Connotation and Technical Progress of Spatial Intelligence and Spatio-Temporal Intelligence

Hao Liu, Hanxin Chen, Hong Ma, Jianfeng Deng, Jing Chen
Chongqing Academy of Surveying and Mapping, Chongqing, China

Key words: Spatial Intelligence, Spatio-Temporal Intelligence, Artificial Intelligence, 3D Understanding, Dynamic Prediction

SUMMARY

As artificial intelligence (AI) shifts from processing digital information to deep integration with the physical world, Spatial Intelligence (SI) and Spatio-Temporal Intelligence (STI) have emerged as two pivotal, yet distinct, frontier branches. This paper systematically elucidates their conceptual distinctions and reviews key technological advancements. SI focuses on enabling machines to perceive, represent, reason about, and interact with the 3D physical world, with its core being the understanding of spatial structures and geometric properties. Key technological progress in SI encompasses 3D perception, neural field-based representation and modeling, spatial reasoning and planning, and content generation for embodied interaction in domains like robotics and the Metaverse. Conversely, STI emphasizes analyzing the dynamic evolution patterns of objects and events within a unified spatio-temporal framework, aiming to mine patterns of change for decision support. Its key technologies involve the fusion of multi-modal spatio-temporal big data, spatio-temporal data mining, deep learning-based prediction models, spatio-temporal large models, and digital twin-driven simulation and scenario analysis. This paper clarifies their fundamental divergence: SI answers "Where is it and how to operate?" while STI addresses "How does it change and how to decide?". Despite their different emphases, they are converging in complex real-world applications like autonomous driving, collectively underpinning the development of more autonomous AI systems and progress toward Artificial General Intelligence (AGI). The paper concludes with a discussion of future challenges and directions concerning data, model robustness, and ethical considerations.

Support Information: Performance Incentive and Guidance Project of Scientific Research Institutions in Chongqing in 2025（CSTB2025JXJL-YFX0008, CSTB2025JXJL-YFX0040）

Connotation and Technical Progress of Spatial Intelligence and Spatio-Temporal Intelligence

Hao Liu, Hanxin Chen, Hong Ma, Jianfeng Deng, Jing Chen, China

1. INTRODUCTION

Artificial Intelligence (AI) technology is penetrating social production models and human lifestyles at an unprecedented rate. The developmental focus of AI is shifting from processing purely digital information (e.g., text, images) towards deep integration with the complex physical world. How to endow machines with human-like capabilities for perceiving, understanding, reasoning, and interacting with the physical world has become a core challenge and a critical opportunity in the field of AI. In this context, Spatial Intelligence and Spatio-Temporal Intelligence, as two frontier branches of AI, have emerged as pivotal technical pillars underpinning this crucial leap, garnering significant attention from both academia and industry.

The core objective of Spatial Intelligence is to equip machines with the ability to perceive, represent, reason about, and ultimately interact effectively with the 3D physical world. Its essence lies in understanding the geometric properties (e.g., shape, location, pose), interrelationships, and underlying physical laws of three-dimensional space.

Spatio-Temporal Intelligence emphasizes the modeling, inference, prediction, and decision support under a unified spatio-temporal framework by fusing multi-source sensing data with AI algorithms. Its core is to capture and analyze the patterns, trends, and dynamics of objects, events, and environments in the physical world across spatial and temporal dimensions, thereby providing a scientific basis for human decision-making(Deng, 2021).

The primary driving force behind the development of Spatial and Spatio-Temporal Intelligence stems from the ever-increasing human need to understand and shape the physical world. While traditional Geographic Information Systems (GIS) have achieved significant success in spatial management and analysis, they exhibit growing limitations in deeply understanding 3D space, performing complex spatio-temporal reasoning, and interacting with the physical world. The rapid advancement of AI, particularly in deep learning, reinforcement learning, and large model technologies, offers new pathways to overcome these bottlenecks. Concurrently, developments in remote sensing, perception technologies, mobile devices, and urban sensing networks have enabled the acquisition of large-scale, high-precision 3D spatial data and spatio-temporal dynamic data. This synergistic progress of "Data + Algorithms + Computing Power" further accelerates the advancement of Spatial and Spatio-Temporal Intelligence.

Despite notable progress in both academic and industrial research on Spatial and Spatio-

Temporal Intelligence, there remains a need for systematic review and clarification, as well as a lack of macro-level understanding of the overall technological ecosystem. This paper provides a systematic analysis of Spatial and Spatio-Temporal Intelligence from the perspectives of their conceptual distinctions, technological advancements, and future prospects, aiming to offer forward-looking references for continuous technological innovation and industrial practice in related fields.

2. CONCEPTUAL DISTINCTIONS

2.1 Concept and Essence of Spatial Intelligence

The early concept of Spatial Intelligence originated from research in cognitive science and psychology on human spatial cognition, initially describing “an individual’s ability to understand, manipulate, and reason about spatial relationships” encompassing core cognitive abilities such as spatial perception and spatial orientation(Hu, 2021). Advancements in AI technology have significantly expanded and deepened the connotation of Spatial Intelligence. In the context of AI, Spatial Intelligence is defined as “the ability of machines to perceive, reason, and act in three-dimensional space and time,” emphasizing its role as a core AI capability beyond linguistic intelligence, crucial for achieving Artificial General Intelligence (AGI). Spatial Intelligence aims to enable machines not only to “see” the world but also to understand it, learning knowledge and taking actions while observing three-dimensional space. Professor Fei-Fei Li explicitly states that Spatial Intelligence is the foundational basis for AI to understand the physical world, complementing language models—where language processes symbolic sequences, Spatial Intelligence handles geometry, physics, and dynamic interaction.

From a technical essence perspective, the core subject of Spatial Intelligence is “three-dimensional space.” Its technical logic primarily includes: (1) Acquiring spatial information of the environment through 3D perception technologies (e.g., stereo vision, laser scanning) and constructing/storing accurate 3D spatial models using representations like point clouds, meshes, or neural fields(Hu, 2021); (2) Utilizing spatial reasoning algorithms to understand spatial and topological logical relationships between objects, predict their motion transformations within space, and complete complex task planning based on this understanding; (3) Generating physically plausible and detailed 3D scenes and digital content based on deep understanding of the physical world, providing high-fidelity virtual environments for the learning and evolution of embodied agents; (4) Achieving natural interaction between machines and the physical world via Extended Reality (XR) technologies, or executing operations in the physical world through physical devices like robots(Wang, 2020).

2.2 Concept and Essence of Spatio-Temporal Intelligence

As a new interdisciplinary direction fusing AI and Geographic Information Science, Spatio-Temporal Intelligence seeks to answer a series of core questions related to spatio-temporal processes—such as "when, where, what, how it changes, why it changes, and how to respond"—by mining the spatio-temporal correlations, dependencies, and evolutionary patterns embedded within big data(Li, 2023). This aims to achieve profound understanding and accurate prediction of complex dynamic systems. Through deep integration of various technologies including GIS, remote sensing, the Internet of Things (IoT), and AI, Spatio-Temporal Intelligence aims to build a comprehensive technological system for dynamic perception, deep cognition, and intelligent service of natural environments and human activities, promoting full-chain intelligence from data collection and analysis to decision-making(Li, 2025).

From a technical essence perspective, the core characteristics of Spatio-Temporal Intelligence are mainly manifested in three aspects: (1) Dynamicity: The core subject of Spatio-Temporal Intelligence is spatio-temporal data that changes continuously or discretely over time (e.g., traffic flow, movement trajectories, urban evolution). (2) Multi-source Fusion: Spatio-Temporal Intelligence emphasizes the integration of massive, multi-source spatio-temporal data to form a comprehensive understanding of complex dynamic phenomena through data fusion and correlation analysis(Li, 2023). (3) Decision-Oriented: The ultimate goal of Spatio-Temporal Intelligence is to provide scientific basis and support for human decision-making in areas like urban governance, public safety, environmental protection, and economic operations through in-depth analysis, modeling, and prediction of spatio-temporal data(Li, 2021).

2.3 Comparison Between SI and STI

Spatial Intelligence and Spatio-Temporal Intelligence exhibit significant differences in core elements, research focus, technological foundations, core objectives, and application domains (Table 1). Their fundamental distinction lies in the different scientific questions they aim to address: the core of Spatial Intelligence is to answer "Where? What? How to interact and operate?", focusing on understanding and interacting with spatial structures; the core of Spatio-Temporal Intelligence is to answer "How does it change? Why does it change? How to respond and make decisions?", focusing on understanding and predicting spatio-temporal processes(Cheng, 2022).

Table 1. Comparison between Spatial Intelligence and Spatio-Temporal Intelligence.

Comparison Dimension	Spatial Intelligence	Spatio-Temporal Intelligence
Core Elements	Spatial structure, geometric properties, topological relations	Spatio-temporal objects, dynamic changes, events, evolutionary patterns
Research Focus	Spatial modeling, spatial reasoning, spatial generation, spatial interaction	Spatio-temporal modeling, spatio-temporal analysis, spatio-temporal prediction, decision support
Technological Foundation	Computer vision, computer graphics, virtual reality, robotics	GIS, remote sensing, IoT, spatio-temporal statistics, data mining
Core Objective	Understanding, interaction, and operation of spatial structures	Perception of dynamic processes, pattern mining, prediction, and decision-making
Typical Applications	Autonomous driving, metaverse, intelligent robots, etc.	Smart cities, public safety, low-altitude economy, etc.

As technology continuously evolves, Spatial Intelligence and Spatio-Temporal Intelligence are showing an increasingly close trend of integration. For example, in autonomous driving scenarios, an intelligent agent not only needs to utilize Spatial Intelligence to perceive and understand the surrounding 3D environmental structure in real-time for precise navigation and obstacle avoidance but also requires Spatio-Temporal Intelligence to predict the future trajectories of traffic participants (vehicles, pedestrians), understand the temporal changes of traffic signals, and accordingly plan and adjust its own action strategies(Yan, 2021). This integration enables agents to demonstrate advanced autonomy and adaptability in more complex and realistic scenarios.

3. KEY TECHNOLOGICAL ADVANCEMENTS

3.1 Key Technological Advancements in Spatial Intelligence

3.1.1 3D Perception

3D perception is the foundation for realizing Spatial Intelligence, aiming to enable machines to acquire and understand information about the geometric structure, object shapes, and spatial relationships of the 3D physical environment. In autonomous driving, accurate 3D perception is crucial for obstacle detection, drivable area recognition, and path planning. Traditional computer vision primarily deals with 2D images, while 3D perception recovers depth information and structure of 3D scenes from 2D inputs (e.g., cameras) or sensor data (e.g., LiDAR point clouds). In recent years, deep learning-based methods have made breakthrough progress in 3D perception. LiDAR technology can directly acquire high-precision point cloud data, providing rich geometric information for 3D scene understanding and has become one of the key technologies for 3D perception. Vision-based 3D perception methods, trained on large-scale depth datasets, can predict depth from single images. Multi-sensor fusion is also an effective approach to improve 3D perception accuracy; by fusing cameras, LiDAR, Inertial Measurement Units (IMU), and other sensors, it overcomes the limitations of single sensors, yielding more reliable environmental perception results.

3.1.2 Spatial Representation and Modeling

Effectively representing and modeling the acquired 3D environmental information is a core issue in Spatial Intelligence. Traditional 3D representation methods include polygonal meshes, voxels, and point clouds. Recently, deep learning-based implicit or hybrid 3D representation methods have achieved breakthroughs. Neural Radiance Fields (NeRF) learn the mapping between any 5D coordinate (3D spatial location and 2D viewing direction) and volume density/radiance color through training a Multilayer Perceptron (MLP), generating high-quality novel views and achieving implicit, continuous scene representation. NeRF and its subsequent improved models have shown remarkable effectiveness in 3D reconstruction and scene rendering. 3D Gaussian Splatting offers efficient rendering quality and training speed, enabling finer and more compact representation of 3D scenes and providing richer information for subsequent spatial reasoning and physical interaction.

3.1.3 Spatial Reasoning and Planning

Spatial reasoning and planning are the core cognitive abilities of Spatial Intelligence, enabling machines to perceive and understand 3D environments, engage in logical thinking, and formulate reasonable action strategies. Spatial reasoning involves understanding spatial relationships between objects, predicting object motion and changes, and having a preliminary grasp of physical laws. Spatial planning involves formulating a series of paths or actions based on environmental understanding and reasoning, including path planning, motion planning, and task planning. Deep learning and reinforcement learning play vital roles in spatial reasoning and planning; deep learning extracts high-level features from perceptual data for reasoning, while reinforcement learning acquires optimal planning strategies through interaction with the environment.

3.1.4 Spatial Creation and Content Generation

Spatial creation and content generation involve using AI technology to automatically or semi-automatically create new 3D scenes, objects, or virtual content, holding immense application value in fields like game development, film special effects, the metaverse, and agent training. Generative Adversarial Networks (GANs), Diffusion Models, and Transformer-based large models demonstrate strong potential in 3D content generation. For instance, some research can convert text descriptions into corresponding 3D models or scenes. Other methods can generate complete 3D objects or scenes from a single or a few 2D images. Professor Fei-Fei Li mentioned in a speech that her team and collaborators developed algorithms that can transform an input image into a 3D shape, even generating infinite explorable spaces from a single picture. Startups like GigaAI are also committed to advancing video generation to 4D world models, enabling AI models to understand, generate, and reason about 4D space, which is crucial for content creation in virtual environments and data generation/cognitive reasoning in physical space. These generative models need to understand not only 3D geometry but also object materials, lighting, textures, and the plausibility and semantic consistency of scenes (e.g., generated indoor scenes should follow common spatial layouts and object placement logic). Challenges lie in generating high-fidelity, highly diverse, and physically plausible 3D content, and providing user-friendly control methods for creators to precisely express their design intent. Additionally, the editability, composability of generated content, and integration with existing 3D content creation pipelines are important research directions.

3.1.5 Spatial Interaction and Embodied AI

Spatial interaction and Embodied AI are the ultimate manifestation of Spatial Intelligence, emphasizing that intelligent agents (e.g., robots, virtual characters) learn and perform tasks through interaction within 3D physical or virtual environments. Embodied AI posits that intelligence does not reside solely in abstract computation within a brain but emerges from the continuous interaction and perceptual feedback between the body and the environment. Spatial interaction technologies include Human-Computer Interaction (HCI), teleoperation, and natural interaction interfaces in Augmented Reality (AR) and Virtual Reality (VR). For example, users can interact more intuitively with virtual 3D objects or remote robots through gesture recognition, voice control, and haptic feedback. In robotics, Embodied AI means robots need not only to perceive and understand the environment but also to move safely and efficiently, manipulate objects, and collaborate with humans or other robots. This requires robots to possess fine motor control, dexterous manipulation skills, and the ability to predict interaction consequences. Reinforcement learning, imitation learning, and learning from demonstration are common methods for training embodied agents. Through extensive interaction with the environment or human demonstrations, agents can learn complex skills and behavioral strategies. Professor Fei-Fei Li notes that the impulse to act is inherent to all spatially intelligent beings; it links perception with action, propelling AI's virtuous cycle from "seeing the world" to "understanding the world and taking action". Challenges include achieving robust and safe interaction in dynamic, uncertain, and complex environments; enabling agents to learn quickly from limited experience and generalize to new scenes and tasks; and realizing natural and efficient human-machine collaboration. For instance, service robots need to understand ambiguous human instructions, adapt to diverse home environments, and safely coexist and work with humans. This necessitates the synergistic development of all subfields of Spatial Intelligence, with tight integration from perception and modeling to reasoning, planning, and interaction.

3.2 Key Technological Advancements in Spatio-Temporal Intelligence

As an emerging interdisciplinary field, Spatio-Temporal Intelligence aims to integrate spatio-temporal data with intelligent computing methods to reveal the dynamic evolution patterns of objects, events, and environments in the physical world and provide a scientific basis for decision-making(Deng, 2021). With the rapid development of IoT, big data, AI, and other technologies, Spatio-Temporal Intelligence has achieved significant advancements in processing multimodal spatio-temporal big data, spatio-temporal data mining, spatio-temporal prediction models, spatio-temporal large models, and simulation/inference. These advancements provide powerful analytical tools and decision support for numerous fields such as smart cities, intelligent transportation, environmental monitoring, and public safety. The core of Spatio-Temporal Intelligence lies in extracting valuable patterns and knowledge from massive, dynamic, multi-source spatio-temporal data to achieve deep insight and accurate prediction of complex systems. This section will elaborate on the latest developments in these key technological directions of Spatio-Temporal Intelligence, discussing core technical methods, challenges, and future application prospects, aiming to showcase the technological vitality and broad application space of this field.

3.2.1 Multimodal Spatio-Temporal Big Data

Multimodal spatio-temporal big data forms the cornerstone of Spatio-Temporal Intelligence development. Its core characteristics are "multi-source heterogeneity," "spatio-temporal correlation," and "dynamic evolution"(Qin, 2022). Multi-source heterogeneity refers to spatio-temporal data originating from diverse sensor platforms (e.g., satellite remote sensing, UAV aerial photography, ground surveillance cameras, GPS trajectories, social media check-ins, weather stations), which exhibit significant differences in structure, resolution, spatio-temporal granularity, and semantic meaning(Run, 2025). For example, remote sensing imagery provides large-scale land cover information but may lack detail and real-time currency, whereas GPS trajectories accurately reflect the real-time location of moving objects but have limited coverage. Spatio-temporal correlation emphasizes the inherent connections between different sources and types of spatio-temporal data across spatial location and time dimensions. For instance, changes in urban traffic flow are closely related to population activity, weather conditions, and public events in specific areas. Dynamic evolution refers to the continuous change of spatio-temporal data over time, reflecting the dynamic processes of geographical phenomena and human activities. Effectively fusing and managing these multimodal spatio-temporal big data is the primary task of Spatio-Temporal Intelligence. This includes data cleaning, alignment, standardization, spatio-temporal index construction, and data quality assessment. In recent years, knowledge graph technology has been used to organize and manage multi-source heterogeneous spatio-temporal data, enabling deeper-level data fusion and correlation analysis by constructing knowledge networks of entities, relationships, and events[14]. Furthermore, the development of spatio-temporal big data platforms provides foundational support for Spatio-Temporal Intelligence applications; these platforms typically offer functionalities for data ingestion, storage, computation, visualization, and service publishing. Challenges lie in designing efficient data models and algorithms to handle the ever-increasing volume of spatio-temporal big data, bridging the semantic gap between different data modalities, and ensuring data timeliness and accuracy to provide high-quality input for subsequent intelligent analysis and decision-making(Qin, 2022).

3.2.2 Spatio-Temporal Data Mining

Spatio-temporal data mining aims to discover valuable patterns, trends, anomalies, and association rules from large-scale spatio-temporal databases. It serves as a critical bridge connecting spatio-temporal data with intelligent analysis. When applying traditional data mining techniques (e.g., clustering, classification, association rule mining) to spatio-temporal data, the spatio-temporal characteristics of the data must be fully considered. For instance, spatio-temporal clustering must consider not only attribute similarity of data objects but also their spatial proximity and temporal adjacency. Common spatio-temporal data mining tasks include:

(1) Spatio-Temporal Pattern Mining: Discovering frequently occurring spatio-temporal sequences or subsequences, e.g., periodic patterns of urban traffic congestion or high-incidence patterns of specific crime types in certain areas and times.

(2) Spatio-Temporal Anomaly Detection: Identifying objects or events deviating from normal behavior in spatio-temporal distribution, e.g., sudden traffic congestion points, anomalous meteorological phenomena, or early signals of disease outbreaks.

(3) Trajectory Data Analysis: Clustering, classifying, pattern recognition, and anomalous behavior detection for moving object trajectories (e.g., people, vehicles, animals). For example, analyzing urban resident commuting patterns or identifying suspicious vehicle trajectories.

(4) Spatio-Temporal Association Rule Mining: Discovering associative relationships between different spatio-temporal events, e.g., the relationship between rainfall in a specific area and river water level changes, or the association between large event hosting and increased traffic flow/commercial activity in surrounding areas.

Recently, deep learning techniques, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) and their variants (e.g., LSTM, GRU), have been widely applied to spatio-temporal data mining tasks. CNNs are adept at extracting spatial features, while RNNs effectively handle time series data. Graph Neural Networks (GNN) also show advantages in processing data with irregular spatial topological relationships (e.g., traffic networks, social networks)(Cheng, 2022). For example, Spatio-Temporal Graph Neural Networks (STGNN) can simultaneously model the topological structure of road networks and the spatio-temporal dynamics of traffic flow for traffic prediction. Challenges include designing model architectures that effectively capture spatio-temporal dependencies, handling uncertainty, noise, and missing values in data, and improving algorithm scalability and efficiency to meet the demands of mining massive spatio-temporal data(Qin, 2022).

3.2.3 Spatio-Temporal Prediction Models

Spatio-temporal prediction is a core task of Spatio-Temporal Intelligence, aiming to forecast the state or attribute value of a specific spatial region at a future time point or period based on historical and current spatio-temporal data. This has wide applications in traffic flow prediction, air quality forecasting, population mobility prediction, meteorological disaster warning, energy demand forecasting, etc. Spatio-temporal prediction models need to capture both spatial dependencies (where a region's value is influenced by its neighboring regions) and temporal dependencies (where the current value is influenced by past values). Traditional spatio-temporal prediction methods include time series models (e.g., ARIMA), spatial statistical models (e.g., geographically weighted regression), and their simple combinations. In recent years, deep learning-based spatio-temporal prediction models have become a research hotspot, demonstrating performance significantly superior to traditional methods. Representative models include:

(1) CNN-based Models: Using convolutional layers to extract spatial features, combined with RNNs or LSTMs to handle the temporal dimension.

(2) RNN/LSTM-based Models: Treating data from spatial regions as time series input to RNNs/LSTMs, or incorporating spatial attention mechanisms.

(3) Graph Neural Network (GNN)-based Models: Particularly suitable for spatio-temporal data with graph structures (e.g., traffic networks, sensor networks), aggregating spatial information from neighboring nodes through graph convolution operations and combining them with time series models for prediction.

(4) Transformer-based Models: Leveraging self-attention mechanisms to capture long-range dependencies in spatio-temporal data, showing excellent performance in many prediction tasks. For example, Spatio-Temporal Transformers can treat spatio-temporal data as sequences or patches, dynamically learning correlations between different spatio-temporal locations via attention mechanisms.

(5) Multimodal Fusion Prediction Models: Fusing spatio-temporal data from different sources (e.g., traffic flow, weather, events, POI data) for joint prediction, often yielding more accurate results.

Challenges include effectively modeling complex spatio-temporal nonlinear relationships, handling the multi-scale characteristics of data (local and global patterns), improving model generalization ability and robustness to unseen situations, and providing interpretability of prediction results so that decision-makers can understand and trust the model's output.

3.2.4 Spatio-Temporal Large Models

Following the remarkable success of Large Language Models (LLMs) in Natural Language Processing (NLP), building "Spatio-Temporal Large Models" or "Spatial Foundation Models" for spatio-temporal data has become an important trend in Spatio-Temporal Intelligence development(Yang, 2023). These models aim to leverage massive, multimodal spatio-temporal data for pre-training to learn general spatio-temporal representations and knowledge. They can then be adapted to various downstream spatio-temporal intelligence tasks via fine-tuning or prompt learning, thereby shifting from "task-specific, modality-specific" paradigms towards a "general paradigm". The core idea of spatio-temporal large models is to borrow the self-supervised learning and transfer learning capabilities of powerful architectures like Transformers, learning from large-scale unlabeled or weakly labeled spatio-temporal data. For example, some research attempts to unify data from multiple modalities such as remote sensing imagery, geographic grid data, trajectory data, and textual descriptions within a single large model for pre-training. The AllSpark model proposes using Language as a Reference Framework (LaRF), mapping ten different modalities of spatio-temporal data (including RGB, multispectral, SAR, hyperspectral, graphs, trajectories, point clouds, etc.) into a linguistic feature space, achieving unified multimodal modeling and demonstrating strong generalization capabilities in tasks like few-shot classification. These pre-trained models can serve as powerful foundations for various downstream spatio-temporal tasks (e.g., land use classification, change detection, traffic prediction, disaster assessment). By fine-tuning with a small amount of labeled data, they can quickly adapt to new tasks and achieve excellent performance. The development of spatio-temporal large models faces numerous challenges, including: designing effective model architectures to handle heterogeneous spatio-temporal modalities; addressing distribution shifts and domain adaptation issues across different spatio-temporal data domains; reducing model training and deployment costs; and ensuring model interpretability and reliability, especially in decision-making applications involving public safety and critical infrastructure. Furthermore, constructing high-quality, large-scale, and diverse spatio-temporal pre-training datasets is also key to advancing spatio-temporal large models.

3.2.5 Simulation and Inference

Simulation and inference represent advanced application forms of Spatio-Temporal Intelligence, utilizing constructed spatio-temporal models and knowledge to perform dynamic simulations of future behaviors of complex systems, assess the impacts of different strategies, and explore "what-if" scenarios(Shen, 2023). This holds significant value in fields like urban planning, emergency management, traffic management, environmental protection, and military simulation. For instance, in urban planning, simulating the long-term impacts of different land-use schemes on traffic flow, air quality, and energy consumption can assist decision-makers in selecting optimal plans. In emergency management, simulating the propagation paths and impact ranges of disasters like floods, fires, or pandemics can provide a scientific basis for resource allocation and evacuation strategy formulation. Digital Twin technology provides a powerful platform for spatio-temporal simulation and inference. Digital Twins construct high-fidelity virtual mappings of physical entities or systems, capable of real-time synchronization with physical world data and performing simulation, analysis, and optimization in virtual space. Combined with AI, Digital Twins can not only perform historical review and current state analysis but also predict and infer future states. Spatio-temporal simulation and inference typically involve complex computational models, including physical models, statistical models, machine learning models, and Agent-Based Modeling (ABM). For example, traffic simulation can model the driving behavior of individual vehicles and their interactions to evaluate the effects of different traffic control strategies. Challenges include constructing high-fidelity, high-efficiency spatio-temporal simulation models, effectively integrating multi-source heterogeneous data and different types of models (physics-driven and data-driven), and handling uncertainty and randomness during simulation. Furthermore, the visualization and interactive analysis of simulation results are key to enhancing the practicality of simulation and inference systems.

4. CONCLUSION AND OUTLOOK

As frontier branches of AI, Spatial Intelligence and Spatio-Temporal Intelligence are advancing at an unprecedented pace, empowering machines with capabilities to perceive, understand, reason about, and interact with the physical world. Spatial Intelligence focuses on endowing machines with a deep understanding of the static structures and dynamic interactions of the 3D physical world. Its key technological advancements span multiple levels, from 3D perception and representation modeling to spatial reasoning, content generation, and embodied AI interaction. Breakthroughs in these technologies lay a solid foundation for the development of fields like robotics navigation, autonomous driving, and the metaverse, marking AI's transition from processing 2D information to truly acting in the 3D world. Spatio-Temporal Intelligence, on the other hand, dedicates itself to fusing multi-source heterogeneous data under a unified spatio-temporal framework, mining dynamic evolution patterns, and achieving precise prediction and intelligent decision-making. Its progress in handling multimodal spatio-temporal big data, spatio-temporal data mining, prediction models, large model construction, and simulation/inference provides powerful analytical tools for applications like smart cities, intelligent transportation, and environmental monitoring, assisting human society in moving towards a data-driven intelligent era.

Although Spatial Intelligence and Spatio-Temporal Intelligence have distinct research focuses and technical pathways—the former more concerned with "what it is" and "how to operate," the latter with "how it changes" and "future trends"—the boundary between them is not rigid but shows an increasingly close trend of integration. In complex real-world application scenarios, such as advanced autonomous driving, intelligent agents require both Spatial Intelligence to precisely perceive and understand the rapidly changing 3D environment and Spatio-Temporal Intelligence to predict the dynamic behaviors of other participants and plan their own actions. This fusion, characterized by "taking space as the substance and spatio-temporal as the function," is key for future AI systems to achieve higher levels of autonomy and adaptability. The concept of a "World Model" proposed by Professor Fei-Fei Li embodies this fusion, requiring AI to build an internal model that understands 3D structure and simulates its dynamic evolution, which relies on the synergistic development of Spatial and Spatio-Temporal Intelligence.

Looking ahead, the development of Spatial and Spatio-Temporal Intelligence still faces numerous challenges. At the data level, acquiring high-quality, large-scale, annotated 3D and spatio-temporal data, especially for complex scenes and rare events, remains a bottleneck. At the model level, improving model interpretability, robustness, generalization capability, and computational efficiency are core scientific problems requiring urgent solutions. For example, purely data-driven models still have shortcomings in representation capability, interpretability, and generalization, while physics-informed machine learning methods offer new avenues to address these issues(Zhang, 2023). Additionally, issues of privacy protection, algorithmic bias, safety, and ethics are becoming increasingly prominent with the widespread application of these technologies, necessitating interdisciplinary collaboration.

Despite the challenges, the development prospects for Spatial and Spatio-Temporal Intelligence remain vast. With continuous algorithmic innovation, sustained growth in computing power, and increasing data abundance, we have reason to believe that AI systems will possess stronger spatial cognition and spatio-temporal understanding capabilities in the near future. They will not merely be tools for processing information but become intelligent partners for humans in perceiving, transforming, and creating within the physical world, profoundly impacting production and lifestyles, and driving society towards a more intelligent and sustainable future. The transition from "Internet intelligence" to "physical world intelligence" hinges on the deep integration and innovative breakthroughs in Spatial and Spatio-Temporal Intelligence.

REFERENCES

Cheng, T., Zhang, Y., and Haworth, J. 2022. Network and Graph-Based Space Time AI: conception, method and Applications. Acta Geodaetica et Cartographica Sinica.

Deng, M., Huang, J. F., Huang, Y., Liu, X. P., Shi, Y., Li, L. L., Wu, Z. F., Xin, Q. C., Sun, Y., Liu, H. M., Zhang, X. C., Li, C. K., Cao, K. B., and Yang, W. T. 2021. Key Technologies and Applications of Geospatial Information Intelligent Processing in New Urbanization Construction. Construction Science and Technology: 92-95.

Guowei, Z., Chunlan, G., and Biao, L. 2024. The Application of Spatio-Temporal Intelligence in Urban Sensing. 2024 Cross Strait Radio Science and Wireless Technology Conference (CSRSWTC): 1-3.

Hu, Y., Xu, K., Qin, L., Yin, Q. J., and Zha, Y. B. 2021. Spatio-Temporal Memory Models in Spatial Cognitive Behavioral Modeling. Journal of System Simulation, 33(2): 249-261.

Li, D. R., Wang, M., Xiao, J., Li, M., Di, K. C., Li, X., and Luo, B. 2025. On Ubiquitous Spatio-

Temporal Intelligence. Journal of Remote Sensing, 29(6): 1388-1398.
Li, Q. Q., Tao, L. F., Chen, X. P., Huang, B., Wang, P., and Pan, M. M. 2023. Key Technologies of Pan-Spatial Intelligent GIS Platform and Its Applications. wf, 48(1): 2023.
Li, R. S., Cheng, Y., Li, Q. J., Zhu, Q., Zhang, Y. J., Li, Y. S., Liu, Y., Tang, Y., Liu, D. P., Yang, S. Z., and Li, H. F. 2025. All Spark: A Multimodal Spatio-Temporal General Intelligence Model with Ten Modalities Via Language As a Reference Framework. IEEE Transactions on Geoscience and Remote Sensing, 63: 2025.
Li, Q. Q., Tao, L. F., Chen, X. P., Huang, B., Wang, P., and Pan, M. M. 2023. Key Technologies of Pan-Spatial Intelligent GIS Platform and Its Applications. wf, 48(1): 2023.
Monjoree, U., and Yan, W. 2025. Assessing AI Models' Spatial Visualization in PSVT:R and Augmented Reality: Towards Enhancing AI's Spatial Intelligence. 2025 IEEE Conference on Artificial Intelligence (CAI): 1-8.
Qin, K., Xu, K., Wu, T., Xu, M., Huang, J., Chen, Y. X., Feng, X., Zhao, P. X., Wang, Y. L., Zhang, Y., Xu, Y. Q., and Bi, Q. 2022. Intelligent Spatial Information Processing and Spatio-temporal Big Data Analysis Exploration. Geospatial Information, 20(12): 2022.
Run, S., Cheng, Y., Qiujun, L., Qing, Z., Yongjun, Z., Yan Sheng, L., Yu, L., Yong, T., Dapeng, L., Shizhong, Y., and Haifeng, L. 2025. All Spark: A Multimodal Spatio-Temporal General Intelligence Model with Ten Modalities Via Language As a Reference Framework. IEEE Transactions on Geoscience and Remote Sensing, 63.
Shen, Y., and Yu Meng, L. 2023. Metaverse and Big Data: Data Insight and Value Connection in Spatio-Temporal Intelligence. Big Data Research, 9(1): 103-110.
Uttamasha, M., and Wei, Y. 2025. Assessing AI Models' Spatial Visualization in PSVT:R and Augmented Reality: Towards Enhancing AI's Spatial Intelligence. 2025 IEEE Conference on Artificial Intelligence (CAI): 1-8.
Wang, Y., Yan, T., and Wang, L. 2022. Development Situation and Trend of Space Intelligent Navigation Technology. Aerospace Control and Application, 48(5): 2022.
Yang, B. S., Chen, Y. P., and Zou, Q. 2023. Opportunities and Challenges of Spatiotemporal Information Intelligent Processing of Surveying and Mapping in the Era of Large Models. Geomatics and Information Science of Wuhan University, 48(11): 2023.
Yan, B., Jiang, J. Y., Lang, L. Y., Xiang, H., and Wang, W. Q. 2021. Research on Key Technologies of Spatial-Temporal Intelligent Application for Energy Internet. Electric Power Information and Communication Technology: 24-30.
Zhang, T., Liu, R. Y., Wang, P. X., Gao, C. L., Liu, J., and Wang, W. S. 2023. Physics-informed Machine Learning and Its Research Prospects in GeoAI. Journal of Geo-information Science, 25(7).

BIOGRAPHICAL NOTES

Liu Hao is director of the innovation center of Chongqing Academy of Surveying and Mapping. His research is on Spatio-Temporal intelligence, City information modeling, and Virtual Geographic Environments. He has led the completion of Chongqing's City Information

Modeling (CIM) basic platform, explored spatial cognitive mechanisms in virtual geographic environments, and promoted the interdisciplinary application of geographic information technology with artificial intelligence and the low-altitude economy.

CONTACTS

Professor Hao Liu
Chongqing Academy of Surveying and Mapping
No.6 Qingzhu East Road, Dazhulin Street, Liangjiang New District
Chongqing
China
Tel. +86 15213356615
Email: 37933748@qq.com