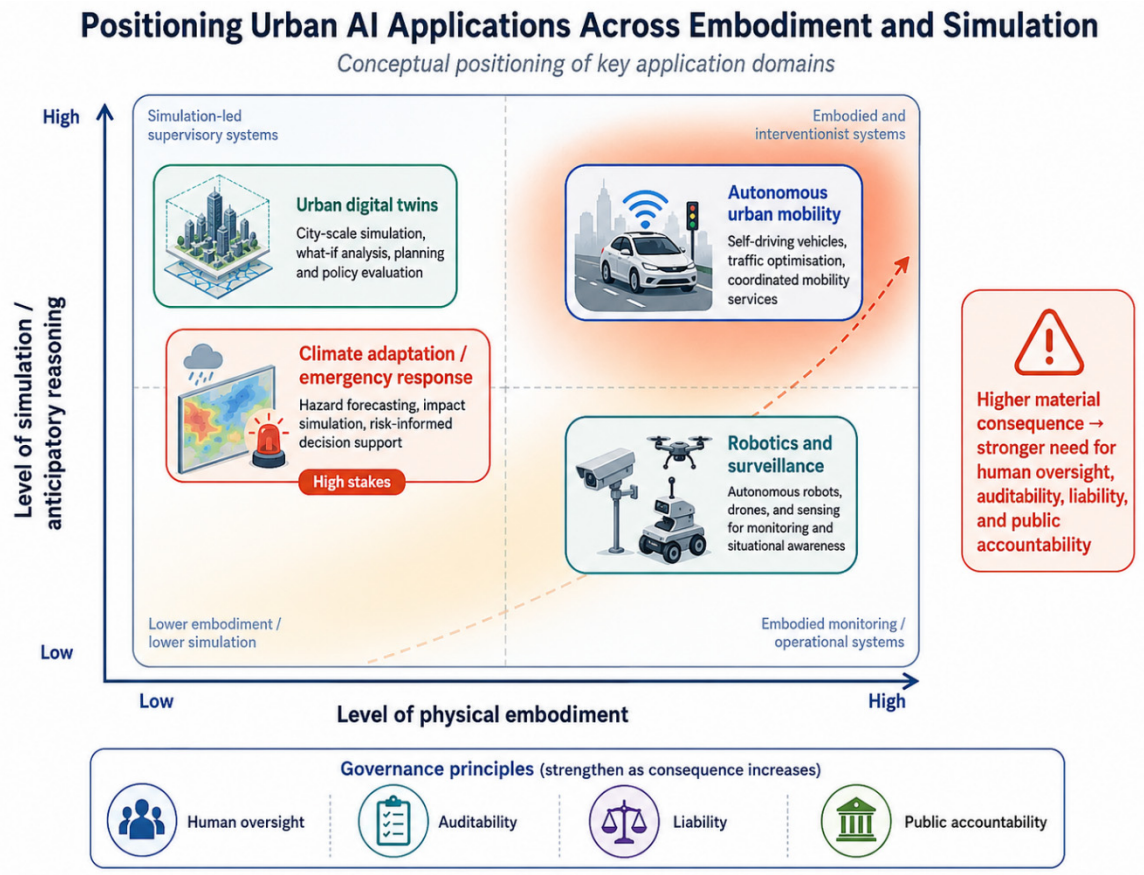


Figure 3. Positioning Urban AI Applications Across Embodiment and Simulation (created using ChatGPT 5.5).

6. Emerging Applications and Case Examples

The four application domains below, mobility, digital twins, robotics and surveillance, and climate adaptation, illustrate configurations that vary in embodiment, temporality, material consequence, and institutional mediation. Rather than repeating the framework in each case, the analysis applies the elements established in Figure 2 and identifies the distinctive governance implication of each configuration.

6.1. Autonomous Urban Mobility as Embodied Closed-Loop Automation

Autonomous urban mobility is a high-embodiment configuration of Urban AI [48,49]. Vehicle-borne and roadside sensing feed onboard inference and control, while connected actuation changes movement in real-time. Multimodal perception, spatiotemporal modelling, and predictive world models are used to interpret lane geometry, pedestrian movement, infrastructure, weather, and regulatory boundaries in dynamic, multi-agent environments.

The decisive challenge is not simply geometric accuracy but norm-aware navigation in shared public space. Negotiated right-of-way, informal pedestrian behaviour, and culturally specific conventions rarely appear fully in formal datasets, yet they structure everyday mobility. Failures, therefore, may arise from mismatches between algorithmic assumptions and lived practice as much as from computational deficiencies.

Waymo's expansion of fully driverless ride-hailing across Phoenix, San Francisco, and Los Angeles illustrates the scaling of urban autonomy. The 2023 suspension of Cruise operations in San Francisco, following incidents involving pedestrians and emergency responders, shows how socio-spatial misalignment, public-safety concerns, and regulatory trust can trigger regulatory intervention and the withdrawal of operating licences [48,50,51].

Governance is consequently part of autonomous-mobility design rather than a post-deployment addition. Liability allocation, fail-safe operation, infrastructure adaptation, transparency, human override, and distributive equity must be considered in advance. The Cruise case shows that direct physical consequences can trigger institutional feedback through licensing decisions and regulatory suspension, supporting the proposition that legitimacy increasingly depends on contestability, oversight, and liability as the consequences of actuation rise.

6.2. Digital Twins as Simulation-Enabled Supervisory Automation

Digital twins exemplify AI–IoT integration at the modelling and supervisory layer by integrating real-time IoT sensor streams with spatial simulations of transportation networks, buildings, infrastructure, and environmental systems. Their role is not usually to actuate directly in physical space, but to support anticipatory modelling, scenario testing, and decision evaluation before intervention.

In Barcelona, large-scale traffic digital twins have been used to simulate congestion redistribution and air-quality scenarios under empirical validation [52]. Comparable AI-driven spatial analytics platforms in Singapore illustrate the maturation of simulation-based planning support [53]. The Virtual Singapore programme extends this trajectory into a city-scale, three-dimensional digital twin that integrates data on building geometry, mobility, and environmental conditions to support planning, accessibility analysis, and scenario testing [7]. European municipal digital twins similarly show how anticipatory simulation is being incorporated into urban planning practice. Helsinki's city-scale digital twin and Energy and Climate Atlas support analyses of solar-energy potential and flood assessment, as well as noise calculations, while the Herrenberg digital twin integrates a 3D urban model with mobility, wind-flow, pollution, volunteered geographic information, and participatory data to support scenario exploration and collaborative planning before physical intervention [54–57].

The boundary between decision support and automation depends on how model outputs are translated into action. A digital twin remains a decision-support instrument when its outputs are implemented through discrete human decisions. It becomes part of an automation system when its outputs are coupled directly, or through pre-authorised standing rules, to actuation. Examples include simulation-derived signal plans automatically transmitted to intersections or model-triggered flood alerts that activate barriers without case-by-case authorisation. Most municipal twins, including those discussed above, remain on the decision-support side of this boundary, which is why Table 3 characterises them as supervisory rather than directly automated.

Digital twins, therefore, occupy a complementary role: they provide anticipatory supervision for interventions enacted elsewhere and support repeated recalibration under non-stationary urban conditions.

6.3. Robotics and Surveillance as Mobile Urban Automation

Recent deployments of humanoid service robots and mobile surveillance systems in Chinese cities illustrate how Physical AI is entering everyday urban environments. In these settings, continuous monitoring and governance turn public space into an actively sensed and regulated technological environment [58]. During the COVID-19 pandemic, Singapore's deployment of the quadruped robot Spot to monitor safe distancing in Bishan–Ang Mo Kio Park offered a widely documented case of machines exercising spatial authority in shared public space [59]. Autonomous security and patrol robots in North American commercial, institutional, campus, airport, and transit settings have likewise extended mobile robotic surveillance into everyday environments [60–63].

Delivery robots, inspection drones, autonomous patrol systems, and AI-enabled monitoring infrastructures operate in shared environments where spatial authority is negotiated among humans, machines, and institutions. Ambiguity, informality, and competing territorial claims make public space fundamentally different from a controlled industrial setting. These systems participate in urban regulation through their mobility, signalling, monitoring, and effects on behaviour.

Initiatives such as AI-enabled CCTV and traffic systems in Hyderabad, developed within an ambition to create a Physical Intelligence City, illustrate how sensor-driven infrastructures are embedded within specific governance imaginaries [64]. In this case, spatial intelligence is technical and also aligned with a broader vision of urban control, efficiency, and real-time oversight. Crowdsourced urban mapping initiatives using 360-degree vision systems similarly recast cities as continuously updated and participatory datasets [43]. These platforms blur the boundaries between sensing and representation, while linking both to governance by transforming urban space into an ongoing stream of collective visual data.

Such systems can redistribute knowledge authority and spatial power by determining what is sensed, classified, and acted upon. Because mobile automation also influences behaviour in public space, proportionality, consent, oversight, and contestability must be treated as design requirements.

The case supports the proposition that where audit, override, and appeal mechanisms are not specified ex ante, the redistribution of spatial authority may advance more rapidly than the accountability arrangements intended to govern it, creating gaps that post-hoc regulation may not fully repair.

6.4. Climate Adaptation as High-Stakes Automated Decision Support

These dynamics intensify in emergency response and climate adaptation, where spatially intelligent AI systems are explored for flood management, wildfire mitigation, and evacuation planning [65]. Here, AI must integrate physical geography, infrastructure interdependencies, environmental change, and social vulnerability under conditions of acute uncertainty. Effective reasoning requires the synthesis of heterogeneous spatial knowledge and the anticipation of cascading effects across interconnected systems [28,66].

Flood Hub by Google couples AI-based flood forecasting with riverine flood maps, water trends, and global data sources to provide flood forecasts up to seven days in advance across river basins in more than 150 countries, translating spatial inference into early warnings for governments, aid organisations, and populations at risk [67,68]. AI-assisted wildfire detection networks in California, such as the camera-based ALERTCalifornia system developed with CAL FIRE and other partners, similarly demonstrate sensor-driven spatial reasoning under acute temporal pressure by using a statewide network of high-definition cameras and AI-supported anomaly detection to help identify emerging wildfire risks [69,70].

In these cases, automation is often advisory or semi-automated rather than fully autonomous. The system senses, predicts, classifies, and issues warnings, but human agencies remain responsible for interpretation, prioritisation, evacuation, and response. The technical loop, therefore, remains deliberately open at the actuation stage: sensing, inference, and warning generation may be automated, but evacuation orders, resource deployment, and fire suppression remain institutionally controlled.

The wider loop nevertheless closes through human agencies. Therefore, forecast latency, institutional response capacity, and uneven sensing density form part of the control path rather than external conditions. In Flood Hub, for example, the automated segment extends from hydrological sensing through forecast generation to warning dissemination,

while the institutional segment extends from warning receipt through prioritisation to evacuation. The effectiveness of the overall system is consequently limited by its slowest or least-resourced segment.

This configuration demonstrates that anticipatory modelling can help systems respond to highly non-stationary environmental conditions, but that institutionally mediated action remains integral to loop performance. Such deployments raise questions about the uneven distribution of benefit and risk, particularly where forecast coverage and sensing density reflect patterns of investment and infrastructural capacity.

Accordingly, climate applications should be evaluated not only by forecast accuracy but also by end-to-end response capacity, distributional coverage, and institutional accountability [71,72]. Prediction has public value only when warnings can be interpreted, contested where necessary, and translated into timely and equitable ac-tion.

7. Implications for Urban Planning as a Supervisory Layer

For planning, the framework shifts attention from producing static plans to supervising systems that adapt to real-time data and changing conditions. Planning enters the control architecture through the definition of objectives, constraints, acceptable operating ranges, and procedures for intervention or suspension. The role of planners consequently expands to specifying, supervising, and contesting automated urban operations.

One consequence is regulatory. Conventional zoning based on discrete land-use categories may be supplemented by continuous, data-rich spatial indicators that are algorithmically interpreted and periodically adjusted [73,74]. Regulatory authority is then distributed across data infrastructures, models, and embedded decision rules rather than residing solely in statutory codes. Planning standards, service objectives, and regulatory thresholds function as explicit constraints and set points. Without transparency, contestability, and democratic oversight, however, adaptive regulation can become technocratic control.

A second consequence is methodological: a shift from representation to simulation. Planners can use generative world models and digital twins to test how policies, infrastructural investments, climate shocks, or control decisions may reshape urban systems before implementation. Planning judgement is thereby embedded in iterative cycles of scenario construction, evaluation, monitoring, and revision rather than in one-off prediction.

Human-in-the-loop and human-on-the-loop arrangements become forms of supervisory control through which planners interrogate assumptions, assess consequences, and decide when machine-generated actions should be constrained or overridden. This expands professional practice from producing spatial artefacts to shaping acceptable system behaviour in accordance with safety, spatial justice, democratic legitimacy, and long-term resilience.

In practical terms, this repositioning requires planners and city officials to specify acceptable operating ranges, such as congestion, noise, air-quality, or flood-risk thresholds, as explicit and publicly documented inputs rather than leaving them embedded in vendor defaults. Procurement contracts should define override procedures, operational-logging requirements, audit rights, and responsibilities before deployment. Automated interventions may also require simulation-based impact assessment analogous to the assessment procedures applied to major physical development. Routine review cycles would then compare system behaviour with stated planning objectives and identify where recalibration or suspension is necessary.

Infrastructure managers, in turn, should treat control rules, model versions, and changes to set-points as regulated artefacts subject to version control, documentation, and institutional approval. This preserves an auditable record of how and why an automated system reached and enacted a particular decision.

8. Policy, Ethics, and Governance Challenges

Once urban intervention is distributed across sensors, models, operators, vendors, and public institutions, governance affects system performance as directly as technical design. Legitimacy and safety, therefore, depend on how authority, responsibility, and rights are allocated throughout deployment and operation.

This section examines six interrelated challenges: accountability, equity and bias, privacy and surveillance, sovereignty and political economy, participation and oversight, and cybersecurity and resilience. Table 4 then maps governance mechanisms and responsible actors onto the relevant stages of the architecture.

Table 4. A governance framework for closed-loop urban automation: risks, mechanisms, and responsible actors by loop stage.

Loop Stage	Principal Risk	Governance Mechanism	Primary Responsible Actors
Sensing	Pervasive surveillance; erosion of privacy and autonomy in public space	Data minimisation, purpose limitation, consent and proportionality frameworks, privacy-preserving and federated learning [44–46,75,76]	Municipal data controllers; vendors; data-protection authorities
Inference	Bias inherited from historical urban data; opacity of models	Independent audits, explainability standards, bias testing against segregation and investment patterns [73,77–79]	Developers; independent auditors; municipal AI offices
Control (set-points)	Technocratic drift; value choices hidden in vendor defaults	Transparent procurement, published set-points and tolerances, participatory review of targets	Municipal authorities; councils; participatory bodies
Actuation	Safety and irreversibility; unequal exposure to automated intervention	Fail-safe protocols, human override, liability allocation, equity impact assessment before deployment	System operators; regulators; emergency services
Feedback and learning	Model drift; entrenchment of inequity; loss of contestability over time	Operational logs, appeal and challenge procedures, periodic revalidation, data rights for affected communities [80–83]	Oversight institutions; affected communities; researchers

8.1. Accountability and Responsibility Diffusion

Delegating zoning adjustments, mobility flows, resource allocations, emergency warnings, or infrastructure responses to algorithmic processes redistributes responsibility across designers, data providers, municipal authorities, platform operators, and public agencies. Automated systems may shape outcomes, but responsibility ultimately remains distributed among the human and institutional actors who design, procure, authorise, operate, and oversee them. Accountability becomes spatially and institutionally diffuse as AI systems acquire greater capacity to shape physical outcomes. This diffusion can produce a structural accountability gap within computational urban governance, particularly when auditability, liability allocation, human override, appeal mechanisms, and public oversight are not specified in advance.

8.2. Equity and Algorithmic Bias

Spatial AI systems are often trained on historically situated urban datasets that encode patterns of segregation, uneven investment, and spatial neglect [77,78]. Without explicit corrective mechanisms, such systems risk translating historically contingent socio-spatial inequalities into scalable and seemingly objective computational logics. Data-driven spatial intelligence, therefore, can reinforce inequitable access and representation when built upon uneven data infrastructures [73,79]. In closed-loop automation, this risk is ampli-

fied because biased or incomplete representations may not only inform analysis but also shape operational decisions, control priorities, and physical interventions. Equity assessment must, therefore, address both the data entering the system and the distributional consequences of the actions it produces.

8.3. Privacy and Infrastructural Surveillance

Spatially intelligent environments depend on pervasive IoT sensing infrastructures that continuously capture and infer behaviour in public space [84,85]. The critical issue is not only data volume, but the continuous inference of movement, association, and social interaction across space and time [86,87]. As monitoring becomes infrastructural rather than episodic, longstanding boundaries between collective benefit and individual autonomy may be weakened or renegotiated [88,89]. This shift requires robust governance frameworks based on consent, proportionality, data minimisation, purpose limitation, and attention to asymmetries of power in data extraction [75,76]. Privacy-preserving architectures, including federated learning and on-device inference, offer partial technical mitigations, although they do not by themselves resolve these ethical questions [44,46].

8.4. Sovereignty and the Political Economy of Urban Data

At a structural level, questions of sovereignty become central [90,91]. Control over urban data, automation infrastructures, AI models, cloud services, and sensor–actuator systems is increasingly concentrated within public–private assemblages, raising concerns about municipal autonomy and long-term dependency on proprietary platforms [92,93]. Urban AI governance thus intersects directly with the political economy of data-driven urbanism, shaping whose spatial knowledge is prioritised, whose interests are marginalised, and who controls the feedback loops through which automated urban systems learn and adapt [93–95]. Municipalities must consequently retain meaningful authority over data access, model modification, procurement conditions, interoperability, and the termination or transfer of automated services.

8.5. Participation and Institutional Oversight

These dynamics have prompted warnings of "ghost cities" governed by closed cybernetic optimisation systems that marginalise human agency [15,96–98]. In response, human-in-the-loop, human-on-the-loop, and participatory approaches reframe intelligence as a distributed urban achievement grounded in civic oversight and institutional mediation [80–83]. Effective oversight requires transparent procurement, explainability standards, independent audits, participatory review, data rights, operational logs, fail-safe protocols, human override mechanisms, and procedures through which affected communities can challenge or appeal algorithmic decisions.

Participation should not be limited to consultation after a system has been designed. It should extend to the definition of objectives, acceptable risk thresholds, set-points, data practices, and conditions under which automated interventions may be suspended or rejected. In this way, public participation becomes part of the control architecture rather than a procedural addition outside it.

8.6. Cybersecurity, Interoperability, and Resilience

Deeper interconnection among urban AI subsystems heightens the importance of cybersecurity, interoperability, and resilience, since vulnerabilities can cascade across infrastructures [47]. A compromised sensor network, biased model, insecure actuator, or poorly validated control rule can produce consequences beyond the digital system itself, affecting mobility, safety, infrastructure reliability, and public trust. Resilience, therefore, must be understood as both an engineering and an institutional capacity. It requires

technical redundancy, secure interfaces, incident-response protocols, fallback operating modes, clear chains of authority, and the ability to suspend automated functions without disrupting essential urban services.

Table 4 translates these principles into an operational governance framework. For each stage of the automation loop, it identifies a principal risk, corresponding governance mechanisms, and the actors responsible for implementation and oversight. In doing so, it connects the control architecture presented in Figure 2 with the consequence–oversight relationship represented in Figure 3.

9. Towards a New Urban AI Framework

The contribution of the framework is not another label for connected urban technologies, but a specification of the relationships among system architecture, spatial reasoning, material consequence, and governance. It identifies the point at which an AI–IoT deployment moves beyond analytics: when model outputs are authorised to alter physical or institutional conditions and those consequences become part of subsequent decision-making.

Three implications follow. First, scale becomes a control problem because the effects of an intervention may travel beyond its point of application through networks, land-use relations, environmental processes, and public behaviour [99]. Second, changing urban conditions require continuous recalibration rather than reliance on stationary models [100]. Third, institutional response, through licensing, procurement, emergency action, public challenge, or suspension, must be analysed as part of system performance rather than as external context [4,5].

The framework also treats human–AI collaboration as an operational requirement. Urban and geospatial knowledge cannot be fully formalised; computational inference must be interpreted alongside tacit expertise, local knowledge, and ethical judgement [18,101,102]. Human agency, therefore, shapes how objectives are selected, models are interpreted, and interventions are constrained or contested.

Norm-aware and affect-aware Physical AI illustrates this requirement at the level of embodied systems, where social conventions and legal expectations influence safe operation [22,103,104]. At the representational level, multimodal urban foundation models can expand spatial reasoning while simultaneously embedding choices about visibility, authority, and whose urban realities are represented [17,84,105–111]. The framework brings these technical and political dimensions into the same analysis.

Taken together, these implications lead to the six propositions below. They pro-vide a basis for comparing deployments according to spatial propagation, non-stationarity, institutional feedback, consequence–oversight coupling, embodiment–simulation complementarity, and ex ante accountability.

9.1. The Framework in Brief

Four elements make the framework operational. Tables 1 and 2 establish the conceptual boundaries and differentiate adjacent frameworks. Figure 2 specifies the architecture and the ways in which urban conditions depart from classical control assumptions. Figures 1 and 3 and Table 3 differentiate configurations by embodiment, simulation, consequence, and oversight. Table 4 allocates risks and responsibilities across system stages. The six propositions then translate these elements into claims that can be examined empirically.

9.2. Six Propositions

To sharpen the framework beyond narrative synthesis, its core claims are stated as six propositions. Each is derived from the preceding synthesis, and each is, in principle, examinable against deployed systems.

P1—Spatial propagation. Automated urban interventions propagate beyond their point of actuation; the effectiveness and safety of closed-loop urban automation, therefore, depend on multi-scalar spatial reasoning rather than on local optimisation alone (Sections 4 and 5).

P2—Non-stationarity. Because urban plants are non-stationary, control loops relying on stationary models degrade over time; simulation, world models, and digital twins function as the anticipatory layer that compensates for this degradation (Sections 4–6).

P3—Institutional feedback. Part of the feedback path in urban automation is institutionally mediated; the delay and contestation in that path are constitutive features of the loop, not noise to be engineered away (Sections 5 and 6).

P4—Consequence–oversight coupling. As the material consequence of actuation increases, the legitimacy of automation depends less on predictive accuracy and more on contestability, auditability, liability, and human oversight (Figure 3; Sections 6 and 8).

P5—Embodiment–simulation complementarity. Embodied Physical AI and simulation-enabled spatial intelligence are complementary poles of a single architecture; mature deployments couple them, using simulation as the supervisory layer for embodied actuation (Table 3; Section 6).

P6—Ex ante accountability. Where auditability, override, liability, and appeal are not specified before deployment, closed-loop urban automation produces structural accountability gaps that post-hoc regulation cannot fully repair (Sections 6.3 and 8).

10. Limitations and Future Research

As a perspective paper, this study is conceptual and its framework has not been empirically tested; the six propositions are theoretical claims awaiting validation rather than established findings. The case material is illustrative rather than representative; it draws disproportionately on well-documented deployments in North America, Europe, Singapore, China, and India, and on English-language sources. This geographical and linguistic concentration limits the demonstrated applicability of the framework to other urban contexts, particularly cities in the Global South where sensing coverage, institutional capacity, infrastructure provision, and informality may differ substantially.

The literature engagement, although transparent in its selection logic, is narrative and purposive rather than systematic. Consequently, the distinctions drawn among adjacent frameworks represent a conceptual interpretation rather than the outcome of a bibliometric or systematic comparative analysis. Finally, the paper uses control-theoretic vocabulary partly by analogy, because the formal identification, modelling, and validation of urban systems as controllable plants remain unresolved technical and conceptual challenges.

These limitations suggest several priorities for future research. First, the six propositions should be tested through comparative case studies of deployed AI–IoT systems, including failures, interruptions, and contested deployments rather than only successful applications. Second, implementation studies of urban digital twins should examine when and how these platforms cross the boundary between decision support and automation identified in this paper. Such studies should trace how model outputs are translated into institutional decisions, standing rules, or direct physical actuation.

Third, public-sector governance research should investigate how set-points, performance tolerances, override procedures, liability provisions, operational-logging requirements, and audit rights are specified in procurement contracts and exercised in practice. Fourth, the governance framework presented in Table 4 should be evaluated with planners, infrastructure operators, regulators, technology providers, and affected communities to assess its practical relevance, institutional feasibility, and distributional implications.

Finally, comparative research should extend the framework to cities across the Global South and other under-represented contexts. In settings characterised by uneven sensing coverage, limited administrative capacity, fragmented infrastructure, or high levels of informality, the institutional segment of the feedback path may exert greater influence on system performance than the technical segment. Such research would clarify which elements of the framework are broadly transferable and which require adaptation to particular governance and infrastructural conditions.

11. Conclusions

This perspective has shown that the urban significance of AI–IoT systems intensifies when computational outputs acquire material and institutional consequences. At that point, model performance alone is insufficient: spatial propagation, changing system conditions, human behaviour, and public authority become part of the analytical problem.

The paper makes three contributions. Conceptually, it distinguishes Physical AI as a system architecture, spatial intelligence as a cognitive capacity, and closed-loop ur-ban automation as an analytical lens, while differentiating this lens from AIoT, cyber-physical systems, embodied AI, and urban digital twins (Tables 1 and 2). Analytically, it identifies four departures from classical control assumptions—non-stationary urban systems, multi-agent disturbances, spatially distributed actuation, and institutionally mediated feedback—and formalises them through six examinable propositions. Practically, it offers planners, infrastructure managers, regulators, and public authorities a governance framework for defining set-points, oversight, override procedures, liability, audit rights, and conditions for suspension before deployment.

The future of automated cities will, therefore, depend less on technological inevitability than on deliberate institutional and design choices that align innovation with spatial justice, infrastructural resilience, public trust, and democratic legitimacy.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/automation7040122/s1, File S1: Prompts used for creating Figures 1–3.

Author Contributions: Conceptualization, A.T.; methodology, A.T.; investigation, A.T. and Y.Q.; formal analysis, A.T.; resources, A.T. and Y.Q.; writing—original draft preparation, A.T.; writing—review and editing, A.T. and Y.Q.; visualization, A.T.; supervision, Y.Q.; project administration, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Acknowledgments: The author acknowledges the use of ChatGPT 5.5 solely in enhancing the clarity and language of this work and creating conceptual images (Figures 1–3). The outputs generated were carefully reviewed and edited by the authors, who accept full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Weng, Q.; Li, Z.; Cao, Y.; Lu, X.; Gamba, P.; Zhu, X.; Xu, Y.; Zhang, F.; Qin, R.; Yang, M.Y.; et al. How will AI transform urban observing, sensing, imaging, and mapping? *npj Urban Sustain.* **2024**, *4*, 50. [CrossRef]
2. Yue, Y.; Yan, G.; Lan, T.; Cao, R.; Gao, Q.; Gao, W.; Huang, B.; Huang, G.; Huang, Z.; Kan, Z.; et al. Shaping future sustainable cities with AI-powered urban informatics: Toward human-AI symbiosis. *Comput. Urban Sci.* **2025**, *5*, 31. [CrossRef]

3. Batty, M. Artificial intelligence and smart cities. *Environ. Plan. B Urban Anal. City Sci.* **2018**, *45*, 3–6. [CrossRef]
4. Cugurullo, F. Urban artificial intelligence: From automation to autonomy in the smart city. *Front. Sustain. Cities* **2020**, *2*, 38. [CrossRef]
5. Mattern, S.; Hopkins, H.R. A City Is Not a Computer: Other Urban Intelligences. *Inf. Cult.* **2022**, *57*, 343–346.
6. Amoore, L. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*; Duke University Press: Durham, NC, USA, 2020.
7. Das, D.; Kwek, B. AI and data-driven urbanism: The Singapore experience. *Digit. Geogr. Soc.* **2024**, *7*, 100104. [CrossRef]
8. Karvonen, A. Smart cities, knowledge generation and the enduring pursuit of urban innovation. In *Politics of Urban Knowledge*; Routledge: New York, NY, USA, 2023; pp. 253–269.
9. Salehi, V. Fundamentals of Physical AI. *J. Intell. Syst. Syst. Lifecycle Manag.* **2025**, *2*. [CrossRef]
10. Jaakkola, E. Designing conceptual articles: Four approaches. *AMS Rev.* **2020**, *10*, 18–26. [CrossRef]
11. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
12. Kitchin, R.; Dodge, M. *Code/Space: Software and Everyday Life*; MIT Press: Cambridge, MA, USA, 2014.
13. Bibri, S.E. The anatomy of the data-driven smart sustainable city: Instrumentation, datafication, computerization and related applications. *J. Big Data* **2019**, *6*, 59. [CrossRef]
14. 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.
15. Cugurullo, F.; Caprotti, F.; Cook, M.; Karvonen, A.; McGuirk, P.; Marvin, S. The rise of AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence. *Urban Stud.* **2024**, *61*, 1168–1182.
16. Cook, M.; Karvonen, A. Urban planning and the knowledge politics of the smart city. *Urban Stud.* **2024**, *61*, 370–382.
17. Zhang, W.; Han, J.; Xu, Z.; Ni, H.; Liu, H.; Xiong, H. Urban foundation models: A survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 6633–6643.
18. Marasinghe, R.; Yigitcanlar, T.; Mayere, S.; Washington, T.; Limb, M. Towards responsible urban geospatial AI: Insights from the white and grey literatures. *J. Geovis. Spat. Anal.* **2024**, *8*, 24. [CrossRef]
19. Xu, E.; Wang, H.; Zhang, Y.; Li, S.; Tang, Y.; Zhou, Z.; Lin, Y.; Yuan, Y.; Fan, X.; Ding, J.; et al. A Survey of Physics-Informed AI for Complex Urban Systems. *arXiv* **2025**, arXiv:2506.13777.
20. Dong, Z.; Wang, H.; Chen, Z.; Long, C.; Peng, Y.; Liu, Y.; Liang, F.; Zhou, J.; Chen, Y.; Zhang, F.; et al. The neural city: A next-generation spatio-temporal intelligence paradigm for urban holistic governance. *Innovation* **2025**, *7*, 101149. [CrossRef]
21. Randle, D.; Rasheed, A.; Cotting, D.; Pivovar, R. *Physical AI: Building the Next Foundation in Autonomous Intelligence*; AWS: Seattle, WA, USA, 2025. Available online: https://aws.amazon.com/blogs/spatial/physical-ai-building-the-next-foundation-in-autonomous-intelligence/ (accessed on 26 March 2026).
22. Ray, P.P. Physical AI: Bridging the sim-to-real divide toward embodied, ethical, and autonomous intelligence. *Mach. Learn. Comput. Sci. Eng.* **2026**, *2*, 1. [CrossRef]
23. Nathan, M.J. Disembodied AI and the limits to machine understanding of students' embodied interactions. *Front. Artif. Intell.* **2023**, *6*, 1148227. [CrossRef] [PubMed]
24. Paolo, G.; Gonzalez-Billandon, J.; Kégl, B. A call for embodied AI. *arXiv* **2024**, arXiv:2402.03824.
25. Hegarty, M. Components of spatial intelligence. In *Psychology of Learning and Motivation*; Academic Press: Cambridge, MA, USA, 2010; pp. 265–297.
26. Gruen, A. SMART Cities: The need for spatial intelligence. *Geo-Spat. Inf. Sci.* **2013**, *16*, 3–6. [CrossRef]
27. Feng, J.; Zeng, J.; Long, Q.; Chen, H.; Zhao, J.; Xi, Y.; Zhou, Z.; Yuan, Y.; Wang, S.; Zeng, Q.; et al. A survey of large language model-powered spatial intelligence across scales: Advances in embodied agents, smart cities, and earth science. *arXiv* **2025**, arXiv:2504.09848.
28. Janowicz, K.; Gao, S.; McKenzie, G.; Hu, Y.; Bhaduri, B. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 625–636.
29. Hao, S.; Han, W.; Jiang, T.; Li, Y.; Wu, H.; Zhong, C.; Zhou, Z.; Tang, H. Synthetic Data in AI: Challenges, Applications, and Ethical Implications. *arXiv* **2024**, arXiv:2401.01629.
30. Chen, J.; Wang, H.; Li, J.; Liu, Y.; Dong, Z.; Yang, B. SpatialLLM: From Multi-modality Data to Urban Spatial Intelligence. *arXiv* **2025**, arXiv:2505.12703.
31. Kakko, A. Embodied Intelligence: Grounding AI in the Physical World for Enhanced Capability and Adaptability. 2025. Available online: https://www.alphanome.ai/post/embodied-intelligence-grounding-ai-in-the-physical-world-for-enhanced-capability-and-adaptability (accessed on 29 March 2026).
32. Zhang, Y.; Tian, J.; Xiong, Q. A review of embodied intelligence systems: A three-layer framework integrating multimodal perception, world modeling, and structured strategies. *Front. Robot. AI* **2025**, *12*, 1668910. [CrossRef] [PubMed]
33. Batty, M. Defining urban science. In *Urban Informatics*; Springer: Singapore, 2021; pp. 15–28.

34. Bai, Y.; Wang, M.; Yan, Y.; Wang, H. Exploring the impact of 2D/3D urban morphology on land surface temperature within the diurnal cycle in Tianjin. *Sci. Rep.* **2025**, *15*, 39740. [CrossRef] [PubMed]
35. Alshahrani, A. Bridging cities and citizens with generative AI: Public readiness and trust in urban planning. *Buildings* **2025**, *15*, 2494. [CrossRef]
36. UN-Habitat. AI for Spatial Mapping and Analysis: GeoAI Toolkit for Urban Planners. 2025. Available online: https://unhabitat.org/sites/default/files/2025/12/ai_for_spatial_mapping_and_analysis.pdf (accessed on 4 April 2026).
37. Seto, K.C.; Güneralp, B.; Hutyra, L.R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16083–16088. [CrossRef] [PubMed]
38. Bai, X.; Dawson, R.J.; Ürge-Vorsatz, D.; Delgado, G.C.; Salisu Barau, A.; Dhakal, S.; Dodman, D.; Leonardsen, L.; Masson-Delmotte, V.; Roberts, D.C.; et al. Six research priorities for cities and climate change. *Nature* **2018**, *555*, 23–25. [CrossRef] [PubMed]
39. Knoblauch, S.; Li, H.; Biljecki, F.; Li, W.; Zipf, A. Urban AI for a sustainable built environment: Progress and future directions. *Environ. Plan. B Urban Anal. City Sci.* **2026**, *53*, 23998083261418565. [CrossRef]
40. Liu, W.; Yang, Z.; Gui, C.; Li, G.; Xu, H. Investigating the Nonlinear Relationship Between the Built Environment and Urban Vitality Based on Multi-Source Data and Interpretable Machine Learning. *Buildings* **2025**, *15*, 1414. [CrossRef]
41. Feng, F.; Ren, Y.; Xu, C.; Jia, B.; Wu, S.; Lafortezza, R. Exploring the non-linear impacts of urban features on land surface temperature using explainable artificial intelligence. *Urban Clim.* **2024**, *56*, 102045. [CrossRef]
42. Dritsas, E.; Trigka, M. Advances in geospatial artificial intelligence for remote sensing applications. *Comput. Sci. Rev.* **2026**, *60*, 100913. [CrossRef]
43. Feng, Z.; Fan, M.; Liu, B.; Xu, S.; Xiong, H. End-to-End Generation of City-Scale Vectorized Maps by Crowdsourced Vehicles. *arXiv* **2025**, arXiv:2507.08901.
44. Kapoor, A.; Kumar, D. Enhancing Smart Cities with Federated Learning: A Framework for Secure, Scalable, and Intelligent Urban Sensing Systems. *IEEE Internet Things Mag.* **2025**, *9*, 96–103. [CrossRef]
45. Raza, A.; Badidi, E.; El Harrouss, O. A Reputation-Aware Adaptive Incentive Mechanism for Federated Learning-Based Smart Transportation. *Smart Cities* **2026**, *9*, 27. [CrossRef]
46. Tariq, A.; Serhani, M.A.; Sallabi, F.M.; Barka, E.S.; Qayyum, T.; Khater, H.M.; Shuaib, K.A. Trustworthy federated learning: A comprehensive review, architecture, key challenges, and future research prospects. *IEEE Open J. Commun. Soc.* **2024**, *5*, 4920–4998. [CrossRef]
47. Lyu, Q.; Liu, S.; Shang, Z. Securing Urban Landscape: Cybersecurity Mechanisms for Resilient Smart Cities. *IEEE Access* **2024**, *13*, 10966–10977.
48. Acheampong, R.A.; Cugurullo, F.; Staricco, L.; Brovarone, E.V. Autonomous mobility transitions—Socio-spatial dimensions and the role of urban planning and policy. *Cities* **2024**, *152*, 105184. [CrossRef]
49. Yigitcanlar, T. *Autonomous Urban Mobility: Understanding Adoption Parameters, Perceptions, Perspectives*; CRC Press: Boca Raton, FL, USA, 2025.
50. Jin, H.; Shepardson, D. *California Sidelines GM Cruise's Driverless Cars, Cites Safety Risk*; Reuters: New York, NY, USA, 2023. Available online: https://www.reuters.com/business/autos-transportation/california-suspends-gm-cruises-driverless-autonomous-vehicle-permits-2023-10-24/ (accessed on 5 July 2026).
51. Waymo. *That's a Wrap! Waymo's 2024 Year in Review*; Waymo: Mountain View, CA, USA, 2024. Available online: https://waymo.com/blog/2024/12/year-in-review-2024/ (accessed on 10 July 2026).
52. Argota Sánchez-Vaquerizo, J. Getting real: The challenge of building and validating a large-scale digital twin of Barcelona's traffic with empirical data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 24.
53. Ghosh, S.; Sarkar, S.K.; Roy, P. Application of automation and artificial intelligence (AI) in green transportation system. In *Driving Green Transportation System Through Artificial Intelligence and Automation: Approaches, Technologies and Applications*; Springer Nature: Cham, Switzerland, 2025; pp. 21–42.
54. City of Helsinki. *Helsinki 3D*; City of Helsinki: Helsinki, Finland, 2025.
55. World Geospatial Industry Council. *Digital Synergy Project, Helsinki, Finland*; WGIC: The Hague, The Netherlands, 2024.
56. Dembski, F.; Wössner, U.; Letzgus, M.; Ruddat, M.; Yamu, C. Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany. *Sustainability* **2020**, *12*, 2307. [CrossRef]
57. HLRS. *Using Digital Twins to Design More Sustainable Cities*; High-Performance Computing Center: Stuttgart, Germany, 2020.
58. Hawkins, A. *Humanoid Workers and Surveillance Buggies: "Embodied AI" Is Reshaping Daily Life in China*; The Guardian: London, UK, 2025.
59. Government Technology Agency of Singapore. *NParks and SNDGG Trial SPOT Robot for Safe Distancing Operations at Bishan–Ang Mo Kio Park*; Government Technology Agency of Singapore: Singapore, 2020.
60. Knightscope. *K5 Autonomous Security Robot*; Knightscope: Sunnyvale, CA, USA, 2026. Available online: https://aihumanlove.com/tools/robots/knightscope-k5/ (accessed on 2 July 2026).
61. Simon, M. *The Tricky Ethics of Knightscope's Crime-Fighting Robots*; Wired: New York, NY, USA, 2017.

62. Burkett, N.J. *Robot to Help Patrol Times Square Subway Station*; ABC7 New York: New York, NY, USA, 2023.
63. Yang, M. *Texas's San Antonio Airport Will Get a 420lb Autonomous Security Robot*; The Guardian: London, UK, 2024.
64. Mahesh, K. *'Physical Intelligence' to Help T Solve Urban Issues*; Times of India: Mumbai, India, 2025. Available online: https://timesofindia.indiatimes.com/city/hyderabad/physical-intelligence-to-help-t-solve-urban-issues/articleshow/125470624.cms (accessed on 31 July 2026).
65. Jain, H.; Dhupper, R.; Shrivastava, A.; Kumar, D.; Kumari, M. AI-enabled strategies for climate change adaptation: Protecting communities, infrastructure, and businesses from the impacts of climate change. *Comput. Urban Sci.* **2023**, *3*, 25. [CrossRef]
66. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
67. Nearing, G.; Cohen, D.; Dube, V.; Gauch, M.; Gilon, O.; Harrigan, S.; Hassidim, A.; Klotz, D.; Kratzert, F.; Metzger, A.; et al. AI increases global access to reliable flood forecasts. *arXiv* **2023**, arXiv:2307.16104.
68. Google Research. *Flood Hub: Expert Mode*; Google Research: Mountain View, CA, USA, 2024. Available online: https://sites.research.google/floodforecasting/?web=1&wdLOR=cA69789BC-DB44-BD4D-B8B7-22D0CAAF4E5B (accessed on 29 March 2026).
69. ALERTCalifornia. *ALERTCalifornia and CAL FIRE's Fire Detection AI Program Named One of TIME's Best Inventions of 2023*; ALERTCalifornia: La Jolla, CA, USA, 2023.
70. California Wildfire & Forest Resilience Task Force. *Using AI to Stay Ahead of Wildfires and Other Disasters*; California Wildfire & Forest Resilience Task Force: Sacramento, CA, USA, 2023.
71. Benjamin, R. *Race After Technology: Abolitionist Tools for the New Jim Code*; Polity Press: Medford, OR, USA, 2019.
72. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. Tackling climate change with machine learning. *ACM Comput. Surv.* **2022**, *55*, 1–96. [CrossRef]
73. Sanchez, T.W.; Brenman, M.; Ye, X. The ethical concerns of artificial intelligence in urban planning. *J. Am. Plan. Assoc.* **2025**, *91*, 294–307. [CrossRef]
74. Erikstad, L.; Simensen, T.; Bakkestuen, V.; Halvorsen, R. Index measuring land use intensity—A gradient-based approach. *Geomatics* **2023**, *3*, 188–204. [CrossRef]
75. Richards, N.M.; King, J.H. Big data ethics. *Wake For. Law Rev.* **2014**, *49*, 393.
76. Taylor, L.; Floridi, L.; van der Sloot, B. *Group Privacy: New Challenges of Data Technologies*; Springer: Cham, Switzerland, 2017.
77. Xu, F.; Wang, Q.; Moro, E.; Chen, L.; Salazar Miranda, A.; González, M.C.; Tizzoni, M.; Song, C.; Ratti, C.; Bettencourt, L.; et al. Using human mobility data to quantify experienced urban inequalities. *Nat. Hum. Behav.* **2025**, *9*, 654–664. [CrossRef] [PubMed]
78. Müürisepp, K.; Järv, O.; Tammaru, T.; Toivonen, T. Activity spaces and big data sources in segregation research: A methodological review. *Front. Sustain. Cities* **2022**, *4*, 861640. [CrossRef]
79. Howe, B.; Brown, J.M.; Han, B.; Herman, B.; Weber, N.; Yan, A.; Yang, S.; Yang, Y. Integrative urban AI to expand coverage, access, and equity of urban data. *Eur. Phys. J. Spec. Top.* **2022**, *231*, 1741–1752. [CrossRef] [PubMed]
80. Herrmann, T.; Pfeiffer, S. Keeping the organization in the loop: A socio-technical extension of human-centered artificial intelligence. *AI Soc.* **2023**, *38*, 1523–1542. [CrossRef]
81. Griffen, Z.; Owens, K. From "Human in the Loop" to a Participatory System of Governance for AI in Healthcare. *Am. J. Bioeth.* **2024**, *24*, 81–83. [CrossRef] [PubMed]
82. Kudina, O.; van de Poel, I. A sociotechnical system perspective on AI. *Minds Mach.* **2024**, *34*, 21. [CrossRef]
83. Johnson, D.G.; Verdicchio, M. The sociotechnical entanglement of AI and values. *AI Soc.* **2025**, *40*, 67–76. [CrossRef]
84. Kitchin, R. The real-time city? Big data and smart urbanism. *GeoJournal* **2014**, *79*, 1–14. [CrossRef]
85. Van Zoonen, L. Privacy concerns in smart cities. *Gov. Inf. Q.* **2016**, *33*, 472–480. [CrossRef]
86. Zuboff, S. Big other: Surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* **2015**, *30*, 75–89. [CrossRef]
87. Ruppert, E.; Isin, E.; Bigo, D. Data politics. *Big Data Soc.* **2017**, *4*, 2053951717717749. [CrossRef]
88. Amoore, L. Data derivatives: On the emergence of a security risk calculus for our times. *Theory Cult. Soc.* **2011**, *28*, 24–43.
89. Lyon, D. *The Culture of Surveillance: Watching as a Way of Life*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
90. Badawy, W. Algorithmic sovereignty and democratic resilience: Rethinking AI governance in the age of generative AI. *AI Ethics* **2025**, *5*, 4855–4862.
91. Susnjara, S.; Smalley, I. *AI Sovereignty*; IBM Think: Boston, MA, USA, 2026. Available online: https://www.ibm.com/think/topics/ai-sovereignty (accessed on 2 July 2026).
92. Sadowski, J. The internet of landlords: Digital platforms and new mechanisms of rentier capitalism. *Antipode* **2020**, *52*, 562–580. [CrossRef]
93. Zuboff, S. The age of surveillance capitalism. In *Social Theory Re-Wired*; Routledge: New York, NY, USA, 2023; pp. 203–213.
94. Sadowski, J.; Pasquale, F. The spectrum of control: A social theory of the smart city. *First Monday* **2015**, *20*, 5903. [CrossRef]

95. Cardullo, P.; Kitchin, R. Being a 'citizen' in the smart city: Up and down the scaffold of smart citizen participation in Dublin, Ireland. *GeoJournal* **2019**, *84*, 1–13. [CrossRef]
96. Kassens-Noor, E.; Hintze, A. Cities of the future? The potential impact of artificial intelligence. *AI* **2020**, *1*, 12. [CrossRef]
97. Tiwari, A. Conceptualising the emergence of Agentic Urban AI: From automation to agency. *Urban Inf.* **2025**, *4*, 13. [CrossRef]
98. Tiwari, A. Beyond automation: The emergence of agentic urban AI. *Automation* **2025**, *6*, 29. [CrossRef]
99. Zhang, S.; Dong, L.; Li, X.; Zhang, S.; Sun, X.; Wang, S.; Li, J.; Hu, R.; Zhang, T.; Wang, G.; et al. Instruction tuning for large language models: A survey. *ACM Comput. Surv.* **2026**, *58*, 1–36. [CrossRef]
100. Liu, Y.; Chen, W.; Bai, Y.; Liang, X.; Li, G.; Gao, W.; Lin, L. Aligning cyber space with physical world: A comprehensive survey on embodied AI. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 7253–7274. [CrossRef]
101. Engin, Z. Human-AI Governance (HAIG): A Trust-Utility Approach. *arXiv* **2025**, arXiv:2505.01651.
102. Santi, P. AI improves the design of urban communities. *Nat. Comput. Sci.* **2023**, *3*, 735–736. [CrossRef] [PubMed]
103. Dell'Anna, D.; Jamshidnejad, A. Sonar: An adaptive control architecture for social norm aware robots. *Int. J. Soc. Robot.* **2024**, *16*, 1969–2000. [CrossRef]
104. Mushkani, R. Urban AI Governance Must Embed Legal Reasonableness for Democratic and Sustainable Cities. *arXiv* **2025**, arXiv:2508.12174.
105. Li, X.; Dang, A. The disruptive effect of AI on urban planning. *Nat. Cities* **2025**, *2*, 568–570. [CrossRef]
106. Sanchez, T.W.; Fu, X.; Yigitcanlar, T.; Ye, X. The research landscape of AI in urban planning: A topic analysis of the literature with ChatGPT. *Urban Sci.* **2024**, *8*, 197. [CrossRef]
107. Balsebre, P.; Huang, W.; Cong, G.; Li, Y. City foundation models for learning general purpose representations from openstreetmap. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 87–97.
108. Beer, D. *The Social Power of Algorithms*; Routledge: New York, NY, USA, 2019; pp. 1–13.
109. Shelton, T.; Zook, M.; Wiig, A. The 'actually existing smart city'. *Camb. J. Reg. Econ. Soc.* **2015**, *8*, 13–25. [CrossRef]
110. Dalton, C.M.; Thatcher, J. Inflated granularity: Spatial "big data" and geodemographics. *Big Data Soc.* **2015**, *2*, 2053951715601144. [CrossRef]
111. Shit, R.C.; Subudhi, S. Privacy-Preserving Federated Learning for Fair and Efficient Urban Traffic Optimization. *arXiv* **2025**, arXiv:2511.06363.

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.