OPEN ACCESS PRESS (OAP)
OpenAccessPress.com/Journal/JSI

Journal of Spatial Intelligence
ISJN: JSI
Volume 1 Issue 1 - 2025
ISSN: Registering

Editorial
DOI: 10.65736/jsi01012

The Architecture of Spatial Intelligence: A Multiscale Framework for Perception, Representation, Reasoning, Learning, and Action

Yafei Zhao[1,*]

[1] Research Institute of Spatial Intelligence (RISI), Qingdao, China
*Corresponding Author: yafei.zhao@risi.ac.cn

Abstract Spatial intelligence requires an architecture, not only a definition. This perspective proposes a multiscale framework organized around five interdependent capabilities: spatial perception, spatial representation, spatial reasoning, spatial learning and generation, and spatial action and interaction. These capabilities should not be read as a linear pipeline. They form a recurrent cycle in which action changes the environment, altered environments produce new perception, and learning reshapes future representation and reasoning. The framework also treats scale as constitutive. Body-object interaction, interiors, buildings, neighborhoods, cities, regions, planetary systems, and extraterrestrial environments all require spatial intelligence, but they differ in geometry, temporality, agency, uncertainty, and consequence. Finally, the article distinguishes human, machine, environmental, collective, and hybrid human-AI spatial intelligence. It argues that a mature field must connect cognitive science, artificial intelligence, robotics, architecture, urban studies, GIScience, and design research through shared questions about how spatial systems sense, encode, infer, adapt, generate, and act. The proposed architecture is intended as a reusable but open scaffold for comparative research.

Keywords spatial perception; spatial representation; spatial reasoning; multiscale systems; embodied AI; spatial learning; human-AI collaboration

1. Introduction

If the first task of spatial intelligence is conceptual clarification, the second is architectural: how does spatial intelligence operate? A broad definition is necessary, but not sufficient. Without a framework of processes and scales, the term risks becoming a loose label applied to any work involving maps, robots, buildings, cities, or visual scenes. This article proposes a reusable structure for asking what a spatially intelligent system does, at what scale it acts, through which representations, and with what consequences.

The framework begins from the definition that spatial intelligence is the capacity of humans, machines, environments, and hybrid systems to perceive, represent, reason about, learn from, generate, interact with, and act within spatial environments across multiple scales. The phrasing matters because spatial intelligence is not a single faculty, method, or artifact. It is a coordinated set of processes. A person navigating a building, a robot mapping a room, a building control system adapting to occupancy, and an urban digital twin forecasting flood exposure may all be spatially intelligent in

RECEIVED: 10 Dec 2025 **REVISED:** 25 Dec 2025 **PUBLISHED:** 31 Dec 2025

different ways, but only if their spatial operations are explicit and accountable.

Earlier theories provide important foundations. Marr (2010) showed that vision can be analyzed through levels that connect computational goals, representations, algorithms, and implementation. Kuipers (2000) demonstrated that robot spatial knowledge can be organized through linked control, causal, topological, and metric layers. Tversky (2019) argued that spatial thinking is deeply tied to bodily action, gesture, diagrams, and the structure of thought itself. These works differ in discipline and method, but together they suggest that spatial intelligence requires the integration of sensing, representation, inference, and action.

2. Five Capabilities as a Recurrent Cycle

The first capability is spatial perception. Perception concerns how a system detects spatial entities, relations, events, and affordances. For humans, perception includes vision, hearing, touch, proprioception, memory, expectation, and attention. For machines, it may include RGB images, depth maps, lidar, inertial sensors, audio, semantic segmentation, remote sensing, or environmental telemetry. For ambient environments, perception may involve distributed sensing through occupancy detectors, energy meters, cameras, weather stations, and building systems. Perception is spatial when it identifies not only what is present, but where it is, how it is arranged, what it occludes, what it affords, and how it may change.

The second capability is spatial representation. Spatial intelligence depends on ways of encoding environments so they can be remembered, communicated, computed, and redesigned. Representations include cognitive maps, sketches, diagrams, architectural plans, building information models, graph structures, meshes, point clouds, ontologies, raster layers, semantic maps, and digital twins. No representation is neutral. Each emphasizes certain relations and suppresses others. A topological graph may support wayfinding while ignoring metric detail. A photorealistic mesh may capture surface appearance while omitting ownership, accessibility, or thermal comfort. A floor plan may show organization while hiding actual patterns of use.

The third capability is spatial reasoning. Reasoning asks what follows from a spatial representation and a spatial context. It includes judging containment, adjacency, visibility, reachability, orientation, distance, support, collision, accessibility, exposure, causality, and possible futures. Human reasoning often combines qualitative relations with embodied simulation: one imagines turning a corner, lifting an object, or crossing a street. Machine reasoning may involve path planning, constraint solving, probabilistic inference, graph search, causal modelling, or language-grounded question answering. Spatial reasoning is strongest when it connects geometry with semantics and purpose. Knowing that two rooms share a wall is different from knowing whether sound, heat, smoke, or social interaction can pass between them.

The fourth capability is spatial learning and generation. Learning enables systems to improve from examples, exploration, simulation, feedback, and failure. It may produce better maps, navigation policies, occupancy predictions, design alternatives, or spatial world models. Generation extends learning toward the production of new spatial possibilities: layouts, routes, scenes, urban scenarios, adaptive control policies, or imagined environments. This capability must be handled carefully. A generated plan is not spatially intelligent merely because it looks plausible. It must satisfy constraints, support use, respect context, and remain open to evaluation.

The fifth capability is spatial action and interaction. Spatial intelligence culminates in intervention, but action should be understood broadly. It may be bodily movement, robotic manipulation, adaptive lighting, a redesign of a building layout, a change in urban policy, or a collaborative decision in a digital twin. Thrun et al. (2005) showed how probabilistic robotics treats perception, localization, mapping, and action under uncertainty as connected problems. Spatial action is always consequential because it changes the environment or changes how others can act within it.

These five capabilities form a cycle: perception leads to representation; representation supports reasoning; reasoning guides learning and generation; learning and generation inform action and interaction; action changes the environment, producing new perception. The cycle is not rigid. Humans often perceive through action, and machines may learn directly from sensorimotor loops. The key point is that no capability should be studied in isolation for too long. A perception model that cannot support action, a representation that cannot support reasoning, or a generative model that cannot be evaluated in use is incomplete.

The cycle also clarifies why spatial intelligence is difficult to benchmark. Each capability can be tested separately, but real spatial competence appears in their coordination. An indoor agent may perceive doors accurately but fail to choose a route. A design model may generate attractive rooms but fail to represent circulation. An urban model may predict movement but fail to explain how an intervention changes behavior. The architecture therefore encourages compound tasks in which perception, representation, reasoning, learning, and action are evaluated together.

3. Spatial Scale and Difference

Spatial intelligence is multiscale because spatial problems change character as scale changes. At the body and object scale, the central issues are shape, orientation, graspability, affordance, material contact, and immediate sensorimotor coordination. A cup handle, a surgical tool, a stair tread, or a rock sample on another planet requires spatial understanding measured in centimeters, forces, and bodily reach.

At the room and interior scale, spatial intelligence concerns visibility, enclosure, furniture, circulation, acoustics, comfort, social distance, and activity patterns. A person recognizes a room through use and atmosphere as much as through geometry. A machine may need to infer which surfaces are walkable, which objects are movable, and which configurations support a task. Interiors show why spatial intelligence cannot be reduced to external shape.

At the building scale, the problem expands to organization, program, vertical circulation, structure, environmental systems, safety, and human behavior. Buildings connect perception and representation through plans, sections, sensor data, codes, and lived experience. They are also sites where hybrid intelligence becomes tangible: human occupants, designers, facility managers, robots, and control systems may all participate in spatial adaptation.

At the street and neighborhood scale, spatial intelligence must address paths, edges, landmarks, public space, mobility, land use, social meaning, and the relation between designed form and everyday practice. At the city scale, it must include networks, flows, densities, infrastructures, environmental exposure, housing, logistics, governance, and collective behavior. The cognitive map literature reviewed by Epstein et al. (2017) is relevant here because navigation links neural coding, landmarks, route planning, and memory, but cities add institutional and infrastructural complexity beyond individual navigation.

At regional and territorial scales, spatial intelligence addresses ecological processes, watersheds, climate risks, logistics corridors, energy systems, migration, and geopolitical boundaries. These problems require geographic reasoning about distance, connectivity, scale, heterogeneity, and uncertainty. At planetary and extraterrestrial scales, spatial intelligence must cope with extreme environments, sparse data, delayed communication, autonomous exploration, and new relations between local action and remote interpretation. A rover on Mars, a satellite observing Earth, and a climate model simulating regional futures are not the same kind of system, yet each depends on spatial perception, representation, reasoning, learning, and action.

Continuity across scales should not erase difference. The same word "navigation" means different things for a hand approaching an object, a person moving through a hospital, a delivery vehicle crossing a city, a migratory route through a region, and a spacecraft traversing unknown terrain. A general theory of spatial intelligence must therefore be comparative. It should ask which spatial operations transfer across scales and which must be redefined.

4. Subjects of Spatial Intelligence

Spatial intelligence also differs by subject. Human spatial intelligence involves perception, memory, imagination, language, gesture, emotion, cultural knowledge, and bodily skill. It is not merely an internal map. It is enacted through movement, tools, collaboration, and situated interpretation. People understand a station, market, temple, apartment, or landscape partly through learned practices and meanings.

Machine spatial intelligence involves computational systems that sense, model, infer, learn, generate, and act. Its strengths include scale, precision, persistence, simulation, and rapid search. Its weaknesses often include brittleness, lack of commonsense, weak causal grounding, poor transfer across places, and limited awareness of human meaning. Machines may process more spatial data than humans, yet still fail to understand what matters in a situation.

Environmental or ambient intelligence refers to built or natural settings equipped with sensing, computation, and adaptive response. Intelligent buildings, responsive interiors, and sensor-rich public spaces are examples. Their intelligence is not located in one mind or model, but in the coupling of space, devices, occupants, and governance. This form of intelligence raises questions about privacy, control, consent, maintenance, and unintended behavior.

Collective spatial intelligence emerges when groups coordinate spatial knowledge and action. A community mapping flood risk, a planning team evaluating alternatives, or a crowd adapting to an evacuation route all create spatial intelligence through shared representations and distributed judgment. Hybrid human-AI spatial intelligence combines human interpretation with computational perception, simulation, or generation. The central challenge is not replacement but coordination: how should human values, machine inference, and environmental feedback shape each other?

5. Toward an Open Architecture

The architecture proposed here is unified but intentionally open. It unifies spatial intelligence through five capabilities and multiple scales, but it does not prescribe one representation, one metric, or one discipline. It invites cognitive scientists to connect mental spatial processes with computational models; AI researchers to connect perception with reasoning and action; roboticists to situate navigation within architectural and urban contexts; architects and urbanists to make spatial theory computable without reducing space to geometry; geographers to connect large-scale spatial analysis with embodied and experiential questions; and design researchers to treat generation as a testable spatial process.

Table 1. Core capabilities of spatial intelligence across spatial scales.

Capability	Central question	Typical inputs	Typical outputs
Spatial perception	What spatial entities and relations can be sensed in the environment?	Images, depth, sound, lidar, maps, traces, gestures, environmental sensors	Detected objects, surfaces, landmarks, affordances, events, and states
Spatial representation	How should spatial information be encoded for memory, communication, computation, and design?	Percepts, surveys, point clouds, plans, graphs, semantic labels, trajectories	Maps, cognitive schemas, meshes, graphs, ontologies, digital twins, design models
Spatial reasoning	What can be inferred about relations, causes, constraints, and possible futures?	Representations, rules, experience, simulation models, contextual knowledge	Routes, explanations, constraints, predictions, alternatives, risk assessments
Spatial learning and generation	How can systems improve from spatial experience and produce plausible spatial alternatives?	Datasets, demonstrations, simulations, feedback, performance objectives	Learned policies, world models, generated layouts, synthetic scenes, adaptive designs
Spatial action and interaction	How can intelligent agents intervene responsibly in spatial environments?	Goals, plans, affordances, constraints, human preferences, control signals	Movements, manipulations, design changes, adaptive environmental responses, collaborative actions

Table 1 summarizes the five capabilities. Figure 1 shows how they form a feedback cycle across scale. The diagram places five recurrent capabilities across seven spatial scales and shows feedback from action and interaction to new perception through changed environmental conditions. The framework is meant to help authors locate their contributions precisely. Does a study improve perception, representation, reasoning, learning, generation, action, or their coupling? Which scale does it address? Which subject of intelligence is being claimed? What changes when the same method moves from a room to a city, or from a simulation to a lived environment?

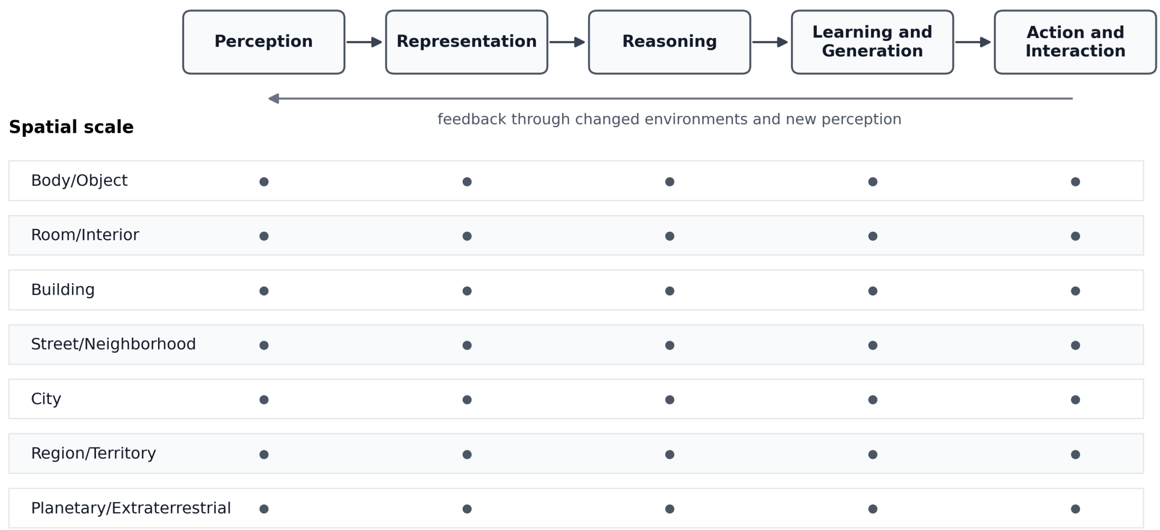

Figure 1. A multiscale architecture of spatial intelligence.

A mature field of spatial intelligence will not be built by choosing between humans and machines, cognition and computation, or design and analysis. It will be built by studying how spatial systems become capable of understanding and acting within environments that are geometric, semantic, embodied, social, ecological, and political at the same time.

This openness is especially important for design and planning fields, where spatial intelligence is often expressed through proposals rather than predictions. A proposal can be treated as an intelligent act when it reveals spatial relations, tests possible futures, and supports accountable choice. The framework therefore treats generation as part of inquiry, not as decoration after analysis.

References

[1] Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. Nature Neuroscience, 20(11), 1504-1513.

[2] Kuipers, B. (2000). The Spatial Semantic Hierarchy. Artificial Intelligence, 119(1-2), 191-233.

[3] Marr, David. Vision: A computational investigation into the human representation and processing of visual information. MIT press, 2010.

[4] Thrun, Sebastian. "Probabilistic robotics." Communications of the ACM 45.3 (2002): 52-57.

[5] Tversky, Barbara. Mind in motion: How action shapes thought. Basic Books, 2019.