

Review

Constructing Embodied Intelligent Spaces from an Architectural Perspective: Technical Frameworks, Integration Mechanisms, and Implementation Pathways

Xin Zhou ¹, Yuping Feng ¹, Xiaokai Yan ² , Jiarui Sun ³, Yian Lu ⁴ and Ji Lu ^{5,*}

¹ The Architectural Design and Research Institute of Zhejiang University Co., Ltd., Hangzhou 310028, China; zzzzzz3611@163.com (X.Z.); fyuping@163.com (Y.F.)

² School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China; yanxiaokai@mail.nwpu.edu.cn

³ Center for Balance Architecture, Zhejiang University, Hangzhou 310028, China; sunjiarui2024@163.com

⁴ Boyer College of Music and Dance, Temple University, Philadelphia, PA 19122, USA; yian.lu@temple.edu

⁵ College of Civil Engineering and Architecture, Zijingang Campus, Zhejiang University, Hangzhou 310058, China

* Correspondence: luji@zju.edu.cn; Tel.: +86-1395717926

Abstract

To address the bottleneck of technologically advanced but cognitively limited smart buildings, this paper introduces embodied intelligence to reconstruct the intelligent paradigm of architectural space. It defines embodied intelligent space and develops a Human–Intelligence–Context framework. In this framework, space is regarded as a collaborative spatial agent with a degree of autonomy. Based on a systematic review of 293 studies published from 2017 to 2026, the paper proposes a technical adaptation loop of embodied perception, cognitive decision-making, and embodied actuation. This loop explains how space gains computability, interpretability, and actionability. The paper further develops a scenario–pathway–implementation framework from an architectural perspective. This framework clarifies how intelligent technologies can be structurally embedded into architectural systems. The findings show that embodied intelligent space supports the shift from a passive container to an autonomous spatial agent. This shift is enabled by multimodal perception, spatial large-model reasoning, and edge–cloud collaborative actuation. The study provides theoretical support and implementation pathways for sustainable and human-centered settlement environments.

Keywords: embodied intelligent space; world model; edge–cloud collaboration; spatial learning; architecture; smart building



Academic Editor: Elena Lucchi

Received: 13 August 2026

Revised: 7 September 2026

Accepted: 24 September 2026

Published: 1 October 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

With recent advances in artificial intelligence, especially large language models (LLMs), edge computing, and multimodal perception, space is evolving from a passive physical container into a collaborative intelligent agent with perception, learning, and adaptive capabilities [1]. In future human settlements, the symbiosis between humans and spatial agents may become a common condition. Space will not only accommodate life but also interact and co-evolve with people and objects [2]. Embodied intelligence theory argues that intelligence is not a computational process isolated from the body and the environment. Rather, it emerges from the coupling of perception, action, and learning in the physical world [2]. This view provides an opportunity for design science to reshape the formal logic

of human settlements. This provides an opportunity for design science to reconceptualize space itself as an intelligent carrier with embodied properties.

However, a closer examination of current intelligent human settlements reveals a growing contradiction between technological accumulation and lagging spatial cognition. Although various IoT terminals and data models have been widely deployed, most conventional intelligent spaces still rely on predefined rule-based triggers, such as If–Then logic [3], and remote passive control. In essence, this approach reduces space to a disembodied and passive container. This limitation can be observed in three aspects. First, current systems are based on disembodied computing [4]. Intelligence is often confined to the cloud or to isolated control terminals. As a result, space cannot respond to the physical environment in real time and at a fine-grained level. When facing complex and nonlinear individual behaviors, these systems usually provide rigid and static responses. Second, the perception–actuation chain remains fragmented [5]. Even in smart building systems, sensors and actuators are not supported by a unified perception–decision layer. The system can record data, but it cannot understand situations. It has not yet developed the autonomous learning and continuous evolution required of an intelligent agent [3]. Third, the relationship between humans and space remains instrumental. Space mainly acts as a tool that responds to human commands [6]. Without an embodied cognition framework, interaction becomes burdensome, and user experience is often fragmented.

Although significant progress has been achieved in various fields such as smart buildings, cognitive buildings, and digital twin-based environments, etc., existing studies mainly focus on individual technological improvements or system-level optimization. A comprehensive understanding of how these technologies can be integrated to enable spatial environments with autonomous perception, reasoning, and adaptive action is still lacking. Therefore, there is a need for a systematic review to clarify the conceptual foundation, technological components, and future development directions of embodied intelligent space.

Moving beyond the conventional logic of automation, this paper introduces embodied AI to redefine space itself. First, it defines the theoretical meaning of embodied intelligent space and proposes a Human–Intelligence–Context system as a technical framework with capacities for computation, understanding, and action. Second, based on this framework, the study systematically reviews 293 papers published from 2017 to 2026. It traces the development trajectory and integration mechanisms of artificial intelligence in human settlement environments. It also clarifies the spatial deployment pathways of embodied perception, behavior prediction and decision-making, and embodied actuation within the Human–Intelligence–Context system. Third, the paper identifies key issues and adaptive challenges for architecture in response to the convergence of artificial intelligence, equipment manufacturing, and related disciplines. It further proposes future scenarios and implementation pathways with a clear architectural structure.

Based on this background, the study addresses the following research questions:

RQ1: What are the conceptual foundations and agent characteristics of embodied intelligent space as a new paradigm for future human settlements?

RQ2: What system architecture and collaborative mechanisms support the operation of spatial embodied cognition?

RQ3: What future design trends may emerge from the deep integration of architectural approaches and embodied intelligence technologies?

The structure of this study is as follows. Section 2 defines the conceptual foundation and mechanisms of embodied intelligent space. Section 3 presents the review method, which combines hierarchical classification with thematic analysis for data extraction, synthesis, and interpretation. Section 4 systematically examines the adaptation mechanism of embodied intelligent space from three aspects: adaptation modes, technical logic, and

implementation framework. It also discusses its implementation pathways in real human settlements. Section 5 explores future development trends of embodied intelligent space. Section 6 examines the architectural challenges that may arise in its application and proposes corresponding strategies. Section 7 concludes the paper.

2. The Basic Framework of Embodied Intelligent Space

2.1. *Reconstructing Space Through Embodied Intelligence: From External Embedding to Endogenous Intelligence*

Traditional intelligent spaces usually rely on smart devices added after construction. These devices are logically separated from the architecture. Their operation is often based on predefined rules. With technological advances, intelligence is no longer a functional layer attached to space. It is becoming a basic attribute, alongside structural and building service systems. It should be integrated into the overall construction logic from the early design stage. In this sense, space shifts from a technical container to an integrated cognition–actuation system. It can continuously perceive the environment, understand situations, and respond actively.

Drawing on relevant definitions of embodied agents [2,5], this paper defines embodied intelligent space as a bounded physical space equipped with intelligent systems. Through multimodal sensors, it continuously perceives itself and its environment, performs real-time computation and decision-making, and actuates devices within the space to interact with indoor environmental conditions, occupants, objects, and the surrounding environment. These processes enable the space to accomplish specific tasks. Its core capacity lies in intelligent behavior that emerges from continuous interaction with these environmental and spatial conditions.

2.2. *From Human–Machine Communication to the Human–Intelligence–Context System*

Human–machine interaction and communication theory originally evolved from classic cybernetic [7] and information processing models, which viewed human interaction with machines as an encoder–decoder channel focused on information exchange and feedback control. Before the formalization of integrated system ergonomics, early cognitive and interaction studies analyzed technical artifacts as physical or cognitive mediating tools supporting the classic human execution–evaluation action cycle [8]. In these foundational paradigms, physical controls and perceptual feedback were often inherently coupled, focusing primarily on isolated operator–machine efficiency and signal accuracy. As systems grew in complexity, communication expanded beyond rigid symbolic exchange to encompass social, affective, and non-symbolic sign interactions. Building upon these cumulative insights into system ergonomics, cybernetics, and human factors, the comprehensive “Man–Machine–Environment” (MME) dynamic framework was officially established in 1981, unifying human cognitive capabilities, mechanical executing systems, and physical operational conditions into a holistic triad [9].

Although the MME framework successfully integrated human factors with mechanical control and physical boundary conditions, its classical formulation exhibits distinct limitations when confronted with modern agentic technologies. Historically, MME treated the machine as a passive executing artifact or a decoupled symbol interface [8], while the environment served primarily as a static physical container or an external background providing environmental stress. Communication was predominantly modeled as a linear, deterministic transmission loop, where non-symbolic cues, ambient spatial intelligence, and dynamic contextual signals were often marginalized or abstracted away [10,11]. Under current spatial intelligence paradigms, however, computing environments are no longer passive physical backdrops; rather, they function as autonomous, sensing–cognitive agents

capable of spatial perception and proactive actuation. Furthermore, modern intelligent systems transcend basic closed-loop control to operate via non-deterministic, agentic, and multimodal human–machine communication. Consequently, the classical MME system struggles to accommodate the spatial agency, socio-affective interaction channels, and dynamic spatial-temporal context characteristic of modern intelligent environments.

Based on this understanding, this paper reinterprets the core elements of the framework while retaining its systemic view. It proposes a structural analytical framework of Human–Intelligence–Context (Figure 1). In this framework, “machine” is integrated into the intelligent space embedded in space. “Environment” is extended from the physical environment to a higher-level contextual interaction field. Together, the three elements form the basic structural unit of embodied intelligent space.

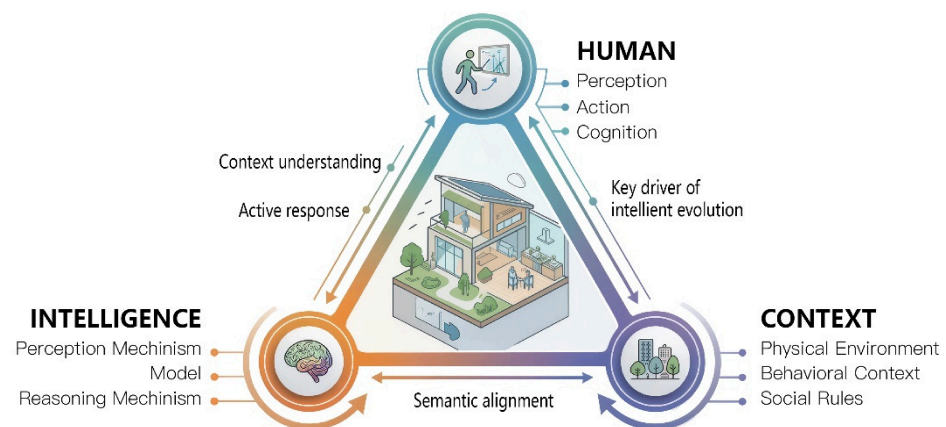


Figure 1. Closed-loop interactive relationship among the three elements: human, intelligence and context.

Human refers to the initiator of behavior and a regulator of the environment. Humans continuously participate in spatial interaction through perception, behavior, and cognition [4]. They therefore become a key driving force for the evolution of spatial intelligence.

Intelligence refers to the intelligent system embedded in space, namely intelligent space. It consists of perception, models, and reasoning mechanisms [12]. These components enable space to understand situations and respond actively.

Context refers to the contextual interaction field in which the human–intelligence system is embedded. It includes the physical environment, behavioral context, social rules, and related factors. It is also dynamically generated and reshaped during interaction.

Together, these three elements form a composite network of behavioral co-evolution, semantic alignment, and strategic coordination. This paradigm follows the embodied cognition logic that cognition is embedded in the body, and the body is embedded in the environment [13]. It further extends this logic to architectural spatial systems and gives space partial capabilities for decision-making and action.

2.3. Aligning Spatial Needs with Technical Support

In the Human–Intelligence–Context system described above, space is no longer regarded as a passive physical container. Instead, it becomes an intelligent carrier with agency that collaborates and coexists with humans. Within this system, spatial intelligence operates through three processes: perceiving the states of humans and the environment, interpreting information through analysis, reasoning, and decision-making, and acting upon the environment. These processes correspond to the core idea of perception–decision-making–actuation coupling in embodied cognition theory [2]. Accordingly, the intelligent adaptation of space requires three key features: computability, interpretability, and action-

ability. These features form the basis for the effective operation of intelligent space and constitute the core requirements for applying embodied intelligence technologies to spatial adaptation (Figure 2).

Mechanism for applying embodied intelligence technologies to spatial adaptation

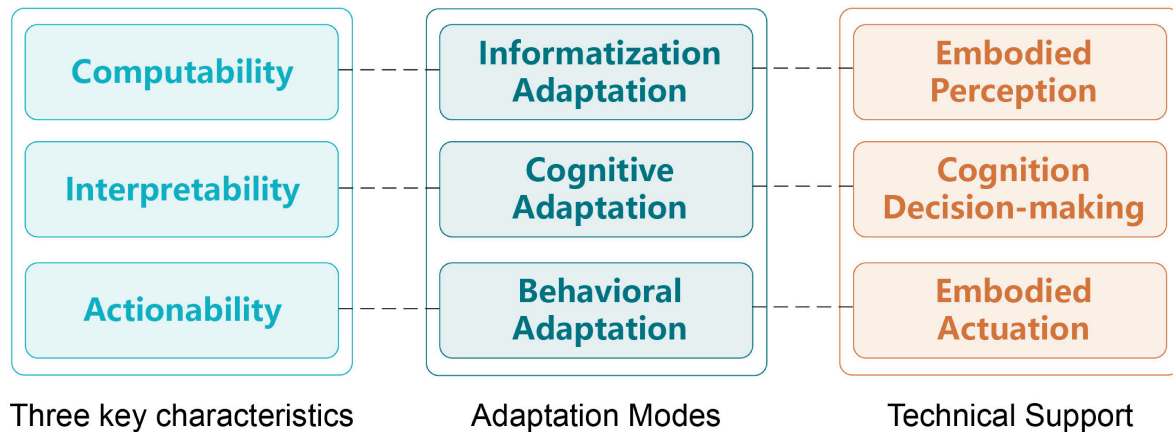


Figure 2. The mechanism for applying embodied intelligence technologies to spatial adaptation.

Spatial computability: Space needs the capacity to continuously monitor the environment through multimodal perception systems. It should collect data in real time and process it through computation. This capacity depends on embodied perception technologies. These technologies use sensor networks to perceive both the internal and external spatial environment [14]. They provide the necessary input for subsequent cognition and action.

Spatial interpretability: Space should not only collect data but also analyze and reason over such data through intelligent systems, thereby interpreting environmental changes and user needs. This feature requires support from cognitive decision-making technologies. Through in-depth analysis and prediction of perceived information, these technologies generate intelligent decisions [15], enabling space to actively regulate and respond in complex and dynamic environments.

Spatial actionability. Space needs to adapt spatial conditions based on cognitive decision-making in response to environmental dynamics and people's needs. This feature requires support from embodied actuation technologies, which translate cognitive decisions into concrete physical operations. Through this process, space can adjust environmental conditions and execute tasks in real-world settings [16].

To assess the feasibility of embodied intelligent space under current technical conditions, the next section provides a systematic review of related studies.

2.4. Distinguishing Embodied Intelligent Space from Existing Intelligent Building Paradigms

Although smart buildings, cognitive buildings, automated buildings, and digital twin-based systems have significantly advanced the intelligence level of the built environment, embodied intelligent space represents a further shift from function-oriented automation toward agent-oriented spatial intelligence. This theoretical shift is particularly suitable for addressing the growing complexity of human–building interaction, where the distinction lies not in the existence of individual technologies, but in the structural integration among perception, cognition, action, and continuous adaptation.

To clarify this difference and provide an analytical framework suitable for evaluating spatial embodied intelligence, this paper proposes four measurable dimensions: perception autonomy, cognitive capability, action autonomy, and evolutionary capability (Table 1).

Table 1. Conceptual differences among major intelligent building paradigms.

Paradigm	Primary Objective	Intelligence Source	Decision Mode	Physical Action	Learning Capability
Automated Building	Operational efficiency	Sensors and control logic	Rule-based [3]	Device-level control	None or limited
Smart Building	Optimization and connectivity [6]	IoT + data analytics	Data-driven optimization	System coordination	Limited adaptation
Cognitive Building	Situation understanding [12]	AI models and knowledge systems	AI-assisted reasoning	Adaptive control	Partial learning
Digital Twin-based Building	Digital representation and simulation [17]	Virtual models + real-time data	Simulation-supported decisions	Indirect control	Model updating
Embodied Intelligent Space	Autonomous spatial adaptation	Multimodal perception + intelligent agents	Context-aware reasoning	Coordinated embodied action	Continual evolution

First, perception autonomy refers to whether a spatial system can continuously acquire and fuse multimodal information rather than relying on isolated sensor inputs. Automated buildings mainly depend on predefined sensors and threshold-based monitoring, while embodied intelligent space integrates distributed sensing, behavioral understanding, and contextual perception.

Second, cognitive capability measures whether the system can interpret environmental states and generate decisions beyond predefined rules [6]. Conventional smart buildings mainly perform optimization based on historical data, while cognitive buildings introduce AI-assisted reasoning. However, embodied intelligent space further emphasizes semantic understanding, contextual reasoning, and prediction through spatial models and intelligent agents.

Third, action autonomy evaluates whether the system can independently translate decisions into coordinated physical responses. Digital twin systems primarily provide virtual representation, simulation, and decision support [17]. In contrast, embodied intelligent space establishes a closed loop from perception to decision-making and embodied actuation, enabling direct interaction with the physical environment.

Fourth, evolutionary capability describes whether intelligence can continuously improve through learning and experience accumulation. Traditional automated systems usually maintain fixed control logic, whereas embodied intelligent space supports continual learning through edge–cloud collaboration, model updating, and multi-agent knowledge sharing [12].

Therefore, embodied intelligent space should not be understood as a replacement for smart building, cognitive building, or digital twin technologies. Instead, it represents an integrated paradigm in which these technologies become embedded components of a spatial agent. The essential difference is the transition from “building as a controllable system” to “space as an adaptive intelligent entity”.

Table 1 summarizes the conceptual differences among major intelligent building paradigms.

This comparison establishes embodied intelligent space as a measurable extension of existing intelligent building paradigms rather than a purely conceptual redefinition.

3. Technical Review of Embodied Intelligent Space

From the technical perspective of embodied intelligence, this paper uses embodied perception–cognitive decision-making–embodied actuation (EPDA) as the analytical framework. Existing studies are classified and reviewed according to this framework. The three components correspond to the key capacities of space: computability, interpretability, and actionability. This establishes a link between technical pathways and spatial adaptation.

3.1. Information Sources and Search Strategy

This study adopts a systematic literature review (SLR) method to examine embodied intelligence-related technologies in urban and architectural fields. The literature search was performed in five academic databases: Scopus, Web of Science (WoS), IEEE Xplore, ACM Digital Library, and SpringerLink. The search covered studies published from 1 January 2017 to the search cut-off date of 18 March 2026. The same conceptual search strategy was applied across the five databases, with syntax adapted where required by individual indexing platforms.

The search strategy was built around two core concept groups: ‘technical concepts’ and ‘application concepts’. The two groups were connected using the Boolean operator ‘AND’. The specific search string was designed as follows.

Technical concept group:

“Embodied Intelligence” OR “Embodied AI” OR “Autonomous Robots” OR “Human-Building Interaction” OR “HBI” OR “Adaptive Architecture” OR “Human-Robot Interaction” OR “Cognitive Environments” OR “Digital Twin” OR “Multimodal Large Model” OR “Distributed Intelligent System” OR “Multi-Agent”

Application concept group:

“Built Environment” OR “Intelligent Buildings” OR “Smart Architecture” OR “Occupant Comfort” OR “Urban Space” OR “Placemaking” OR “Interactive Urban Environments” OR “Human-Centered Space” OR “Sustainable Cities”

The initial search identified 2519 relevant records. After the removal of 11 duplicate records, 2508 unique records underwent title and abstract screening, in which 2193 records were excluded according to the predefined eligibility criteria: (1) unrelated to human settlements or the built environment ($n = 1042$); (2) non-original research, such as reviews, editorials, and conference abstracts ($n = 587$); (3) absence of AI-based or computational-intelligence methods ($n = 498$); and (4) focus on pure content generation without interaction with physical space ($n = 66$). A further 22 articles were excluded after full-text assessment (pure content generation without physical-space interaction, $n = 9$; unrelated to human settlements, $n = 5$; non-original research, $n = 5$; absence of AI-based or computational-intelligence methods, $n = 3$), resulting in a final corpus of 293 studies. To avoid double counting, each excluded record was assigned one primary exclusion reason according to a predefined screening hierarchy, in which eligibility (publication type) was assessed first, followed by thematic relevance, methodological content, and physical-space interaction. The complete selection flow is reported in the PRISMA-style diagram (Figure 3).

3.2. Data Extraction and Structured Analysis Strategy

To avoid limiting the conceptual review to subjective interpretation, this study further converts the literature sample into a computable structured dataset under the EPDA framework, namely embodied perception, cognitive decision-making, and embodied actuation. Statistical results and visual outputs are then generated through a reproducible data analysis workflow.

The structured analysis uses the title, abstract, and author-provided keywords of each included article as the minimum unit of analysis. Titles were retained because they represent

the most consistently available bibliographic field across databases; however, they were not treated as the sole basis for classification. Abstracts and keywords were incorporated to capture methodological details, technical contributions, and cross-disciplinary terminology that may not be explicitly represented in article titles. Full texts were additionally consulted for records with insufficient metadata or ambiguous classifications.

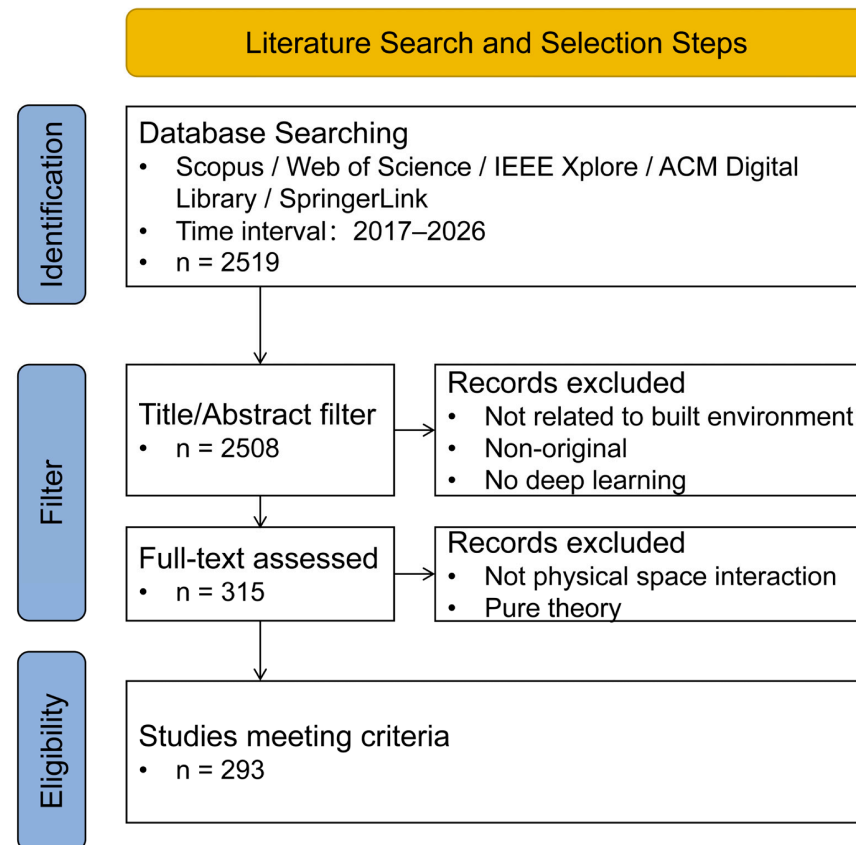


Figure 3. PRISMA-style flow diagram of literature identification, screening, eligibility assessment, and inclusion. The diagram summarizes records identified from five databases (Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and SpringerLink; $n = 2519$), duplicate removal ($n = 11$), title/abstract screening ($n = 2508$), full-text assessment ($n = 315$), and final inclusion ($n = 293$), with the number of records excluded under each criterion annotated at each stage.

The textual fields were normalized using a predefined preprocessing pipeline consisting of filename-suffix and punctuation removal, underscore replacement, lowercasing, whitespace collapsing, and tokenization. Each study was then coded under the embodied perception–cognitive decision-making–embodied actuation (EPDA) framework using a hybrid procedure that combined rule-based technical dictionaries with TF-IDF-based text weighting. The rule layer counted the number of matches between the normalized corpus text and three stage-specific keyword dictionaries (25 perception terms, 23 decision-making terms, and 25 actuation terms). The TF-IDF layer was configured with English stop-word removal, unigram-to-bigram features, sublinear term-frequency scaling, and a document frequency ceiling of 0.95, and stage scores were computed as the summed TF-IDF weights of dictionary-matched terms. Rule-based and TF-IDF scores were min–max normalized per record and combined with equal weights (0.5/0.5) into a composite stage score. A stage was assigned when its composite score reached or exceeded a threshold of 0.5, which allows multi-label assignment.

EPDA coding was performed at two levels. First, studies were allowed to receive multiple stage labels when their substantive contributions covered more than one compo-

nent of the perception–decision–actuation loop; this multi-label representation was used to characterize cross-stage integration. Second, a primary EPDA label was assigned to each study for mutually exclusive distribution and temporal analyses. The primary label was determined by the highest composite classification score; when the score margin between the top two stages fell below 0.08, or when rule-based and TF-IDF rankings disagreed, the record was referred to manual adjudication based on the full abstract and, where necessary, the full text. In total, 23 records required manual adjudication.

To assess the robustness of the computational coding procedure, a stratified random validation subset of 59 studies (approximately 20% of the corpus, stratified by primary EPDA label) was independently re-coded by two coders using the title, abstract, and keywords. Inter-rater reliability was substantial (Cohen’s $\kappa = 0.79$; percentage agreement = 86.4%), and the eight disagreements were resolved through discussion with reference to the full text.

Together, the transparent screening protocol, the reproducible coding scheme, the multilabel treatment of cross-stage studies, and the independent validation strengthen the methodological rigor of the review. The structured procedure enables heterogeneous research in smart buildings, embodied AI, human–building interaction, spatial reasoning, and distributed actuation to be examined within a common perception–decision–actuation framework, providing an evidence base for assessing where current research is concentrated, where cross-stage integration is emerging, and where technical gaps remain in translating artificial intelligence into architectural spatial agency.

After the 293 studies were fully coded with the expanded corpus and documented rules, the literature classification system was regenerated, as shown in Figure 4 [14,16,18–35]. The annual publication trend according to primary EPDA labels was also generated, as shown in Figure 5. Together, the classification system and publication trend provide the data basis and structural framework for the following analysis, trend assessment, and challenge mapping.

3.3. Statistical Analysis of Research Directions

Based on the structured coding analysis of the literature, the three stages of embodied perception, cognitive decision-making, and embodied actuation show clear differences and complementarity in research attention and technical pathways (Figure 6). In terms of the overall distribution of primary labels, cognitive decision-making accounts for the largest share of coded studies (40%, $n = 118$), followed by embodied actuation (38%, $n = 111$) and embodied perception (22%, $n = 64$). This indicates that the algorithmic core of context interpretation and decision generation has become a major research focus in recent years, while the substantial share of actuation studies reflects the rapid growth of edge–cloud collaboration and intelligent building-control research. Under the multi-label coding, 70 studies (23.9%) substantively covered more than one stage, with perception–decision combinations being the most frequent ($n = 21$), followed by decision–actuation ($n = 19$), perception–actuation ($n = 15$), and studies covering all three stages ($n = 15$). The emergence of these cross-stage studies further suggests that a systematic closed-loop linkage is taking shape. Embodied perception mainly focuses on occupant behavior and IoT-based environmental perception. Its technical pathways are dominated by CNNs and HBI perception frameworks, and are shifting from data acquisition toward the spatial representation of behavioral semantics. Cognitive decision-making mainly relies on LLMs and VLMs, and is moving from conventional optimization toward spatial relational reasoning based on multimodal large models. Embodied actuation focuses on edge–cloud collaboration, task orchestration, safety, and trustworthiness, with an emphasis on engineering issues such as low-latency implementation at the edge and autonomous action.

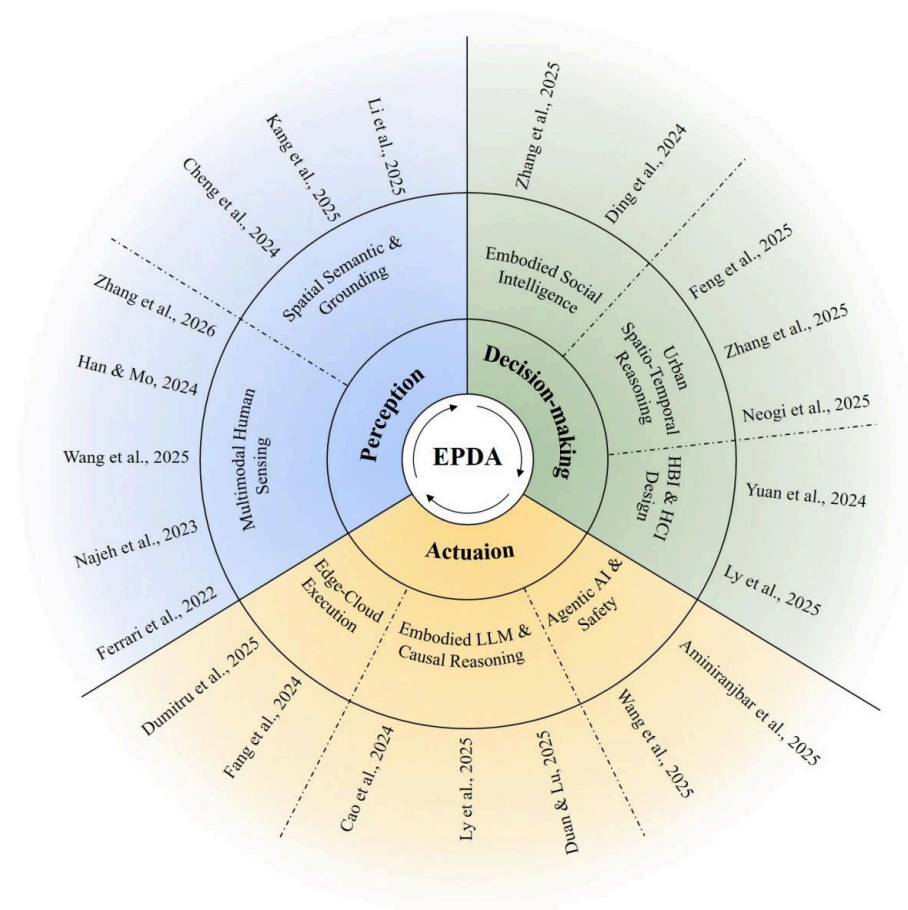


Figure 4. The diagram maps representative technical domains and studies onto embodied intelligent space—Perception, Decision, and Action—showing how data acquisition feeds analytics and decision-making to drive operational services [14,16,18–35].

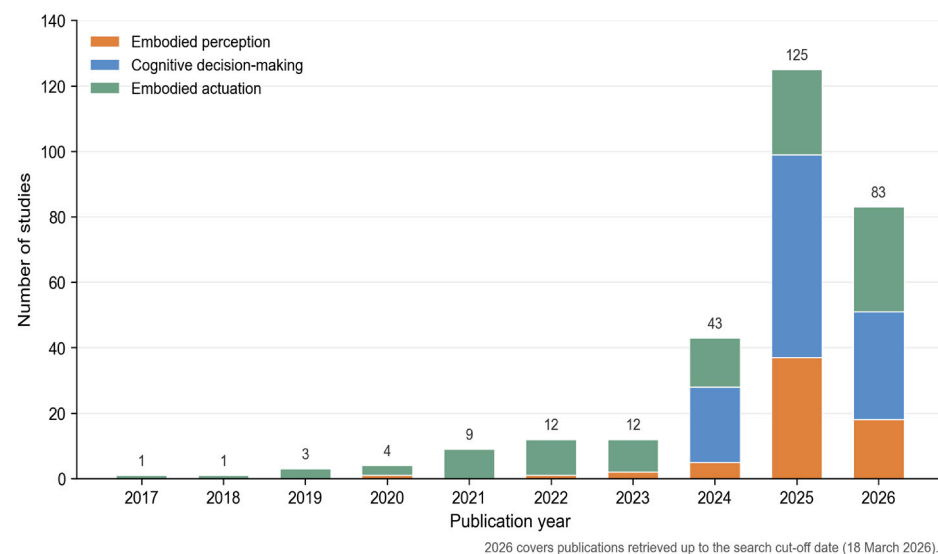


Figure 5. Annual distribution of reviewed studies according to their primary EPDA classification (2017–2026). Stacked bars show counts of reviewed studies classified as embodied perception, cognitive decision-making, and embodied actuation, highlighting the recent rise in decision-making research and the continuous growth of actuation-oriented studies. The 2026 column covers publications retrieved up to the search cut-off date (18 March 2026).

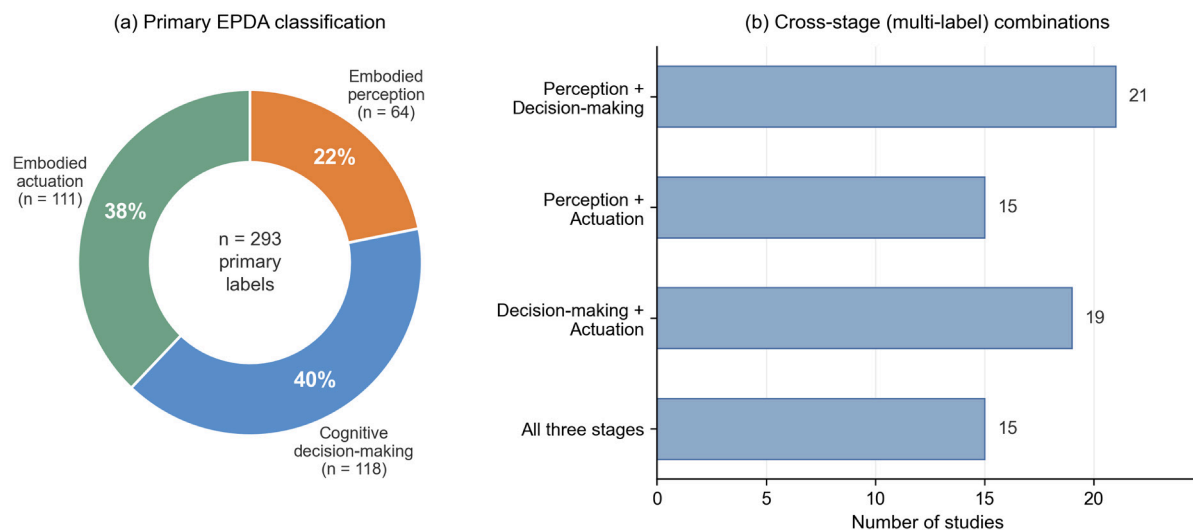


Figure 6. Distribution of reviewed studies according to their primary EPDA classification (a) and cross-stage multi-label combinations (b). Percentages in (a) refer to primary labels used for mutually exclusive statistics; combinations in (b) characterize studies substantively covering more than one EPDA stage.

4. Adaptation Mechanisms and Implementation Pathways for Embodied Intelligent Space in Architecture

4.1. Development Trajectory of Artificial Intelligence in Human Settlement Spaces

The development of artificial intelligence in human settlement spaces is, in essence, a process in which architectural space evolves from an automated system into an embodied intelligent system with cognitive and decision-making capacities [36,37]. It is also a process of continuous adaptation between intelligent technologies and architectural space.

4.1.1. Stage 1 (1990–2010): Formation of Smart Buildings

During the period from 1990 to 2010, the application of artificial intelligence in human settlement spaces was mainly reflected in the formation of smart building systems. The core aim was to improve building operational efficiency through automation technologies and information systems [38].

Research in this stage mainly focused on building automation systems and environmental control technologies. It emphasized the use of sensor networks and control systems to balance occupant comfort and energy consumption. Later, smart buildings gradually introduced optimization algorithms and behavior modeling methods. These methods allowed buildings to control energy use and regulate indoor environments under dynamic conditions [39].

Studies show that building operation needs to consider environmental parameters, energy prices, user behavior, and equipment status. This led to a control framework based on computational optimization [39]. Overall, the operational logic of smart buildings in this stage still depended on predefined rules. It lacked autonomous learning and cognitive capabilities.

4.1.2. Stage 2 (2010–2020): Introduction of Artificial Intelligence into Architecture

With the development of the Internet of Things, big data, and machine learning, artificial intelligence began to be deeply integrated into building systems. This promoted a shift from automation-oriented intelligent buildings to data-driven smart buildings.

Studies show that artificial intelligence technologies, such as neural networks, expert systems, and predictive algorithms, have been widely used in building energy prediction,

equipment control, and environmental optimization. These technologies enable buildings to adapt to complex environments [40].

At the same time, smart building systems gradually shifted from conventional automation control to decision-making systems based on computational intelligence. Dynamic optimization was achieved through behavior modeling and environmental prediction [39]. In this stage, the integration between artificial intelligence and the built environment became deeper. Farzaneh et al. [41] noted that space began to show a basic level of cognitive capability. It could reason and make decisions based on historical data and real-time information.

In this stage, artificial intelligence became an important part of building systems. Space shifted from a passive control system to an intelligent system capable of autonomous optimization.

4.1.3. Stage 3 (2020–Present): Embodied Intelligence-Driven Smart Spaces

In recent years, advances in deep learning, edge computing, and multimodal perception have further expanded the role of artificial intelligence in human settlement spaces. These technologies have promoted the evolution of buildings toward smart spaces and autonomous spaces.

Luo's [36] bibliometric study shows that artificial intelligence is becoming a core direction in smart building research. Building systems are gradually gaining the capacities for self-learning, adaptation, and cross-system collaboration. This supports the transition from single control systems to integrated intelligent systems. Based on this trend, Genkin and McArthur [17] proposed the B-SMART structure as a reference model for autonomic smart buildings. This model integrates artificial intelligence, natural language processing, and big data technologies to support autonomous operation and collaborative management of building systems. Further studies show that AI-driven smart buildings are evolving into spatial intelligence systems. Space can perceive user needs through multi-source data. It can also support dynamic optimization and autonomous decision-making across the building lifecycle [37].

In this stage, artificial intelligence shifts from an auxiliary technology to an active intelligence embedded within human settlement spaces, driving the evolution of human settlement spaces toward embodied intelligent systems.

4.2. *Adaptation Modes of Embodied Intelligence in Human Settlement Spaces*

4.2.1. Spatial Informatization Adaptation

Spatial informatization adaptation is the basis for constructing embodied intelligent space. Its core aim is to transform the physical environment into a digital entity that can be processed and fed back in real time [42]. In this process, AI supports a shift from traditional passive monitoring to data-driven active design [43]. Through the wide deployment of distributed intelligent sensor networks and Internet of Things (IoT) devices, space gains the ability to collect multidimensional real-time data. These data include visual information, temperature and humidity, pressure, and human movement trajectories [44].

This process of making space computable depends on efficient communication networks and edge processing algorithms [45]. For example, in smart living scenarios, AI can monitor indoor illuminance, CO₂ concentration, and occupants' physiological states in real time. Machine learning models can then adjust HVAC energy supply strategies and lighting levels dynamically [41]. This form of adaptation improves energy efficiency. It also aligns the indoor microenvironment with human behavior patterns [37]. The shift from automation to autonomy indicates that space is no longer only a carrier of daily life. It is evolving into an intelligent agent that can perceive conditions and optimize living experience on its own [36].

4.2.2. Spatial Cognitive Adaptation

Spatial cognitive adaptation marks the shift in smart buildings from basic information interconnection to advanced logical understanding. Artificial intelligence functions not only as a computational tool but also as a cognitive center that enables space to interpret complex environmental semantics [46]. By integrating multimodal sensors for vision, audio, motion, and biometrics, embodied intelligent systems can capture and analyze space use intensity, human dynamics, and subtle environmental changes in real time through machine learning and deep learning methods [47]. This deeper cognitive capacity allows space to build more accurate profiles of user behavior patterns and potential preferences [36].

Whether predicting user needs [41] or monitoring emotional changes and adjusting light color temperature and background music accordingly [43], these functions enable space to evolve actively [46].

4.2.3. Spatial Action Adaptation

Spatial action adaptation refers to the active, context-oriented response formed by space after completing perception and understanding [16,39,44,48]. Existing studies show that this type of response mainly takes four forms:

1. Space can regulate the microenvironment based on occupancy recognition and environmental prediction. Typical actions include automatic control of lighting, temperature, and shading. In peak-load or power outage scenarios, the system can also coordinate energy storage and load scheduling. This helps balance comfort and resilience [43,49,50].
2. Space can execute strategies through digital twins and autonomic management. This enables buildings to take corrective actions in fault diagnosis, performance optimization, and continuous recommissioning [17].
3. Space can respond to health, safety, and circulation needs. Examples include biometric access control, motion-triggered device linkage, and indoor mobility management [51].
4. Space can support reconfigurable actions related to spatial configuration. AI has been used for layout generation, robot-actuator coordination, and off-site manufacturing automation [37,42].

Overall, action adaptation marks the shift in space from a controllable object to an active agent capable of execution.

4.3. Technical Logic and Spatial Deployment Pathways of Embodied Intelligence

4.3.1. Spatial Deployment of Embodied Perception

Embodied perception requires the deployment of a multidimensional perception network. This network should capture both the physical environment and human states in a comprehensive way.

In terms of sensor deployment, existing studies emphasize dual coverage on the environment side and the wearable side. On the environment side, distributed sensor networks are used to collect thermal, air quality, and lighting parameters in real time [18]. This distributed deployment does not only focus on the macro environment. It also uses deep learning locally to compress and encrypt multidimensional time-series data [52]. At the same time, wearable and vision-based sensing technologies are used to capture biomechanical data through wearable devices integrated with MEMS modules. For example, the lightweight system designed by Wang et al. [19] achieved synchronized gait recognition across multiple body parts. In addition, eye-tracking technology has also been widely introduced [20,21].

The advancement of embodied perception lies in multimodal data fusion. For heterogeneous data, Wang et al. [19] proposed a CNN–Transformer–Attention framework. This framework enables high-accuracy recognition through cross-modal fusion. For streaming data, Najeh et al. [18] used a method based on dynamic segmentation and trace encoding. It transforms spatiotemporal trajectories into directed weighted networks as inputs for CNNs. This method effectively identifies overlapping activities. Lee et al. [53] proposed a new human–building interaction (HBI) system. The system integrates users' subjective feedback with indoor environmental sensor data. It supports a shift from passive data collection to active understanding.

In summary, the spatial deployment of embodied perception not only enables the collection of physical parameters. It also forms a bidirectional information flow between humans and the built environment.

4.3.2. Spatial Deployment of Cognitive Decision-Making

Cognitive decision-making needs to process large volumes of heterogeneous spatial perception data. It then generates accurate predictions and adaptive decisions.

At present, the data basis for cognitive decision-making shows a high level of cross-modal fusion. To overcome the cognitive limitations of single-modal data in complex spaces, Feng et al. [22] proposed the UrbanLLaVA framework. This framework integrates urban images, geographic text, and structured spatiotemporal sequences. It improves the global reasoning capability of multimodal large language models (MLLMs) for complex buildings and urban environments. For high-dimensional time-series data generated by building IoT sensors, conventional systems often fail to identify the root causes of anomalies. To address this problem, the InsightBuild model introduces a causal reasoning module based on large language models (LLMs). It shifts the analysis from simple data correlation to the automatic interpretation of causal logic behind system anomalies [23].

In the complex dynamics of human settlement, the decision-making paradigm is rapidly moving toward multimodal semantic understanding and spatial relational reasoning. Li et al. [24] proposed the SeeGround framework. It enables 2D VLMs to perform zero-shot 3D visual understanding and accurate target grounding without large-scale 3D-specific annotations. Kang et al. [25] introduced the task of 3D intention grounding. This task allows the system to locate targets in RGB-D scans based only on implicit human intentions expressed in natural language. In addition, Cheng et al. [26] showed that, after integrating depth information and region prompts, large models can accurately interpret complex spatial relations. These relations include absolute distance and relative orientation.

Together, these technical pathways show that new-generation decision-making systems can do more than identify space. They are beginning to perform human-like spatial reasoning and construct implicit cognitive maps. This enables deeper adaptation to complex human settlement scenarios.

4.3.3. Spatial Deployment of Embodied Actuation

The core of embodied actuation is to translate high-dimensional semantic decision instructions into physical actions in real space. These actions need to be low-latency, reliable, and accurate.

In the complex physical context of smart buildings, conventional full-cloud inference can no longer meet the high computational demand of embodied multimodal models (MLMs) and world models (WMs) for 3D spatial reasoning [54]. Therefore, deep neural network partitioning has become a core technique in current embodied actuation architectures. This technique dynamically divides complex computational graphs into multiple subtasks according to real-time computing resources. These subtasks are then deployed

in a collaborative manner [55]. This process can significantly reduce end-to-end inference latency [2].

At the same time, efficient collaboration among heterogeneous actuators is needed in distributed architectures. For this purpose, systems introduce communication protocols for edge-oriented embodied intelligence, such as Agent2Agent, and standardized task orchestration mechanisms [27]. Through lightweight metadata exchange and semantic alignment, these mechanisms support seamless atomic action scheduling and collective collaboration [56]. In terms of actuation safety, AI-driven dynamic resource allocation seeks an optimal balance between latency and power consumption. It also works with blockchain-based self-owned smart building frameworks. This helps define data privacy boundaries in actuation scenarios involving multiple agents [19,28].

In summary, the spatial deployment of embodied actuation forms a closed loop from digital instructions to the physical world through edge–cloud collaboration and distributed agent cooperation.

4.4. Architectural Framework for Adaptation and Implementation

Although the closed loop of embodied perception–cognitive decision-making–embodied actuation reveals the operational logic, it is not sufficient for guiding spatial construction. The framework must return to the social attributes and behavioral structure of space. In this way, technical logic can be translated into an operational mechanism that can be embedded into space (Figure 7).

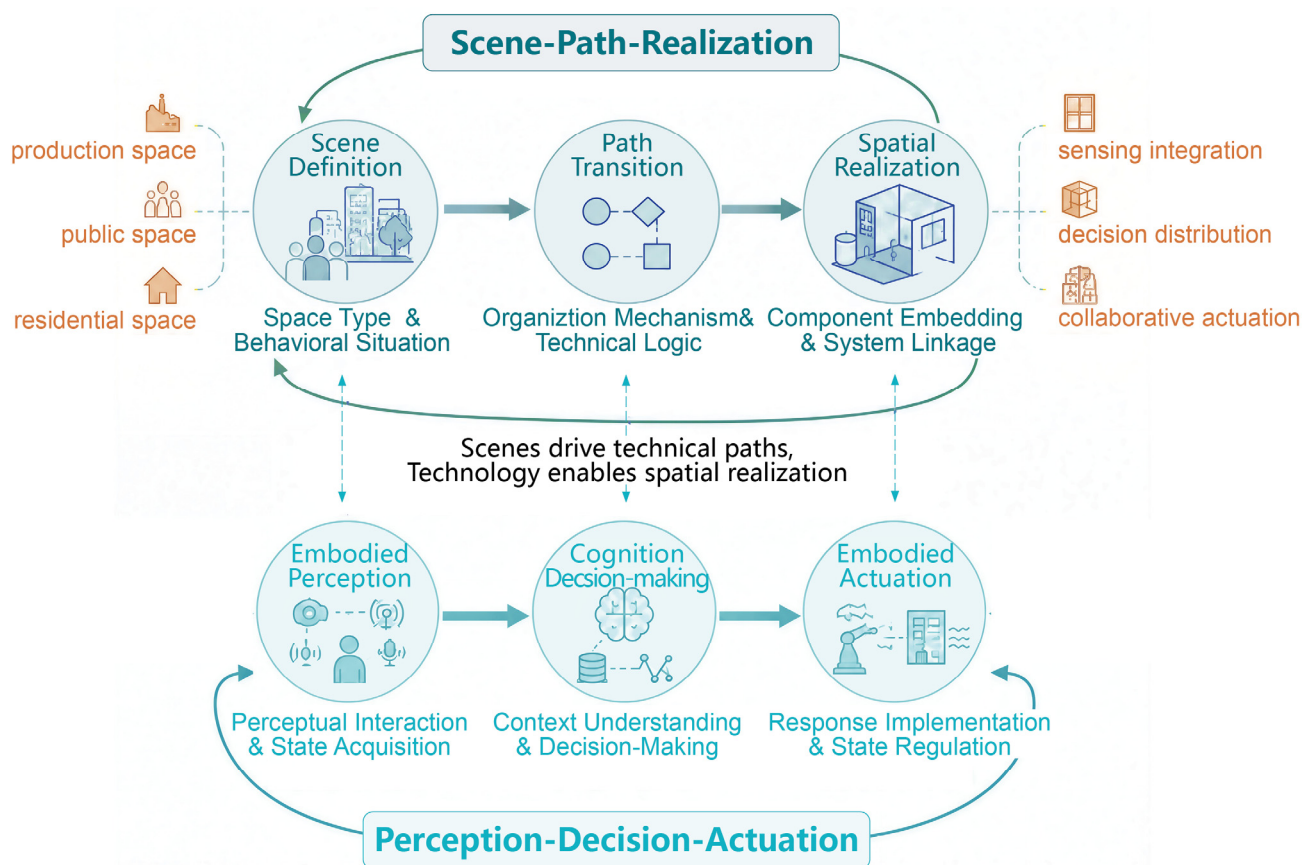


Figure 7. The two-tier paradigm framework of embodied intelligent space. The upper layer reflects the translation process from spatial context to system realization, while the lower layer reveals the technical operation mechanism. The two layers realize the structural embedding from technical logic to space through mapping relations.

4.4.1. Spatial Needs Generation

Differences in needs, and the spatial differences that result from them, are the basic premise for applying embodied intelligent technologies in architecture. Madanipour [57] examined how the concepts of “public” and “private” space are defined, constructed, and understood. Mitchell [58] discussed how digital technologies blur and reshape spatial divisions. Both studies show that different spatial types lead to different user needs, use conditions, and social relations. Accordingly, this paper classifies architectural space into three categories: production space, public space, and residential space.

The three types of space differ in their behavioral structures and interaction logics. Production space is task- and efficiency-oriented. It is characterized by highly coordinated process structures. Public space is shaped by open interaction and mobility. It is therefore dynamic and uncertain. Residential space is organized around individual life and it reflects stability and privacy [57]. These differences directly shape the response strategies of embodied intelligent systems.

4.4.2. Translation from Scenarios to Technical Pathways

Based on the spatial types, embodied intelligence needs to complete the translation from technology to space through a “scenario–pathway–implementation” framework. Here, scenario refers to a typical behavioral situation in a specific space. It is the concrete expression of the social attributes of space. Pathway refers to the organization of embodied perception, cognitive decision-making, and embodied actuation around this situation. It integrates perception and interaction, prediction and decision-making, and response and actuation into a task-oriented technical logic. Implementation refers to the physical expression of this logic within space.

Through this framework, the EPDA mechanism is transformed from an abstract process into a spatial operating structure. Embodied perception enables continuous acquisition of and interaction with human–environment states through distributed sensing and embedded interfaces. Cognitive decision-making supports situation understanding and state prediction through data-driven analysis and model reasoning. It then generates control strategies. Embodied actuation translates these decisions into specific responses through the coordination of spatial components and building systems. In this way, the technical mechanism can be structurally embedded into space.

4.4.3. Implementation Framework of Embodied Intelligent Space

At the perception level, information acquisition shifts from discrete devices to the integration of intelligent components, spatial interfaces, and building materials, allowing space itself to become a carrier of perception and interaction. At the decision-making level, the system moves from centralized control toward scenario-oriented, distributed decision-making across the entire space. Through data accumulation and model learning, it enables contextual interpretation and prediction. At the actuation level, spatial response expands from device control to coordinated changes in components and environmental conditions, transforming actuation into the adjustment of spatial states and forms.

5. Future Trends and Technologies of Embodied Intelligent Space

5.1. Trend 1: Reconstructing Human Settlement Pathways Toward Embodied Intelligent Space

The future development of embodied intelligent space will be deeply embedded in the social and economic logic of different spatial types. Production space, public space, and residential space will therefore generate distinct demand characteristics and technical priorities.

5.1.1. Production Space: High-Reliability Task Collaboration and Safety Response Logic

In industrial manufacturing and complex office scenarios, the future development of embodied intelligent space will focus on high-reliability task scheduling, human–machine collaboration, and dynamic safety protection. Under this trend, production space will be integrated into an intelligent manufacturing architecture based on perception, cognition, actuation, and feedback. This architecture enables semantic fusion and collaborative reasoning among heterogeneous production factors [5]. At the application and actuation level, spatial agents need to manage highly complex production scheduling. One example is multi-robot task allocation (MRTA) under limited buffer capacity and blocking conditions [59]. They also need to continuously optimize workflows and balance spatial layout with human–machine interaction [60]. In addition, spatial agents will use AI for situational awareness. They can adaptively adjust interior design patterns and microenvironmental facilities. This can improve production efficiency while supporting users' physical and mental comfort [61] (Figures 8 and 9).

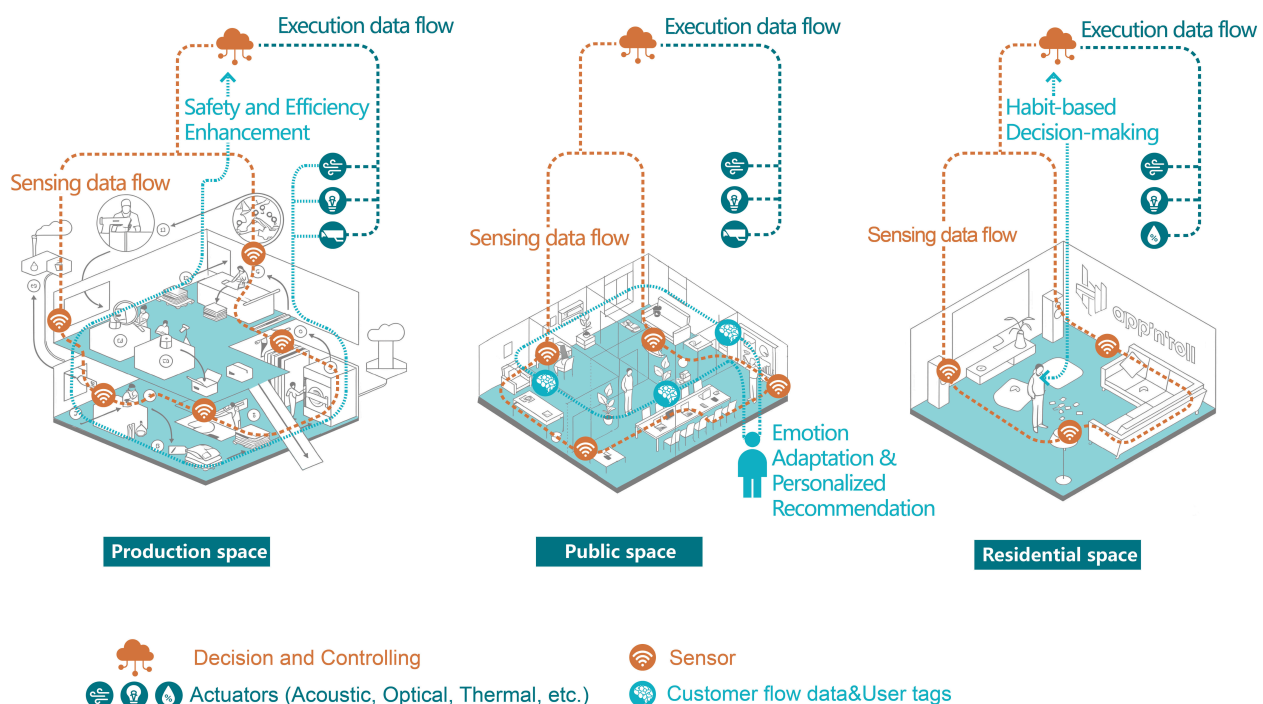


Figure 8. Application Scenarios of Embodied AI in Production, Public, and Residential Spaces. Dashed lines represent the information transmission paths in space, and arrows indicate the interactions between users and the cognitive decision-making system.

5.1.2. Public Space: Experience Enhancement, Flow Guidance, and Commercial Value Creation

The demand for embodied intelligence in public spaces varies by building type. In large commercial facilities, embodied intelligence can analyze large-scale pedestrian flow data and multidimensional user profiles. It can then provide scenario-based and targeted intelligent recommendations. This can improve visitor conversion rates [62,63]. Immersive experience design based on multi-sensory interaction and embodied cognition is reshaping exhibition buildings [64,65]. In high-flow nodes such as airports, AI systems can combine facial recognition with behavior prediction to improve spatial operation efficiency [66]. In addition, social service robots will play important roles in navigation and interaction in public spaces [67]. At a larger scale, the introduction of geospatial intelligent interaction design can support the extraction and optimization of public building spatial structures.

This allows public buildings to better meet the needs of smart city development [68] (Figures 8 and 9).

Information Guidance	AR Equipment Repairing Guide		Indoor Navigation		Medication Reminders							
	AGV Material Handling Guide		Smart Guide Facilities		Home Environmental Info Panel							
	Digital Work Instruction		Dynamic Digital Signage		Children's Storytelling Projector							
	Personnel and Item Positioning		Interactive Inquiry Kiosks		AR Home Decoration Guidance System							
	others		others		others							
cognitive support & emotional care	Operation Error Warning		Smart Virtual Inquiry Assistant		Smart Voice Assistant							
	Fatigue Monitoring and Intervention		Mood-Regulating Lighting		Cognitive Training Game Platform							
	Virtual Mental Wellness Pod		Emergency Soothing Broadcast		Smart Photo Album							
	AR Support for High-Risk Tasks		Lost Child Assistance System		Emotional Companionship							
	others		others		others							
Unobtrusive Monitoring	Machinery Vibration & Acoustic Analysis		Crowd Heat Mapping		Fall Detection Radar							
	Chemical Leak Detection		Washroom Environment Sensing		Water Leak and Gas Sensors							
	High-Temp Thermal Imaging		Noise Monitoring and Controlling		Sleep Quality Monitoring							
	Dust Concentration Sensing		Special Zone Intrusion Monitoring		Gait Monitoring							
	others		others		others							
Wearable Assistive Devices	Industrial Exoskeleton		Intelligent Visual Aid Glasses		Smart Wristband							
	Anti-Fatigue Monitoring Helmet		Vital Signs Monitoring Vest		Smart Hearing Aid							
	Smart Safety Wristband		Real-time Translation Earbuds		Intelligent Sleep Eye Mask							
	Ring Scanner		AR Experience Glasses		Exoskeleton Robot							
	others		others		others							
Robotics	Collaborative Robot		Delivery Robots		Smart Refrigerator							
	Inspection and Cleaning Robot		Smart Surveillance Cameras		Robot Vacuum							
	Autonomous Mobile Robot		Cleaning and Disinfection Robots		Smart Surveillance Cameras							
	Smart Surveillance Cameras		Inventory and Sorting Robots		Companion Robot							
	others		others		others							
	Factory	Plant	Workshop	Mill	Education	Transportation	Commercial	Culture	House	Apartment	Villa	Dormitory
	Foundry	Refinery	Warehouse	others	Stadium	Healthcare	Office	Hotel	others	Condominium	Nursing Home	others
	Production Space				Public Space				Residential Space			

Figure 9. This figure lists the major technologies or devices of embodied intelligence applied in production spaces, public spaces and residential spaces, covering five dimensions: information guidance technology, cognitive support & emotional care, unobtrusive monitoring, wearable assistive devices, and robotics.

5.1.3. Residential Space: Unobtrusive Protection and Everyday Life Assistance

In future residential spaces, embodied intelligence will support unobtrusive protection and everyday life assistance. In energy management and microenvironment regulation, Edge AI can perform anomaly detection and predictive temperature control on local devices with sub-50 ms latency [69]. This supports imperceptible indoor temperature adjustment. Deep learning (DL) is also widely used to monitor and optimize energy use and distribution in smart homes [70]. It can greatly reduce the attention required from occupants for active environmental control. In home health care, embodied intelligent space can continuously and unobtrusively collect vital-sign data. It can also generate anonymized synthetic health profiles to address the shortage of rare pathological data [71]. In emergency situations, it can improve response speed by 25% [71] (Figures 8 and 9).

5.2. Trend 2: Spatial Cognitive Reconstruction Based on World Models

The improvement of embodied intelligent space depends on the system's deep understanding and prediction of physical laws. World models act as internal simulators for embodied agents. They capture the dynamic features of the environment and support both forward and counterfactual reasoning [24].

5.2.1. Scenario Rehearsal: World Model-Based Physical Simulation and Actuation Preview

An embodied intelligent space can use a global latent vector or a spatial latent grid built by world models to obtain a high-fidelity virtual testbed [24]. In complex architectural environments, the system can use world models to simulate and rehearse different actuation plans before physical action [4]. It can also predict environmental states after multiple action steps. This reduces the cost of trial and error in physical actuation. It also improves the operational certainty of embodied agents in unstructured spaces [72].

5.2.2. Multidimensional Scheduling: Building Environmental Regulation, Facility Assignment, and Coordination with Social Systems

The effectiveness of embodied actuation depends not only on the accuracy of single actions. It also depends on the coordinated scheduling of multidimensional spatial elements. Future embodied intelligent spaces will introduce action attention masking to reduce error accumulation in long-sequence action execution [73]. In complex scheduling tasks, the system can use trajectory representations generated by diffusion models. These representations support the fine-grained regulation of environmental media, such as temperature, humidity, and airflow. They also support the dynamic assignment of intelligent facilities and coordination with social systems [74]. This multidimensional scheduling mode can greatly improve the capacity of space to support complex and dynamic tasks [75,76].

5.2.3. Predictive Response: Architectural Spatial Intervention Mechanism Based on World Models

The causal reasoning capability of world models is the key driver of anticipatory intervention in embodied intelligent space. Through self-supervised learning, world models can capture the latent evolution logic of spatial environments. They can identify the deeper causes of environmental change and predict possible abnormal states [24,77]. In residential spaces and urban public spaces, this mechanism supports early responses to energy loss, facility faults, and safety risks [78]. This predictive actuation logic based on world models improves system resilience under extreme conditions. It also improves the safety and resource efficiency of urban spatial operation at the underlying logic level [79,80] (Figure 10).

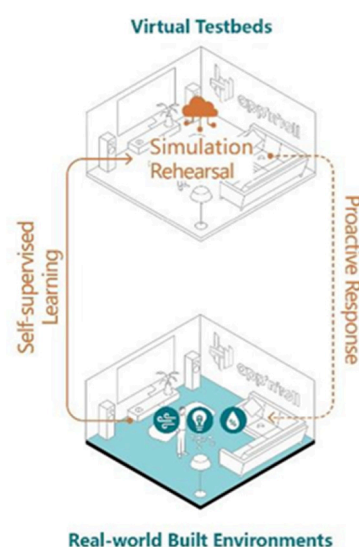


Figure 10. Based on the world model, virtual testbeds learn information from the real-world built environment in a self-supervised manner, preview possible outcomes so as to identify the deeper causes of environmental change and predict potential abnormal states, and then respond proactively.

5.3. Trend 3: Construction and Improvement of Spatial Continual Learning Mechanisms

Embodied intelligent space is gradually shifting from a static automation system to a dynamic system with continual optimization capacity. This shift is supported by edge–cloud collaboration (ECC). It changes the spatial response mode from predefined rules to experience-driven adaptation [24,81]. It also enables sustainable technical iteration in dynamic urban environments [82,83].

5.3.1. Real-Time Optimization at the Edge: Low-Latency Spatial Response and Privacy Protection

In spatial operation, tasks related to safety warnings and privacy protection need to be completed at the edge. The edge node clustering algorithm (ENCA), based on communication range, divides nearby devices into logical clusters. This enables dynamic balancing of local resources [84]. Together with the scheduling algorithm based on maximum matching (SAMM), latency-sensitive tasks, such as fall detection and fire warning, can be preferentially assigned to low-load nodes. This can increase the task completion rate from 32% to 47.2% [84]. In addition, small language models (SLMs) deployed at the edge can perform intention reasoning while keeping data local. This supports both privacy protection and sub-second response [85]. These practices show that real-time optimization at the edge provides the basic capacity for rapid spatial response [82,86].

5.3.2. Long-Term Cloud Iteration: Deep Mining of Spatial Operation Data and Strategy Update

If the edge is responsible for immediate response, the cloud supports long-term data storage, complex reasoning, model optimization, and deep data mining [81,85]. Through the cloud training–edge inference paradigm, the cloud uses aggregated and desensitized data for instruction tuning. General capabilities are then transferred to the edge through knowledge distillation and parameter-efficient fine-tuning (PEFT) [85]. For example, the cloud can analyze energy consumption patterns across building clusters and generate global strategies. It can also reuse validated interaction logic through CrowdTransfer, reducing the cost of repeated learning [83,87]. This mechanism allows the intelligence level of space to improve gradually as data accumulate [81] (Figure 11).

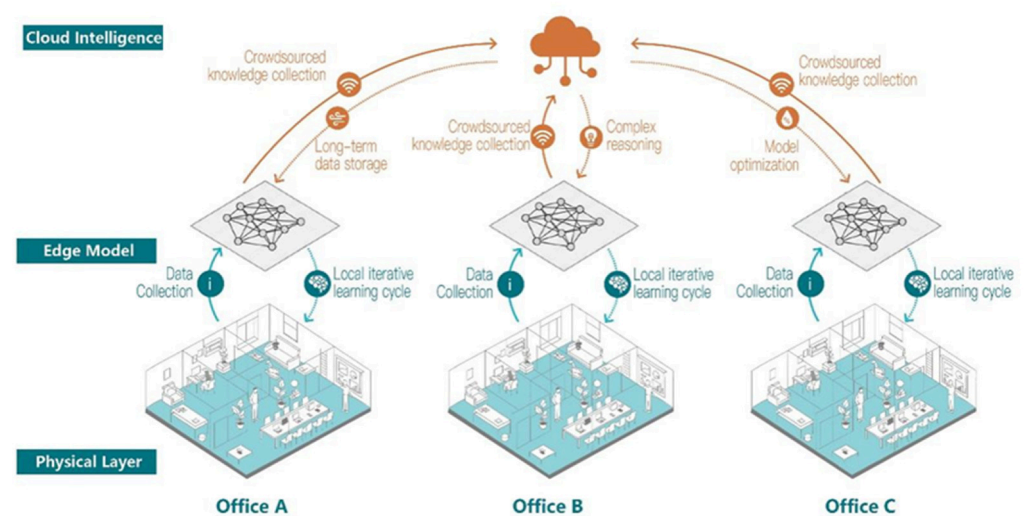


Figure 11. Distributed learning and continuous evolution through edge–cloud collaboration.

5.3.3. Adaptive Evolution: Spatial Response Mode from Predefined Rules to Experience-Driven Adaptation

When facing new situations, space can use retrieval-augmented generation (RAG) to dynamically call cloud knowledge and edge records. It then generates adaptive strategies instead of mechanically executing predefined scripts. This supports a shift from reactive response to predictive response [86]. In addition, swarm intelligence can use a collective knowledge transfer framework. Successful strategies from a single space, such as responses to extreme climate, can be quickly extended to city-level clusters. This enables the derivation, sharing, and integration of knowledge [87]. This shift from individual adaptation to collective coordination helps address complex problems such as regional energy scheduling and traffic coordination [83,88]. In summary, edge–cloud collaboration drives space toward a system with growth capacity. Its intelligence level can continue to improve as application scenarios expand. This supports the resilient development of human settlement environments [81,82].

5.4. Trend 4: Expansion of Spatial Organization Forms for Multi-Agent Systems

Feng et al. [89] examined the cross-scale characteristics of spatial intelligence in navigation, urban planning, remote sensing, and geoscience. Their study shows the potential of embodied intelligent space to move beyond the boundary of a single building. It can extend toward multi-agent cluster collaboration at community and urban scales. In recent years, large language model (LLM)-based multi-agent systems (MAS) have provided a new paradigm for building resilient urban networks [15,90]. In the future, embodied intelligent spaces will act as urban nodes. Through collaboration, they can support cross-space resource scheduling and functional complementarity. This will promote the transformation of smart cities from single-point automation to collective intelligence [91,92].

5.4.1. Distributed Collaborative Decision-Making: Overcoming the Local Optimum of Individual Intelligence

Decision-making by a single agent based on local data can easily fall into a local optimum. Multi-agent collaboration uses a distributed architecture. It allows each agent to remain autonomous while reaching global consensus [93].

For city-scale complex tasks, LLM-driven multi-agent systems show clear advantages. With routing and retrieval-augmented generation (RAG), these systems can reach 94–99% accuracy in heterogeneous data processing. Their response quality is also higher than that of independent models [15]. This framework reduces the risk of single-point failure. It also improves system adaptability to sudden disturbances. For data security, privacy-preserving AI system architectures allow agents to conduct collaborative reasoning without sharing sensitive raw data. This helps address the problem of data silos [94]. These advances show that distributed collaborative decision-making is a key pathway for addressing the complexity of urban systems and achieving global optimization [95].

5.4.2. Dynamic Role Assignment and Task Orchestration: Adapting to Complex and Changing Urban Scenarios

The highly dynamic nature of urban environments requires flexible task allocation mechanisms. Multi-agent systems support flexible resource configuration through dynamic role assignment [90].

The collaboration among heterogeneous agents is critical for handling complex tasks. Agents with different functions can interoperate through standardized semantic models and form temporary task coalitions [29,94]. In smart building operation, LLM-driven virtual assistants can understand natural language instructions (Figure 12). They can analyze real-time occupancy patterns, adjust environmental conditions, and support cross-zone

energy scheduling [29,96]. This mechanism allows agent clusters to flexibly reorganize their structure according to environmental changes. For example, in emergency evacuation scenarios, building agents can serve as information relay nodes and jointly plan optimal routes [15]. In mobile-assisted learning scenarios, AI agents can also adjust service content according to user location and behavior. This reflects a high level of adaptive context awareness [97].

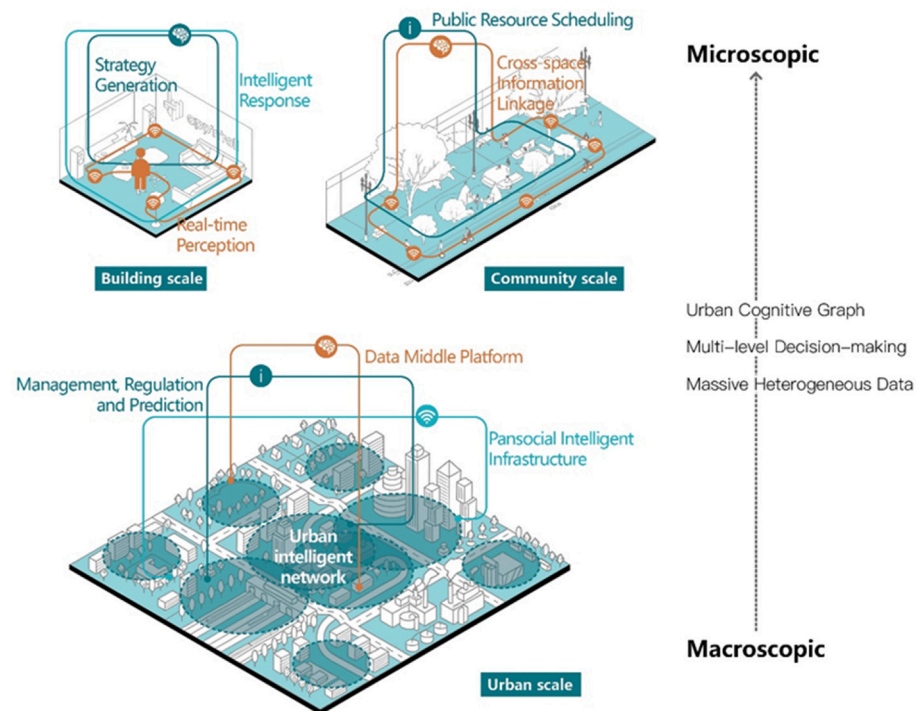


Figure 12. Collective intelligence enables embodied intelligent spaces to process more complex information flows and conduct more diverse interactions, thus realizing the leap of macro governance from individual buildings to communities and further to the entire “AI city”.

5.4.3. Emergence of Collective Intelligence: From Micro-Level Interaction to Macro-Level Urban Governance

The high-level form of multi-agent collaboration is the emergence of collective intelligence. It means that ordered behavior appears at the macro level through local interaction rules. This allows the system to efficiently process massive heterogeneous sensor data at the urban scale [91].

Collective intelligence provides a new approach to urban governance. Through a cross-scale spatial intelligence framework, micro-level interaction data can be aggregated into macro-level urban cognitive maps. These maps can support multi-level decision-making [89]. In blue-green infrastructure planning, AI-driven collective analysis can identify seasonal change patterns. It provides a precise tool for ecosystem service assessment [98]. This bottom-up governance mode reduces administrative costs and strengthens the adaptability of urban systems. With technological development, the vision of an “AI city” that integrates big data and collective intelligence is gradually becoming practical. This marks an inevitable trend in which embodied intelligent space evolves toward an urban ecosystem [15,92] (Figure 12).

6. Discussion

6.1. Theoretical Significance of Constructing Embodied Intelligent Space

First, this paper redefines the meaning of spatial intelligence from an architectural perspective. It shifts the technical framework of embodied intelligence away from the conventional focus on algorithms and devices toward the coordinated operation of the Human–Intelligence–Context system. Existing studies on smart buildings have mainly focused on automated control, energy optimization, and device interconnection. In contrast, this paper emphasizes space itself as an active agent with capacities for perception, cognition, and actuation, thereby transforming architecture from a technological container into an embodied intelligent agent. This shift not only extends the discussion on how digital technologies reconstruct spatial boundaries [58], but also responds to the need for adaptation arising from differences in the social attributes and behavioral structures of space [57].

Second, this paper proposes an analytical framework of embodied perception–cognitive decision-making–embodied actuation (EPDA). It further develops a two-level scenario–pathway–implementation paradigm to translate the technical logic of artificial intelligence into the operational logic of architectural space. Existing studies often remain at the level of single technical modules, such as multimodal perception, edge computing, or digital twins. In contrast, this paper links perception networks, cognitive reasoning, and actuation coordination with spatial types, behavioral structures, and social relations in an integrated manner. This allows technical mechanisms to be structurally embedded into the architectural system. The resulting integration between technical systems and spatial organization provides an architectural pathway for implementing embodied intelligence in space.

Finally, this paper examines the development of embodied intelligent space across a continuous scale from individual buildings to urban collective intelligence. It further incorporates a systematic discussion of implementation pathways for future technologies in application scenarios, thereby extending the traditional paradigm of single-point intelligence in smart city research. By introducing world models, multi-agent collaboration, and edge–cloud continual learning mechanisms, space is no longer an isolated intelligent unit. Instead, it becomes an urban-scale network node capable of knowledge transfer, collaborative reasoning, and dynamic evolution [15,89]. This framework not only strengthens the interdisciplinary connections among architecture, artificial intelligence, and urban systems, but also provides a new theoretical foundation for future research on AI cities and resilient human settlement environments.

6.2. Potentials for Practical Applications: A New Architectural Design Agenda

Based on the theoretical framework and systematic review findings established in this study, the construction of embodied intelligent space extends beyond theoretical innovation to offer actionable potentials for practical architectural applications. The transition from static physical containers to active spatial agents fundamentally reshapes the architectural design agenda, shifting the design focus from mere spatial composition to long-term spatial operation. Specifically, the review findings unlock practical application potentials across three key architectural dimensions.

6.2.1. Human–Space Symbiosis: From Passive Control to Proactive Behavioral Adaptation

In practical architectural design, the integration of multimodal perception and spatial cognitive reasoning transforms human–building interaction from reactive, command-driven control into continuous, unobtrusive behavioral symbiosis. Practically, architects can leverage spatial readability, material semantics, ambient lighting, and acoustic cues

to design environments that proactively adapt to human intent. Rather than relying on intrusive physical interfaces or rigid rule-based triggers, embodied intelligent spaces can progressively improve their ability to interpret occupancy patterns, behavioral states, and contextual information to regulate indoor environmental conditions, optimize lighting, and enhance human comfort and well-being in real-time. From an evaluation perspective, this capability can be assessed through indicators such as perception accuracy, response latency, occupant satisfaction, and adaptability under changing behavioral conditions.

6.2.2. Architectural-Equipment Integration: From Add-On Terminals to Embedded Spatial Operating Systems

The review findings demonstrate that actuators, robotic systems, dynamic building components, and smart terminals should no longer be treated as isolated post-occupancy add-ons. Instead, practical applications mandate their integration into an embedded spatial operating system from the schematic design stage. Architectural design must proactively incorporate spatial pathways for autonomous mobile agents, standardized data/power interfaces, dynamic structural components, and reconfigurable partition units. This structural embedding enables building spaces to flexibly reconfigure their spatial layouts, environmental services, and functional zones in response to shifting operational demands—such as multi-robot task allocation in industrial facilities or dynamic crowd guidance in public transit hubs. The maturity of this integration can be evaluated through criteria including interoperability between spatial components and intelligent systems, response reliability, system scalability, and maintenance requirements throughout the building lifecycle.

6.2.3. Virtual–Physical Twin Synergy: From Static Completion to Continual Operational Evolution

By leveraging digital twin technologies and spatial world models, embodied intelligent spaces gain mirror cognitive capabilities [99], enabling building design to evolve from a static post-construction state into a dynamic, continuously learning operational process. In practical implementation, this synergy allows for high-fidelity behavior–environment predictive simulation [100] during the design phase, real-time closed-loop feedback during operations, and long-term strategy optimization via edge–cloud continual learning. Buildings can gradually accumulate operational knowledge through continuous data acquisition and model updating, allowing spatial agents to dynamically update energy scheduling, maintenance protocols, and safety management strategies based on accumulated empirical data over their entire lifecycle [101]. Potential measurable criteria include prediction accuracy, energy performance improvement, fault detection efficiency, and the effectiveness of long-term operational optimization.

6.3. Challenges in Implementing Embodied Intelligence in Real Architectural Environments

However, the current implementation of embodied intelligent space should not be interpreted as a fully autonomous architectural paradigm. Instead, existing applications remain constrained by technological maturity, data availability, system reliability, and regulatory considerations. While embodied intelligence opens up new possibilities for autonomous spatial operation, its application in real architectural environments remains constrained by multiple practical conditions.

In complex spatial environments, the robustness of behavior perception remains a major challenge. Large buildings are often affected by signal attenuation, multipath effects, and non-line-of-sight (NLOS) interference, all of which can reduce the accuracy of crowd behavior recognition [102]. Crowd behavior is also highly dynamic and context-dependent. Trajectory recognition alone is therefore insufficient for understanding emotions, inten-

tions, and group-level semantics. Without an effective correspondence between physical behavioral data and spatial semantics, spatial responses may become biased and may even increase the risk of congestion in specific scenarios [103].

Long-term autonomous learning in embodied intelligent space also raises issues of controllability. Spatial systems need to keep learning and adjusting over time and across changing scenarios. However, catastrophic forgetting in multimodal continual learning may cause a system to lose established architectural control logic while adapting to new demands [104]. This cognitive discontinuity may weaken the stable execution of basic energy-saving strategies, safety management, and disaster-avoidance mechanisms. It may also lead the system to overreact to local needs in complex social settings while neglecting the overall order of spatial operation. In addition, spatiotemporal coupling systems with multiple failure modes further increase the uncertainty of building safety assessment [105].

Privacy and ethical concerns form another bottleneck in the wider adoption of embodied intelligent space. High-precision behavior analysis often depends on continuous video and sensor data collection. Large-scale sensing systems may therefore cross the boundary of individual anonymity in less visible ways [106]. At the same time, the black-box nature of AI decision-making creates tension with the interpretability and perceptibility expected in architectural design [107]. When space is able to autonomously regulate environmental conditions or even change its physical layout, decisions that users cannot understand may weaken their sense of trust and safety.

The future implementation of embodied intelligent space therefore requires not only advances in artificial intelligence technologies but also systematic evaluation frameworks, long-term validation in real architectural environments, and appropriate governance mechanisms that balance autonomous operation with human control. These considerations are essential for transforming embodied intelligent space from a conceptual framework into a reliable and socially acceptable architectural practice.

7. Conclusions

As artificial intelligence shifts from disembodied computing to embodied intelligence, the intelligent development of human settlements is reaching a critical turning point. Based on a systematic review of the literature from 2017 to 2026, this study addresses the current challenges of “advanced technology but limited cognition” and the disconnection between perception and actuation in smart buildings. The core findings of this study directly address the three research questions.

In response to RQ1 regarding conceptual foundations and agent characteristics, embodied intelligent space establishes a new paradigm for future human settlements by driving a spatial evolution from a passive container into a proactive agent. Grounded in the Human–Intelligence–Context (HIC) framework, this paradigm marks a fundamental shift from external technology embedding to endogenous intelligence. Under this framework, space is reconstructed as a collaborative intelligent agent endowed with autonomy, allowing it to perceive, respond, and adapt to human activities continuously. Much like a living organism, space engages in a continuous interaction with occupants, thereby fostering a long-term symbiotic relationship between humans and their built environment.

In response to RQ2 concerning system architecture and collaborative mechanisms, the operation of spatial embodied cognition is sustained by an integrated closed-loop triad of embodied perception, cognitive decision-making, and embodied actuation. This technical closed loop establishes three fundamental spatial capabilities. First, multimodal perception coupled with edge computing endows the environment with computability, enabling real-time quantification and processing of dynamic spatial data. Second, spatial foundation models and world models provide interpretability, serving as a cognitive core

that allows space to reason within its physical context. Third, edge–cloud collaboration and multi-agent coordination deliver actionability, allowing space to execute proactive micro-environmental regulation and physical–spatial reconfigurations.

In response to RQ3 regarding future design trends, the technical closed loop serves as the catalyst for a new architectural agenda. By structurally embedding embodied intelligence into architectural logic through a scenario–pathway–implementation framework, the boundary of architectural design is significantly expanded. The deep integration of architectural approaches and embodied intelligence technologies is set to drive three definitive design trends: a paradigm shift toward dynamic human–space interaction design, seamless space–equipment physical–digital collaboration, and comprehensive lifecycle virtual–physical twin design.

The main contributions of this review can be summarized as follows. First, this study clarifies the conceptual characteristics of embodied intelligent space and its distinction from existing intelligent building approaches. Second, it synthesizes key technological components, including perception, cognition, and actuation, which support autonomous spatial intelligence. Third, it discusses future opportunities for applying embodied intelligent space in human-centered built environments.

However, several limitations should be acknowledged. As embodied intelligent space is still an emerging concept, most related technologies remain at experimental or early implementation stages, and comprehensive evaluation methods are still under development. Future research should focus on empirical validation in real-world environments, quantitative evaluation indicators for spatial intelligence, and the ethical and operational challenges associated with autonomous decision-making.

Author Contributions: Conceptualization, X.Z.; methodology, X.Z.; software, X.Y.; validation, X.Z., X.Y. and J.S.; formal analysis, X.Z. and Y.L.; investigation, X.Z., X.Y. and J.S.; resources, Y.F.; data curation, X.Y.; writing—original draft preparation, X.Z.; writing—review and editing, J.S., Y.F. and Y.L.; visualization, X.Z.; supervision, J.L.; project administration, J.L.; funding acquisition, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors acknowledge the support of the Center for Balance Architecture, Zhejiang University.

Conflicts of Interest: Authors Xin Zhou and Yuping Feng were employed by the company The Architectural Design and Research Institute of Zhejiang University Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EPDA	Embodied Perception–Cognitive Decision-making–Embodied Actuation
CNNs	Convolutional Neural Networks
HBI	Human–Building Interaction
LLM	Large Language Model
VLM	Vision–Language Model
HVAC	Heating, Ventilation and Air Conditioning
RAG	Retrieval-augmented Generation
SLR	Systematic Literature Review
TF-IDF	Term Frequency–Inverse Document Frequency

MLLM	Multimodal Large Language Model
MLM	Multimodal Language Model
WM	World Model
MRTA	Multi-Robot Task Allocation
DL	Deep Learning
ECC	Edge–Cloud Collaboration
ENCA	Edge Node Clustering Algorithm
SLM	Small Language Model
PEFT	Parameter-Efficient Fine-Tuning
MAS	Multi-Agent Systems
NLOS	Non-Line-Of-Sight

References

1. Rao, S.; Good, J. What do we design for when we design “smart buildings”?—A scoping review of human experience design research in buildings. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 26 April–1 May 2025. [\[CrossRef\]](#)
2. Liu, H.; Guo, D.; Cangelosi, A. Embodied intelligence: A synergy of morphology, action, perception and learning. *ACM Comput. Surv.* **2025**, *57*, 186. [\[CrossRef\]](#)
3. Zhang, H.; Sediq, A.B.; Afana, A.; Erol-Kantarci, M. Generative AI-in-the-loop: Integrating LLMs and GPTs into the next generation networks. *arXiv* **2024**, arXiv:2406.04276. [\[CrossRef\]](#)
4. Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. A survey on large language model based autonomous agents. *Front. Comput. Sci.* **2024**, *18*, 186345. [\[CrossRef\]](#)
5. Tan, J.; Shi, J.; Wu, L.; Chen, B.; Tang, H.; Zhang, C.; Zhang, W.; Wang, S.; Wan, J. Embodied intelligence empowering customized manufacturing: Architecture, opportunities, and challenges. *IEEE Access* **2025**, *13*, 92740–92755. [\[CrossRef\]](#)
6. Amangeldy, B.; Imankulov, T.; Tasmurzayev, N.; Dikhanbayeva, G.; Nurakhov, Y. A review of artificial intelligence and deep learning approaches for resource management in smart buildings. *Buildings* **2025**, *15*, 2631. [\[CrossRef\]](#)
7. Wiener, N. *Cybernetics or Control and Communication in the Animal and the Machine*; The MIT Press: Cambridge, MA, USA, 2019. [\[CrossRef\]](#)
8. Norman, D.A. Design Principles for Cognitive Artifacts. *Res. Eng. Des.* **1992**, *4*, 43–50. [\[CrossRef\]](#)
9. Long, S.; Dhillon, B. (Eds.) Man-machine-environment System Engineering. In *Proceedings of the 16th International Conference on MMESE*; Springer: Berlin/Heidelberg, Germany, 2016. [\[CrossRef\]](#)
10. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; The University of Illinois Press: Urbana, IL, USA, 1949; pp. 1–117.
11. Kubota, M. What is “Communication”?—Beyond the Shannon & Weaver’s Model. *Int. J. Educ. Media Technol.* **2019**, *13*, 54–65.
12. Chen, F.; Wang, Z. Intelligent system architecture based on system theory. *Chin. J. Inf. Fusion* **2025**, *2*, 1–13. [\[CrossRef\]](#)
13. Asada, M.; Cangelosi, A. *Developmental Robotics: From Babies to Robots*, 2nd ed.; Cangelosi, A., Schlesinger, M., Eds.; MIT Press: Cambridge, MA, USA, 2021. [\[CrossRef\]](#)
14. Ferrari, A.; Micucci, D.; Mobilio, M.; Napoletano, P. Deep learning and model personalization in sensor-based human activity recognition. *J. Reliab. Intell. Environ.* **2022**, *9*, 27–39. [\[CrossRef\]](#)
15. Kalyuzhnaya, A.; Mityagin, S.; Lutsenko, E.; Getmanov, A.; Aksenkin, Y.; Fatkhiev, K.; Fedorin, K.; Nikitin, N.O.; Chichkova, N.; Vorona, V.; et al. LLM agents for smart city management: Enhancing decision support through multi-agent AI systems. *Smart Cities* **2025**, *8*, 19. [\[CrossRef\]](#)
16. Cao, Z.; Wang, Z.; Xie, S.; Liu, A.; Fan, L. Smart Help: Strategic opponent modeling for proactive and adaptive robot assistance in households. *arXiv* **2024**, arXiv:2404.09001. [\[CrossRef\]](#)
17. Genkin, M.; McArthur, J.J. B-SMART: Building systems management autonomic reference template for smart buildings. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106063. [\[CrossRef\]](#)
18. Najeh, H.; Lohr, C.; Leduc, B. Convolutional neural network bootstrapped by dynamic segmentation and stigmergy-based encoding for real-time human activity recognition in smart homes. *Sensors* **2023**, *23*, 1969. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Wang, J.; Liu, N.; Xie, Y.; Que, S.; Xia, M. A multimodal CNN–Transformer network for gait pattern recognition with wearable sensors in weak GNSS scenarios. *Electronics* **2025**, *14*, 1537. [\[CrossRef\]](#)
20. Han, J.; Mo, Y. A framework for investigating smart building and occupant behavior through eye-tracking technology. In *Computing in Civil Engineering*; American Society of Civil Engineers: New York, NY, USA, 2024; Volume 2024, pp. 1038–1046. [\[CrossRef\]](#)

21. Ding, W.; Li, F.; Ji, Z.; Xue, Z.; Liu, J. ATOM-Bot: Embodied fulfillment of unspoken human needs with affective theory of mind. *arXiv* **2024**, arXiv:2406.08455. [[CrossRef](#)]
22. Feng, J.; Wang, S.; Liu, T.; Xi, Y.; Li, Y. UrbanLLaVA: A multimodal large language model for urban intelligence with spatial reasoning and understanding. *arXiv* **2025**, arXiv:2506.23219. [[CrossRef](#)]
23. Neogi, P.P.G.; Mohammadshirazi, A.; Ramnath, R. InsightBuild: LLM-powered causal reasoning in smart building systems. *arXiv* **2025**, arXiv:2507.08235. [[CrossRef](#)]
24. Li, R.; Li, S.; Kong, L.; Yang, X.; Liang, J. SeeGround: See and ground for zero-shot open-vocabulary 3D visual grounding. *arXiv* **2025**, arXiv:2412.04383. [[CrossRef](#)]
25. Kang, W.; Qu, M.; Kini, J.; Wei, Y.; Shah, M.; Yan, Y. INTENT3D: 3D object detection in RGB-D scans based on human intention. In Proceedings of the International Conference on Learning Representations (ICLR), Singapore, 24–28 April 2025. [[CrossRef](#)]
26. Cheng, A.-C.; Yin, H.; Fu, Y.; Guo, Q.; Yang, R.; Kautz, J.; Wang, X.; Liu, S. SpatialRGPT: Grounded spatial reasoning in vision-language models. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 10–15 December 2024. [[CrossRef](#)]
27. Duan, Q.; Lu, Z. Agent communications toward agentic AI at edge: A case study of the Agent2Agent protocol. *arXiv* **2025**, arXiv:2508.15819. [[CrossRef](#)]
28. Dumitru, M.-C.; Caramihai, S.-I.; Dumitrascu, A.; Pietraru, R.-N.; Moiescu, M.-A. AI-enabled dynamic edge-cloud resource allocation for smart cities and smart buildings. *Sensors* **2025**, *25*, 7438. [[CrossRef](#)] [[PubMed](#)]
29. Ly, R.; Shojaei, A.; Gao, X. Smart building operations and virtual assistants using LLM. In *Companion Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*; ACM: San Francisco, CA, USA, 2025; pp. 1683–1689. [[CrossRef](#)]
30. Zhang, Z.; Kang, Z.; Zhao, R.; Feng, Y.; Jiang, B.; Ji, H.; Liu, L. ProAct: A dual-system framework for proactive embodied social agents. *arXiv* **2026**, arXiv:2602.14048. [[CrossRef](#)]
31. Zhang, W.; Zhou, Z.; Zeng, X.; Liu, X.; Fang, J.; Gao, C.; Cui, J.; Li, Y.; Chen, X.; Zhang, X.-P. Open3D-VQA: A benchmark for embodied spatial concept reasoning with multimodal large language model in open space. In Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), Dublin, Ireland, 27–31 October 2025; pp. 12784–12791. [[CrossRef](#)]
32. Yuan, Y.; Ding, J.; Feng, J.; Jin, D.; Li, Y. UniST: A prompt-empowered universal model for urban spatio-temporal prediction. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), Barcelona, Spain, 25–29 August 2024; pp. 4095–4106. [[CrossRef](#)]
33. Aminiranjbar, Z.; Tang, J.; Wang, Q.; Pant, S.; Viswanathan, M. DAWN: Designing distributed agents in a worldwide network. *arXiv* **2024**, arXiv:2410.22339. [[CrossRef](#)]
34. Wang, H.; Hunhevicz, J.J.; Hall, D.M. From automation to agency: Prototype for self-owning intelligent buildings enabled by blockchain. *Autom. Constr.* **2025**, *177*, 106309. [[CrossRef](#)]
35. Fang, W.; Zhu, C.; Zhang, W. Toward secure and lightweight data transmission for cloud–edge–terminal collaboration in artificial intelligence of things. *IEEE Internet Things J.* **2024**, *11*, 105–113. [[CrossRef](#)]
36. Luo, J. A bibliometric review on artificial intelligence for smart buildings. *Sustainability* **2022**, *14*, 10230. [[CrossRef](#)]
37. Luan, B.; Feng, X. Artificial intelligence in smart building engineering: A review. *J. Asian Archit. Build. Eng.* **2025**, *25*, 4241–4265. [[CrossRef](#)]
38. Wong, J.K.W.; Li, H.; Wang, S.W. Intelligent building research: A review. *Autom. Constr.* **2005**, *14*, 143–159. [[CrossRef](#)]
39. Mofidi, F.; Akbari, H. Intelligent buildings: An overview. *Energy Build.* **2020**, *223*, 110192. [[CrossRef](#)]
40. Panchalingam, R.; Chan, K.C. A state-of-the-art review on artificial intelligence for smart buildings. *Intell. Build. Int.* **2019**, *13*, 203–226. [[CrossRef](#)]
41. Farzaneh, H.; Malehmirchegini, L.; Bejan, A.; Afolabi, T.; Mulumba, A.; Daka, P. Artificial intelligence evolution in smart buildings for energy efficiency. *Appl. Sci.* **2021**, *11*, 763. [[CrossRef](#)]
42. Zhang, Q.; Wang, L. AI in smart buildings and construction 4.0: Implementation areas and influencing factors in construction organizations. *Autom. Constr.* **2026**, *181*, 106623. [[CrossRef](#)]
43. Habiba, U.E.; Ahmed, I.; Asif, M.; Alhelou, H.H.; Khalid, M. A review on enhancing energy efficiency and adaptability through system integration for smart buildings. *J. Build. Eng.* **2024**, *89*, 109354. [[CrossRef](#)]
44. Masroor, M.; Rezazadeh, J.; Ayoade, J.; Aliehyaei, M. A survey of intelligent building automation with machine learning and IoT. *Adv. Build. Energy Res.* **2023**, *17*, 345–378. [[CrossRef](#)]
45. Himeur, Y.; Elnour, M.; Fadli, F.; Meskin, N.; Petri, I.; Rezgui, Y.; Bensaali, F.; Amira, A. AI-big data analytics for building automation and management systems: A survey, actual challenges and future perspectives. *Artif. Intell. Rev.* **2023**, *56*, 4929–5021. [[CrossRef](#)] [[PubMed](#)]
46. Oulefki, A.; Kheddar, H.; Amira, A.; Kurugollu, F.; Himeur, Y.; Bounceur, A. Innovative AI strategies for enhancing smart building operations through digital twins: A survey. *Energy Build.* **2025**, *335*, 115567. [[CrossRef](#)]
47. Al-Garadi, M.A.; Mohamed, A.; Al-Ali, A.K.; Du, X.; Ali, I.; Guizani, M. A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1646–1685. [[CrossRef](#)]

48. O'Neill, Z.; Wen, J. Artificial intelligence in smart buildings. *Sci. Technol. Built Environ.* **2022**, *28*, 1115. [\[CrossRef\]](#)
49. Alam, S.M.M.; Ali, M.H. An overview of state-of-the-art research on smart building systems. *Electronics* **2025**, *14*, 2602. [\[CrossRef\]](#)
50. Ekanayaka Gunasinghalge, L.U.G.; Alazab, A.; Talukder, M.A. Artificial intelligence for energy optimization in smart buildings: A systematic review and meta-analysis. *Energy Inform.* **2025**, *8*, 135. [\[CrossRef\]](#)
51. Arun, M.; Barik, D.; Chandran, S.R.S.; Praveenkumar, S.; Tudu, K. Economic, policy, social, and regulatory aspects of AI-driven smart buildings. *J. Build. Eng.* **2025**, *99*, 111666. [\[CrossRef\]](#)
52. Kumari, P.; Gupta, H.P.; Mishra, R.; Das, S.K. An energy-efficient smart space system using LoRa network with deadline and security constraints. In *Proceedings of the 24th International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*; Association for Computing Machinery: San Francisco, CA, USA, 2021; pp. 79–86. [\[CrossRef\]](#)
53. Lee, M.J.; Zhang, R.; Shojaei, A.; Roofigari-Esfahan, N. A Human-Building Interaction (HBI) system for multidimensional occupant feedback integration and predictive modeling. *J. Build. Eng.* **2025**, *112*, 113839. [\[CrossRef\]](#)
54. Wang, Y.; Guo, S.; Pan, Y.; Su, Z.; Chen, F.; Luan, T.H.; Li, P.; Kang, J.; Niyato, D. Internet of agents: Fundamentals, applications, and challenges. *IEEE Trans. Cogn. Commun. Netw.* **2026**, *12*, 4476–4498. [\[CrossRef\]](#)
55. Xu, D.; He, X.; Su, T.; Wang, Z. A survey on deep neural network partition over cloud, edge and end devices. *arXiv* **2023**, arXiv:2304.10020. [\[CrossRef\]](#)
56. Marro, S.; La Malfa, E.; Wright, J.; Li, G.; Shadbolt, N.; Wooldridge, M.; Torr, P. A scalable communication protocol for networks of large language models. *arXiv* **2024**, arXiv:2410.11905. [\[CrossRef\]](#)
57. Madanipour, A. *Public and Private Spaces of the City*; Routledge: London, UK, 2003. [\[CrossRef\]](#)
58. Mitchell, W.J. *City of Bits: Space, Place, and the Infobahn*; MIT Press: Cambridge, MA, USA, 1995. [\[CrossRef\]](#)
59. Shakeri, Z.; Benfriha, K.; Varmazyar, M.; Talhi, E.; Quenehen, A. Production scheduling with multi robot task allocation in a real industry 4.0 setting. *Sci. Rep.* **2025**, *15*, 1795. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Kilari, S.D. Use artificial intelligence into facility design and layout planning work in manufacturing facility. *Eur. J. Artif. Intell. Mach. Learn.* **2025**, *4*, 27–34. [\[CrossRef\]](#)
61. Alabadleh, O.S.; Al-Karablieh, M.A. The impact of artificial intelligence on the development of design thinking for interior design patterns in commercial spaces. *Dirasat Hum. Soc. Sci.* **2025**, *52*, 8123–8135. [\[CrossRef\]](#)
62. Guo, X.; Yuan, K. Based on the realization path of accurate analysis and intelligent push of shopping mall users driven by AI agents. In *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (DECS '25)*; IEEE: New York, NY, USA, 2026; pp. 1410–1418. [\[CrossRef\]](#)
63. Iyer, S.S. Usefulness of AI in Dubai Mall and future trends. *J. Pioneer. Artif. Intell. Res.* **2025**, *1*, 1–15. [\[CrossRef\]](#)
64. Lin, X. Five-sense interaction and emotional experience design of commercial exhibition space in the era of digital intelligence. In *Proceedings of the 2025 2nd International Conference on Digital Society and Artificial Intelligence (DSAI '25)*; ACM: New York, NY, USA, 2025. [\[CrossRef\]](#)
65. Zhang, J.; Zhu, T.; Hu, C. Application model of museum cultural heritage educational game based on embodied cognition and immerse experience. *ACM J. Comput. Cult. Herit.* **2025**, *18*, 33. [\[CrossRef\]](#)
66. Ku, E.C.S. Contactless service: Artificial intelligence applications of airports. *Int. J. Hum.-Comput. Interact.* **2024**, *41*, 8884–8896. [\[CrossRef\]](#)
67. Oruma, S.; Colomo-Palacios, R.; Gkioulos, V. Architectural views for social robots in public spaces: Business, system, and security strategies. *Int. J. Inf. Secur.* **2024**, *24*, 1120–1135. [\[CrossRef\]](#)
68. Feng, Y.; Chen, K. Optimization of architectural space in public places under the concept of geographic intelligent interactive design. *GeoJournal* **2025**, *90*, 109–122. [\[CrossRef\]](#)
69. Reis, M.J.C.S.; Serôdio, C. Edge AI for real-time anomaly detection in smart homes. *Future Internet* **2025**, *17*, 179. [\[CrossRef\]](#)
70. Ikegwu, A.C.; Obianuju, O.J.; Nwokoro, I.S.; Kama, M.O.; Ebem, D.U. Investigating the impact of AI/ML for monitoring and optimizing energy usage in smart home. *Artif. Intell. Evol.* **2025**, *6*, 30–45. [\[CrossRef\]](#)
71. Naseer, F.; Addas, A.; Tahir, M.; Khan, M.N.; Sattar, N. Integrating generative adversarial networks with IoT for adaptive AI-powered personalized elderly care in smart homes. *Front. Artif. Intell.* **2025**, *8*, 1520592. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Ha, D.; Schmidhuber, J. Recurrent world models facilitate policy evolution. *Adv. Neural Inf. Process. Syst.* **2018**, *32*. [\[CrossRef\]](#)
73. Cen, J.; Yu, C.; Yuan, H.; Jiang, Y.; Huang, S.; Guo, J.; Li, X.; Song, Y.; Luo, H.; Wang, F.; et al. WorldVLA: Towards Autoregressive Action World Model. *arXiv* **2025**, arXiv:2506.21539. [\[CrossRef\]](#)
74. Chi, C.; Feng, S.; Du, Y.; Xu, Z.; Cousineau, E.; Burchfiel, B.; Song, S. Diffusion policy: Visuomotor policy learning via action diffusion. *Int. J. Rob. Res.* **2025**, *44*, 1684–1704. [\[CrossRef\]](#)
75. Reed, S.; Zolna, K.; Parisotto, E.; Colmenarejo, S.G.; Novikov, A.; Hoffman, G.; Giménez, M.; Sulsky, Y.; Kay, J.; de Freitas, N. A generalist agent. *Trans. Mach. Learn. Res.* **2022**, *44*, 1684–1704. [\[CrossRef\]](#)
76. Brohan, A.; Brown, N.; Carbajal, J.; Chebotar, Y.; Chen, X.; Choromanski, K.; Ding, T.; Driess, D.; Dubey, A.; Zitkovich, S. RT-2: Vision-language-action models transferred to real-world robotic control. *arXiv* **2023**, arXiv:2307.15818. [\[CrossRef\]](#)

77. Hafner, D.; Pasukonis, J.; Ba, J.; Lillicrap, T. Mastering diverse domains through world models. *arXiv* **2023**, arXiv:2301.04104. [CrossRef]
78. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 16000–16009. [CrossRef]
79. Han, X.; Xu, H. Causal intervention and counterfactual reasoning for multimodal pedestrian trajectory prediction. *J. Imaging* **2025**, *11*, 379. [CrossRef] [PubMed]
80. Wu, Z.; Liu, S.; Gao, J. Anticipatory intervention systems in smart cities via causal world modeling. *Sustain. Cities Soc.* **2024**, *105*, 105321. [CrossRef]
81. Liu, J.; Du, Y.; Yang, K.; Wu, J.; Wang, Y.; Hu, X.; Wang, Z.; Liu, Y.; Sun, P.; Boukerche, A.; et al. Edge-cloud collaborative computing on distributed intelligence and model optimization: A survey. *IEEE Commun. Surv. Tutor.* **2026**, Advance online publication. [CrossRef]
82. Trigka, M.; Dritsas, E. Edge and cloud computing in smart cities. *Future Internet* **2025**, *17*, 118. [CrossRef]
83. Zeng, L.; Ye, S.; Chen, X.; Yang, Y. Implementation of big AI models for wireless networks with collaborative edge computing. *arXiv* **2024**, arXiv:2404.17766. [CrossRef]
84. Long, S.; Wang, C.; Long, W.; Liu, H.; Deng, Q.; Li, Z. An efficient task scheduling algorithm in the cloud and edge collaborative environment. *Chin. J. Electron.* **2024**, *33*, 1296–1307. [CrossRef]
85. Li, S.; Wang, H.; Xu, W.; Zhang, R.; Guo, S.; Yuan, J.; Zhong, X.; Zhang, T.; Li, R. Collaborative inference and learning between edge SLMs and cloud LLMs: A survey of algorithms, execution, and open challenges. *arXiv* **2025**, arXiv:2507.16731. [CrossRef]
86. Mittal, A. The evolution of Edge AI: A new paradigm in decentralized cloud computing. *Iconic Res. Eng. J.* **2025**, *8*, 2185–2186.
87. Liu, Y.; Guo, B.; Li, N.; Ding, Y.; Zhang, Z.; Yu, Z. CrowdTransfer: Enabling crowd knowledge transfer in AIoT community. *arXiv* **2024**, arXiv:2407.06485. [CrossRef]
88. Ali, A.; Ullah, I.; Singh, S.K.; Sharafan, A.; Jiang, W.; Sherazi, H.I.; Bai, X. Energy-efficient resource allocation for urban traffic flow prediction in edge-cloud computing. *Int. J. Intell. Syst.* **2025**, *2025*, 1863025. [CrossRef]
89. 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. [CrossRef]
90. Zhuge, M.; Wang, W.; Kirsch, L.; Faccio, F.; Khizbullin, D.; Schmidhuber, J. GPTSwarm: Language agents as optimizable graphs. In *Proceedings of the 41st International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2024; Volume 235, pp. 1–15. [CrossRef]
91. Srivastava, A.K.; Archana, M.; Saidulu, D.; Manellore, P.K.R.; Rao, K.P.; Reddy, J.R. Swarm intelligence for scalable IoT data fusion in smart cities. In *Proceedings of the International Conference on Sustainable Communication Networks and Application*; IEEE: Piscataway, NJ, USA, 2025; pp. 31–38. [CrossRef]
92. Wu, Z. Beyond smart city: The AI city is coming. In *The AI City*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 23–45. [CrossRef]
93. Cui, S.; Xiao, J.-W. Game-based peer-to-peer energy sharing management for a community of energy buildings. *Int. J. Electr. Power Energy Syst.* **2020**, *123*, 106204. [CrossRef]
94. Dedeoglu, V.; Zhang, Q.; Li, Y.; Liu, J.; Sethuvenkatraman, S. BuildingSage: A safe and secure AI copilot for smart buildings. In *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*; ACM: San Francisco, CA, USA, 2024; pp. 369–374. [CrossRef]
95. Fan, C.; Xiao, F.; Wang, H. Smart buildings: State-of-the-art methods and data-driven applications. In *Intelligent Building Fire Safety and Smart Firefighting*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 43–68. [CrossRef]
96. He, T.; Jazizadeh, F. Context-aware LLM-based AI agents for human-centered energy management systems. *arXiv* **2025**, arXiv:2512.25055. [CrossRef]
97. Kukulska-Hulme, A.; Ilıc, P. MALL in the age of AI. In *The Palgrave Encyclopedia of Computer-assisted Language Learning*; Springer: Berlin/Heidelberg, Germany, 2025. [CrossRef]
98. Hasan, M.M.; Pramanik, M.; Alam, I.; Kumar, A.; Avtar, R.; Zhran, M. Assessing the efficacy of artificial intelligence based city-scale blue green infrastructure mapping using Google Earth Engine in the Bangkok metropolitan region. *J. Urban Manag.* **2025**, *14*, 434–450. [CrossRef]
99. National Academies. Foundational Research Gaps and Future Directions for Digital Twins. 2024. Available online: <https://www.nationalacademies.org/projects/DEPS-BMSA-21-03/> (accessed on 6 May 2026).
100. Tang, M.; Nikolaenko, M.; Alrefai, A.; Kumar, A. Metaverse and Digital Twins in the Age of AI and Extended Reality. *Architecture* **2025**, *5*, 36. [CrossRef]
101. Lee, M.S.; Kim, D.H.; Choi, Y. Enhancing EEG-Based Emotion Recognition Using Sparse Dynamic Graph CNN with $\ell_{2,1}$ -Norm. *IEEE Sens. J.* **2025**, *25*, 41472–41480. [CrossRef]

102. Panja, A.K.; Sasidhar, K.; Roy, M.; Chowdhury, C. A survey on crowd behavior analysis through indoor localization. *J. Locat. Based Serv.* **2025**, *19*, 216–255. [[CrossRef](#)]
103. Heda, L.; Sahare, P. Design of an iterative method for crowd behavior analysis integrating faster R-CNN, YOLOv8, and graph convolutional networks. *Signal Image Video Process.* **2025**, *19*, 553–575. [[CrossRef](#)]
104. Liu, W.; Zhu, F.; Wei, L.; Tian, Q. C-CLIP: Multimodal continual learning for vision-language model. In Proceedings of the International Conference on Learning Representations (ICLR 2025), Singapore, 24–28 April 2025. Available online: <https://link.wtturl.cn/?target=https%3A%2F%2Fopenreview.net%2Fforum%3Fid%3Dsb7qHFYwBc&scene=im&aid=497858&lang=zh> (accessed on 2 March 2026).
105. Zhan, H.; Xiao, N.C. A new active learning surrogate model for time- and space-dependent system reliability analysis. *Reliab. Eng. Syst. Saf.* **2025**, *253*, 110536. [[CrossRef](#)]
106. Ilyas, A.; Bawany, N. Crowd dynamics analysis and behavior recognition in surveillance videos based on deep learning. *Multimed. Tools Appl.* **2024**, *84*, 26609–26643. [[CrossRef](#)]
107. Ye, X.; Yigitcanlar, T.; Goodchild, M.; Huang, X.; Li, W.; Shaw, S.L.; Fu, Y.; Gong, W.; Newman, G. Artificial intelligence in urban science: Why does it matter? *Ann. GIS* **2025**, *31*, 181–189. [[CrossRef](#)] [[PubMed](#)]

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.