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/jsi01013

Building the Field of Spatial Intelligence: Grand Challenges, Benchmark Tasks, and an Open Research Agenda

Yafei Zhao[1,*]

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

Abstract Spatial intelligence is not yet a fully unified research field. It is distributed across cognitive science, computer vision, robotics, GIScience, architecture, urban analytics, spatial computing, and design research, where similar terms often refer to different scales, methods, and evaluation criteria. This perspective proposes an open research agenda for the next decade. It identifies why current work remains fragmented: disciplinary definitions diverge, benchmark tasks are scale-bound, computer vision often privileges recognition over spatial causality, robotics can understate architectural and urban context, design fields hold rich spatial theories but uneven computational models, and GeoAI does not always address embodied experience. The article then proposes eight grand challenges, from multimodal perception and 3D/4D representation to causal reasoning, embodied interaction, generative world models, human experience, multiscale territorial intelligence, and responsible spatial systems. It argues that future benchmarks must evaluate geometry, semantics, topology, causality, task success, feasibility, environmental performance, usability, robustness, fairness, explainability, and transfer across places. The field will mature through open datasets, reproducible workflows, transparent benchmarks, shared ontologies, responsible governance, and international collaboration.

Keywords research agenda; benchmark tasks; spatial AI; embodied navigation; digital twins; responsible AI; open science

1. Introduction

Spatial intelligence has become a necessary concept before becoming a settled field. The problems are visible everywhere: models that can describe images but struggle with spatial causality, robots that navigate simulated rooms but fail in complex buildings, urban analytics platforms that predict flows without understanding lived experience, and design systems that generate forms whose spatial validity remains difficult to evaluate. These problems are not isolated technical gaps. They indicate that spatial intelligence requires shared definitions, benchmark tasks, evaluation criteria, datasets, and governance practices.

The field should not be forced into premature unity. Its strength lies in the fact that it brings together human spatial cognition, artificial intelligence, robotics, architecture, urban studies, GIScience, spatial computing, digital twins, and environmental governance. Yet fragmentation now limits comparison and accumulation. Two studies may both claim to

RECEIVED: 18 Dec 2025 **REVISED:** 29 Dec 2025 **PUBLISHED:** 31 Dec 2025

address spatial reasoning while one tests language descriptions of object relations, another evaluates route planning, and a third simulates land-use change. A shared agenda does not require one method. It requires a way to locate methods in relation to common spatial capabilities and scales.

This article proposes such an agenda. Spatial intelligence is understood here as 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 agenda asks what research tasks, benchmarks, and open science practices would allow that definition to become empirically productive.

2. Why the Field Remains Fragmented

The first source of fragmentation is terminology. Spatial cognition, spatial ability, GeoAI, embodied AI, spatial computing, digital twin, urban intelligence, and architectural intelligence all contain claims about space and intelligence, but they do not use the same units of analysis. Cognitive science may focus on individual memory and navigation. GeoAI may focus on geospatial prediction and knowledge discovery, as Janowicz et al. (2020) argue in their account of spatially explicit artificial intelligence. Robotics may emphasize localization, control, and task completion. Architecture may foreground program, atmosphere, inhabitation, and performance. These differences are productive, but only if they are visible.

The second source is benchmark specialization. Data resources such as Matterport3D provide rich building-scale RGB-D scenes for 3D understanding (Chang et al., 2017), while embodied AI platforms such as Habitat make it possible to test navigation and interaction at scale in simulation (Savva et al., 2019). These resources have been transformative, yet they also reveal a gap. Benchmarks are often organized around particular modalities, tasks, or scales. Indoor visual navigation, urban remote sensing, floor-plan analysis, and regional environmental modelling may advance independently without shared measures of spatial transfer.

The third source is a mismatch between recognition and reasoning. Computer vision has made extraordinary progress in detection, segmentation, reconstruction, and scene classification. Spatial intelligence needs those capacities, but it also asks why spatial arrangements matter. A model may detect a stair, wall, door, and crowd without reasoning about evacuation, accessibility, acoustic separation, privacy, or risk propagation. A spatial system must understand relations, constraints, and possible interventions, not only labels.

The fourth source is uneven treatment of context. Robotics and embodied AI often evaluate agents in rooms or simulated environments, but built environments are not generic containers. They have codes, thresholds, maintenance regimes, social norms, and institutional purposes. Architecture and urban studies have deep theories of such spatial context, but their concepts are not always expressed in computable, testable forms. GIScience and GeoAI contribute strong methods for large-scale spatial data, yet they may underrepresent embodied experience, situated perception, and design intent. A field of spatial intelligence must make these partial strengths interoperable.

Finally, data governance remains immature. Spatial data often involves people, property, mobility, ecology, and security. Dataset documentation, as argued by Gebru et al. (2021), is not a bureaucratic afterthought; it is part of responsible model development. Spatial intelligence needs documentation practices that address scale, sampling, uncertainty, consent, cultural meaning, and downstream use.

3. Eight Grand Challenges

The first grand challenge is multimodal spatial perception. Future systems must integrate images, depth, lidar, maps, language, sound, motion, environmental sensors, and human input. The challenge is not only sensor fusion. It is the construction of coherent spatial situations from partial, noisy, and socially interpreted signals. A benchmark might ask systems to integrate indoor scans, exterior street scenes, floor plans, and natural-language descriptions into a common spatial account.

The second challenge is 3D and 4D spatial representation. Spatial intelligence needs representations that combine geometry, semantics, topology, time, uncertainty, and agency. Static 3D models are insufficient for changing buildings, streets, territories, and planetary environments. Digital twin research shows the importance of linking physical and virtual systems, but Fuller et al. (2020) also identify open challenges in enabling technologies and standardization. Spatial intelligence should extend digital twins toward reasoning, accountability, and human interpretability.

The third challenge is commonsense and causal spatial reasoning. Systems should infer not only that two entities are near, but why proximity matters. Does it support interaction, create risk, imply ownership, enable visibility, increase exposure, or constrain movement? Benchmark tasks should test counterfactual spatial questions: What happens if this door is closed, this corridor is blocked, this shade structure is added, or this transport link fails?

The fourth challenge is embodied navigation, manipulation, and interaction. Agents must act in spaces that include obstacles, affordances, people, rules, and changing goals. Navigation should be evaluated beyond shortest paths. It should include safety, comfort, social legibility, energy use, accessibility, and recovery from error. Manipulation should be linked to material and spatial constraints, not treated as isolated object control.

The fifth challenge is generative design and spatial world models. Generative systems can produce images, plans, routes, scenes, and scenarios, but generation must be tied to spatial validity. A plausible-looking floor plan may violate circulation, daylight, egress, structure, or cultural use. A generated urban scenario may optimize one metric while worsening heat exposure or displacement. Spatial world models should support imagination, simulation, and critique, not only synthesis.

The sixth challenge is human spatial experience and cognitive-behavioral modelling. Spatial intelligence must account for perception, memory, emotion, attention, habit, accessibility, and social difference. Human movement and behavior prediction should not reduce people to particles. It should represent purposes, constraints, uncertainty, and dignity. Benchmarks in this area should combine trajectories with surveys, experiments, ethnographic interpretation, and privacy-preserving data governance.

The seventh challenge is multiscale urban and territorial intelligence. Buildings affect streets, streets affect neighborhoods, cities depend on regions, and regions are embedded in planetary systems. Spatial intelligence should connect scales rather than treating them as separate domains. Cross-scale reasoning might link a building retrofit to grid demand, heat exposure, mobility patterns, and regional emissions.

The eighth challenge is responsible, inclusive, and trustworthy spatial intelligence. Spatial systems can exclude, surveil, misclassify, or intensify unequal exposure to risk. Responsible spatial intelligence requires fairness across places and populations, explainability for spatial decisions, robustness under environmental change, and mechanisms for contestation. Ethics should not be placed after technical work. It should shape problem definition, data collection, benchmarking, and deployment.

4. Benchmark Tasks and Evaluation

A field matures when its claims can be compared without erasing disciplinary richness. Spatial intelligence needs benchmark tasks that connect capabilities and scales. Candidate tasks include spatial relation understanding, floor-plan and building-layout reasoning, indoor-outdoor scene integration, embodied navigation, urban scene interpretation, human movement and behavior prediction, spatial question answering, environmental performance prediction, generative architectural and urban design, real-time digital-twin decision-making, human-robot spatial collaboration, and cross-scale spatial reasoning.

The next generation of benchmarks should be modular. A building-layout benchmark could include geometric validity, semantic program relations, code constraints, simulated environmental performance, and human usability. An embodied navigation benchmark could include path efficiency, collision avoidance, instruction following, recovery behavior, and social acceptability. A digital-twin benchmark could evaluate whether a system correctly updates state, predicts consequences, supports intervention, and communicates uncertainty to decision-makers.

Evaluation cannot rely only on classification accuracy. Accuracy remains useful when a task is classification, but spatial intelligence also requires geometric accuracy, semantic consistency, topological validity, causal reasoning ability, task success, physical feasibility, environmental performance, human usability, robustness, fairness, explainability, and transferability across scales and places. Some metrics will be quantitative, such as localization error, energy performance, graph validity, or success-weighted path length. Others will require expert review, user studies, participatory evaluation, or policy analysis.

Benchmark design should also distinguish between closed-world and open-world spatial tasks. In a closed-world task, all relevant objects, rules, and goals may be known in advance. In an open-world task, systems encounter incomplete maps, ambiguous language, changing occupancy, unexpected weather, contested boundaries, or social norms that were not encoded in the training data. Spatial intelligence will be overestimated if benchmarks reward only performance in clean, fully specified environments.

Transfer is especially important. A system trained on one city, culture, building type, climate, or simulator may fail elsewhere. Benchmarks should therefore report place diversity, sampling assumptions, temporal coverage, annotation protocols, and known limitations. Dataset datasheets and model cards should be adapted for spatial data, including documentation of coordinate systems, scale, privacy risk, sensor bias, missing populations, and conditions of legitimate use.

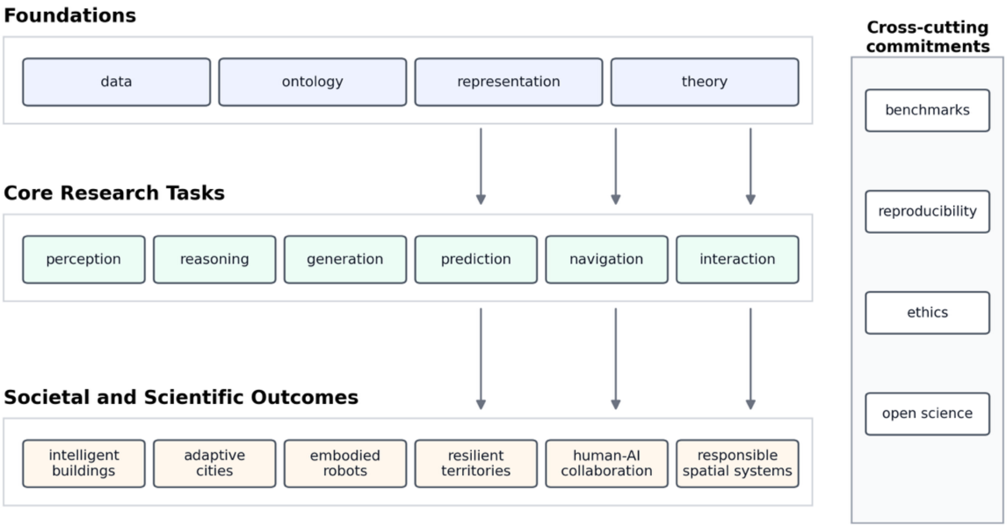

Figure 1. An open research agenda for spatial intelligence.

Figure 1 presents the agenda as a layered structure from foundations to tasks to outcomes, with benchmarks, reproducibility, ethics, and open science as cross-cutting commitments. The diagram frames field building through foundations, core research tasks, and societal and scientific outcomes, with benchmarks, reproducibility, ethics, and open science as cross-cutting commitments.

5. Open Science and Field Building

The field of spatial intelligence should be built through open science rather than closed claims of novelty. Open datasets are necessary, but not sufficient. Datasets need clear licenses, spatial metadata, provenance, uncertainty documentation, privacy review, and mechanisms for community correction. Reproducible workflows should include code, model settings, preprocessing steps, simulation assumptions, and evaluation scripts. Transparent benchmarks should make failure visible, not merely rank systems by a single score.

Shared ontologies and interdisciplinary terminology will be equally important. Researchers need ways to align concepts such as affordance, topology, place, accessibility, scene, territory, route, public space, and environmental performance. These terms cannot be flattened into one vocabulary, but their relationships can be made explicit. International and cross-cultural collaboration is essential because spatial intelligence is always shaped by local environments, social norms, legal systems, and cultural meanings.

Open science also has an institutional dimension. Journals, conferences, laboratories, public agencies, and communities should reward careful negative results, benchmark failures, replication studies, and dataset audits. These forms of work are sometimes less visible than new models, but they are essential for a field whose claims may affect buildings, streets, mobility, infrastructure, and environmental risk.

The next decade should therefore be organized around common problems: how spatial systems perceive, represent, reason, learn, generate, interact, and act responsibly across scales. The invitation is broad but not vague. Cognitive scientists, AI researchers, roboticists, geographers, architects, urbanists, designers, environmental scientists, and policy scholars can build the field together by making their assumptions legible and their evidence comparable. Spatial intelligence will mature not when one discipline defines it for all others, but when shared research practices allow diverse forms of spatial knowledge to meet, disagree, and improve one another.

References

[1] Chang, Angel, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. "Matterport3d: Learning from rgb-d data in indoor environments." arXiv preprint arXiv:1709.06158 (2017).

[2] Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE Access, 8, 108952-108971.

[3] Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. "Datasheets for datasets." Communications of the ACM 64, no. 12 (2021): 86-92.

[4] Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625-636.

[5] Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., Straub, J., Liu, J., Koltun, V., Malik, J. and Parikh, D., 2019. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9339-9347).