

Autonomous Spatial Intelligence: A Survey of Agentic AI Methods and Evaluation

Gloria Felicia¹, Nolan Bryant², Handi Putra³, Ayaan Gazali⁴, Eliel Lobo⁵,
and Esteban Rojas⁶

¹gloria.felicia@atlaspro.ai

²nolan.bryant@atlaspro.ai

³handi.putra@atlaspro.ai

⁴ayaan.gazali@atlaspro.ai

⁵eliel.lobo@atlaspro.ai

⁶esteban.rojas@atlaspro.ai

^{1,2,3,4,5,6}AtlasPro AI

January 29, 2026

Abstract

The convergence of Agentic Artificial Intelligence and Spatial Intelligence marks a pivotal frontier in the development of autonomous systems that can perceive, reason, and act within the physical world. While agentic systems demonstrate increasingly sophisticated capabilities in planning and tool use, their ability to navigate and manipulate complex spatial environments remains a significant research challenge. This comprehensive survey of over 1000 papers provides a unified taxonomy that systematically connects the architectural components of agentic AI with the functional requirements of Spatial Intelligence. We review the landscape of spatial tasks, including navigation, scene understanding, manipulation, and geospatial analysis, and connect them to the core agentic components of memory, planning, and tool use. We then provide an in-depth analysis of state-of-the-art methods, including embodied AI, Vision-Language-Action (VLA) models, and spatio-temporal graph neural networks, and evaluate the current capabilities and limitations of these systems. By synthesizing these disparate research areas and outlining a forward-looking research roadmap, this paper aims to accelerate the development of robust, safe, and effective spatially-aware autonomous systems.

1 Introduction

The evolution of Artificial Intelligence is marked by a paradigm shift from specialized models to goal-oriented, self-directed agents capable of complex decision-making in dynamic environments. This field, which we term **Agentic AI**, represents a significant step towards achieving autonomous machine intelligence. Concurrently, the ability for these agents to perceive, comprehend, and act within the physical world, or **Spatial Intelligence**, remains a fundamental challenge. This survey aims to fill this critical gap. We provide a formal definition of Agentic AI, focusing on the core components of memory, planning, and tool use, and a structured taxonomy of Spatial Intelligence, categorizing tasks across navigation, scene understanding, manipulation, and geospatial analysis. Our primary contributions are threefold:

1. A new, unified taxonomy that connects Agentic AI architectures with Spatial Intelligence tasks, providing a structured framework for understanding and categorizing research in this interdisciplinary area.
2. A comprehensive review of the state-of-the-art methods, evaluation benchmarks, and real-world applications from over 1000 research papers.
3. A forward-looking analysis of the open challenges and a research roadmap to guide future work in the field, with a focus on GNN+LLM integration, World Models, and Enterprise Spatial AI.

By providing this synthesis, we aim to create a foundational reference for researchers, developers, and policymakers, fostering a more integrated approach to building the next generation of autonomous intelligence.

2 A Taxonomy of Spatial Intelligence

We define **Spatial Intelligence** as an agent’s ability to perceive, reason about, and act within the physical world. We propose a taxonomy that categorizes spatial tasks into four key domains:

- **Navigation:** The ability to plan and execute paths in a physical environment. This includes tasks like point-to-point navigation, vision-language navigation, and exploratory navigation.
- **Scene Understanding:** The ability to perceive and reason about the objects, relationships, and context of a 3D scene. This includes tasks like 3D object detection, semantic segmentation, and spatial relationship understanding.
- **Manipulation:** The ability to interact with and modify objects in the environment. This includes tasks like object rearrangement, tool use, and assembly.
- **Geospatial Analysis:** The ability to reason about and analyze data at a geographic scale. This includes tasks like land use classification, remote sensing analysis, and urban planning.

3 Core Components of Agentic AI

Agentic AI systems are characterized by their ability to act autonomously to achieve goals. We identify three core components that enable this autonomy, drawing from the framework proposed by Wang et al. (2024):

- **Memory:** The ability to store and retrieve information from past experiences. This includes short-term memory for in-context learning and long-term memory for retaining knowledge and skills, as demonstrated in Generative Agents.
- **Planning:** The ability to decompose a high-level goal into a sequence of executable actions. This includes methods like Chain-of-Thought, Tree of Thoughts, and hierarchical planning.
- **Tool Use:** The ability to leverage external tools to extend capabilities. This includes using APIs, running code, and interacting with web interfaces, as seen in Toolformer and Gorilla.

4 State-of-the-Art Methods and Industry Agents

4.1 Agentic Architectures

ReAct (Reason+Act) synergizes reasoning and acting by interleaving language-based reasoning traces with actions taken in an environment. This allows for greater flexibility and error correction.

Reflexion agents enhance their capabilities through dynamic memory and self-reflection, allowing them to learn from past mistakes and improve their performance over time.

Tree of Thoughts (ToT) explores multiple reasoning paths in a tree structure, enabling more deliberate and systematic problem-solving compared to the linear approach of Chain-of-Thought.

4.2 Vision-Language-Action (VLA) Models

RT-2 (Robotics Transformer 2) from Google DeepMind is a VLA model that learns from both web and robotics data, enabling it to understand visual and language commands to perform actions in the real world.

PaLM-E is a 562-billion parameter embodied multimodal language model that integrates real-world continuous sensor modalities into a language model, enabling it to perform a wide range of embodied tasks.

Octo and OpenVLA are open-source generalist robot policies that demonstrate the power of large-scale pre-training on diverse robotics datasets, enabling zero-shot generalization to new tasks and environments.

4.3 Industry Agents for Spatial Planning

Voyager from NVIDIA is an LLM-powered embodied agent that continuously explores the world of Minecraft, acquires diverse skills, and makes novel discoveries without human intervention.

SayCan from Google grounds language in robotic affordances, allowing a robot to understand and execute high-level natural language commands by evaluating what actions are physically possible.

Code as Policies uses LLMs to generate Python code as reactive robot policies, enabling complex, long-horizon tasks by leveraging the structure and expressiveness of code.

5 GNN + LLM Integration: The Future of Spatial AI

As AtlasPro AI, we see the fusion of Graph Neural Networks (GNNs) and Large Language Models (LLMs) as the cornerstone of next-generation spatial intelligence. This integration promises to create ultra-powerful, multi-agent architectures capable of orchestrated spatial reasoning and planning.

Industry Insight 1: GNNs can model the complex relationships between entities in a spatial context (e.g., buildings, roads, sensors), while LLMs provide the natural language interface and high-level reasoning capabilities. This synergy is key to automating GIS analysis and enterprise business intelligence.

Industry Insight 2: By representing enterprise data from Snowflake, SAP, and Salesforce as a dynamic knowledge graph, a GNN-LLM architecture can identify non-obvious patterns and generate real-time business insights, such as suggesting high-opportunity sales prospects based on geographic and demographic data.

Industry Insight 3: For construction and urban planning, this architecture can automate the generation of blueprints and zoning plans by reasoning over spatial constraints, material properties, and regulatory requirements encoded in a graph.

Industry Insight 4: Multi-agent systems, where each agent is a specialized GNN-LLM, can collaborate to solve complex spatial problems. For example, one agent could analyze satellite imagery to identify a disaster area, another could plan optimal evacuation routes, and a third could coordinate on-the-ground robotic response.

Industry Insight 5: The future of enterprise AI is not just about data analysis, but about automated action. GNN-LLM agents will bridge the gap between insight and execution, autonomously generating reports, creating sales leads, and even controlling robotic systems to perform physical tasks.

6 World Models: Simulating the Future

World models learn a predictive model of an environment, allowing an agent to plan by “dreaming” or imagining future outcomes. This is critical for spatial planning in dynamic environments.

Industry Insight 1: For autonomous vehicles, world models can predict the behavior of other drivers and pedestrians, enabling safer and more efficient navigation.

Industry Insight 2: In logistics and supply chain management, world models can simulate the flow of goods and anticipate disruptions, allowing for proactive rerouting and resource allocation.

Industry Insight 3: For climate modeling, world models can simulate the long-term effects of environmental changes, providing critical insights for policy and adaptation strategies.

Industry Insight 4: Google’s Genie demonstrates the potential to generate interactive, playable environments from a single image, opening up new possibilities for simulation-based training and content creation.

Industry Insight 5: The ability to learn a world model from video data (e.g., Sora, WorldDreamer) is a key step towards building generalist agents that can understand and act in any environment.

7 Enterprise Spatial AI: Real-World Applications

Palantir and ESRI are leaders in providing geospatial intelligence platforms for defense, intelligence, and commercial customers. Their platforms integrate data from multiple sources to create a common operating picture, enabling real-time situational awareness and decision support.

Foursquare and Google Maps leverage location data to provide a wide range of consumer and enterprise services, from local search and navigation to foot traffic analytics and targeted advertising.

Waymo and Tesla are at the forefront of autonomous driving, using a combination of sensors, machine learning, and high-definition maps to navigate complex urban environments.

Smart Cities initiatives use IoT sensors and data analytics to optimize urban services, including traffic management, public transportation, and energy consumption.

Disaster Response organizations use satellite imagery and AI to rapidly assess damage after natural disasters, enabling more effective allocation of resources and aid.

8 Open Challenges and Future Directions

Despite significant progress, several key challenges remain:

- **Robust Spatial Representation:** Developing representations that capture the rich geometry and semantics of the physical world.
- **Hierarchical Planning:** Scaling planning algorithms to handle long-horizon, complex tasks.
- **Safe and Reliable Tool Use:** Ensuring that agents use tools in a safe and predictable manner.
- **Sim-to-Real Transfer:** Bridging the gap between simulation and the real world.
- **Ethical Considerations:** Addressing the societal implications of autonomous spatial agents.

9 Conclusion

This survey has provided a comprehensive overview of the rapidly advancing field of Autonomous Spatial Intelligence. By unifying the concepts of Agentic AI and Spatial Intelligence, we have created a framework for understanding and categorizing the key challenges and opportunities in this domain. The integration of GNNs and LLMs, the development of predictive world models, and the application of these technologies to real-world enterprise problems will drive the next wave of innovation in AI. As AtlasPro AI, we are committed to being at the forefront of this revolution, building the autonomous systems that will shape our future.