Novelty Assessment Report

Paper: What Lies Beyond the View? Actively Constructing Spatial Beliefs in Foundation Models
PDF URL: https://openreview.net/pdf?id=8iPwqr6Adk
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-08

Abstract

Current foundation models can answer spatial reasoning questions about a given image or text, yet they lack the fundamental ability to build a genuine spatial understanding of an environment through active exploration. This reflects a critical blind spot in prevailing evaluation protocols, which predominantly test passive reasoning on curated data rather than the active construction of knowledge under uncertainty. To address this, we introduce Theory of Space (ToS), a new framework analogous to the Theory of Mind. While Theory of Mind concerns an agent's ability to model the hidden mental states of others, ToS concerns its ability to construct, update, and utilize an internal belief about the unobserved structure of its spatial environment from local, incomplete observations. We implement ToS with a comprehensive benchmark featuring both text-based and visual environments. Instead of performing specific tasks in such environments, the primary objective is to build a complete and accurate spatial belief through curiosity-driven exploration. A core innovation of our framework is the direct probing of this internal belief: we prompt models to explicitly present their cognitive map at each step, allowing us to measure not only task performance but also the quality, consistency, and evolution of the underlying spatial model itself. By evaluating state-of-the-art models as both active explorers and passive reasoners (using logs from scripted proxy agents), we disentangle exploration strategy from reasoning ability. Our analysis reveals common failure modes in spatial belief management, such as egomotion update errors and the inability to maintain a globally consistent map. The ToS framework provides the concepts and tools necessary to evaluate and build agents with more robust, human-like spatial intelligence.

Core Task Landscape

This paper addresses: **Actively Constructing Spatial Beliefs Through Exploration in Foundation Models**

A total of **48 papers** were analyzed and organized into a taxonomy with **14 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Spatial Belief Construction and Cognitive Mapping**
- **Exploration Strategies and Active Perception**
- **Embodied Navigation and Object Search**
- **Reinforcement Learning and Interactive Decision-Making**
- **Benchmarks and Evaluation Frameworks**
- **Foundation Models for Robotics and Embodied AI**
- **Domain-Specific and Applied Spatial Systems**
- **Cognitive and Learning Sciences Perspectives**

Complete Taxonomy Tree

- Actively Constructing Spatial Beliefs Through Exploration in Foundation Models Survey Taxonomy
- Spatial Belief Construction and Cognitive Mapping
  - Theory-Driven Spatial Belief Frameworks ★ (3 papers)
  - [0] What Lies Beyond the View? Actively Constructing Spatial Beliefs in Foundation Models (Anon et al., 2026) View paper
  - [4] Exploring Spatial Schema Intuitions in Large Language and Vision Models (Wicke, 2024) View paper
  - [13] Assessing adaptive world models in machines with novel games (Ying, 2025) View paper
  - Memory-Augmented Spatial Reasoning (3 papers)
  - [18] STMA: A spatio-temporal memory agent for long-horizon embodied task planning (Zhao Yiming, 2025) View paper
  - [19] BeliefMapNav: 3D Voxel-Based Belief Map for Zero-Shot Object Navigation (Zhou Zibo, 2025) View paper
  - [30] Vision to Geometry: 3D Spatial Memory for Sequential Embodied MLLM Reasoning and Exploration (Zhongyi Cai, 2025) View paper
  - Spatial Representation Learning (3 papers)
  - [1] SpatialVLA: Exploring Spatial Representations for Visual-Language-Action Model (Qu, 2025) View paper
  - [17] Spatial Representations for Learning Robotic Mobile Manipulation (Wu, 2025) View paper
  - [26] Actial: Activate Spatial Reasoning Ability of Multimodal Large Language Models (Zhan XiaoYu, 2025) View paper
- Exploration Strategies and Active Perception
  - Foundation Model-Guided Exploration (4 papers)
  - [15] Sensei: Semantic exploration guided by foundation models to learn versatile world models (Sancaktar, 2025) View paper
  - [20] Semantic Exploration from Language Abstractions and Pretrained Representations (Tam, 2022) View paper
  - [34] Foundation Models for Semantic Novelty in Reinforcement Learning (Gupta Tarun, 2022) View paper
  - [35] PIGEON: VLM-Driven Object Navigation via Points of Interest Selection (Cheng Peng, 2025) View paper
  - Uncertainty-Driven and Information-Theoretic Exploration (3 papers)
  - [6] Explore until confident: Efficient exploration for embodied question answering (Allen Z. Ren, 2024) View paper

    - [7] Can foundation models actively gather information in interactive environments to test hypotheses? (Ke, 2024) View paper
    - [10] Seeing is Believing: Belief-Space Planning with Foundation Models as Uncertainty Estimators (Zhao, 2025) View paper
    - Active Perception and Adaptive Sensing (3 papers)
    - [21] Seal: Self-supervised embodied active learning using exploration and 3d consistency (Chaplot, 2021) View paper
    - [37] SpatialReasoner: Active Perception for Large-Scale 3D Scene Understanding (Hongpei Zheng, 2025) View paper
    - [40] Active-O3: Empowering Multimodal Large Language Models with Active Perception via GRPO (Zhu Muzhi, 2025) View paper
- Embodied Navigation and Object Search
    - Zero-Shot Object Navigation (4 papers)
    - [2] Voronav: Voronoi-based zero-shot object navigation with large language model (Pengying Wu, 2024) View paper
    - [8] SSR-ZSON: Zero-Shot Object Navigation via Spatial-Semantic Relations within a Hierarchical Exploration Framework (Meng Xiang-yi, 2025) View paper
    - [9] LLM-Guided Zero-Shot Visual Object Navigation with Building Semantic Map (Jin Shi, 2025) View paper
    - [39] Multi-Floor Zero-Shot Object Navigation Policy (Lingfeng Zhang, 2024) View paper
    - Vision-Language Model-Based Navigation (3 papers)
    - [12] Wmnav: Integrating vision-language models into world models for object goal navigation (Guo, 2025) View paper
    - [16] Towards Autonomous UAV Visual Object Search in City Space: Benchmark and Agentic Methodology (Ji, 2025) View paper
    - [38] ImagineNav++: Prompting Vision-Language Models as Embodied Navigator through Scene Imagination (Zhao Xin-xin, 2025) View paper
    - Hierarchical and Multi-Level Navigation (4 papers)
    - [27] IndustryNav: Exploring Spatial Reasoning of Embodied Agents in Dynamic Industrial Navigation (Yifan Li, 2025) View paper
    - [29] Stairway to Autonomy: Hierarchical Decision-Making for LLM-Guided Planning, Bandit-Driven Exploration, and Multi-Agent Navigation (Nayak, 2025) View paper
    - [36] From Prompts to Paths: Large Language Models for Zero-Shot Planning in Unmanned Ground Vehicle Simulation (K Olaiya, 2025) View paper
    - [42] From Prompts to Paths: Large Language Models for Zero-Shot Planning and Simulation (Kelvin Olaiya, 2025) View paper
- Reinforcement Learning and Interactive Decision-Making
    - RL with Foundation Model Integration (3 papers)
    - [3] Embodied-r: Collaborative framework for activating embodied spatial reasoning in foundation models via reinforcement learning (Wang Zi-you, 2025) View paper
    - [5] Pre-trained language models for interactive decision-making (Li Shuang, 2022) View paper
    - [48] Language-Guided Semantic Affordance Exploration for Efficient Reinforcement Learning (Ma, n.d.) View paper
- Benchmarks and Evaluation Frameworks
    - Task-Level Planning and Reasoning Benchmarks (3 papers)
    - [11] Et-plan-bench: Embodied task-level planning benchmark towards spatial-temporal cognition with foundation models (Zhang Ling-Feng, 2024) View paper
    - [32] ASTRA: Autonomous Spatial-Temporal Red-teaming for AI Software Assistants (Xu, 2025) View paper
    - [33] Thinking on Maps: How Foundation Model Agents Explore, Remember, and Reason Map Environments (Zhiwei Wei, 2025) View paper
- Foundation Models for Robotics and Embodied AI (2 papers)
    - [14] Foundation Models for Robotic Tasks: Survey, Challenges and Future Directions (M. A. Viraj J. Muthugala, 2025) View paper
    - [41] The Foundation Model Path to Open-World Robots (Shah, 2024) View paper
- Domain-Specific and Applied Spatial Systems (4 papers)
    - [22] Research on an autonomous exploration strategy for hydroelectric generator inspection robots driven by large language models (Y Dong, 2025) View paper
    - [23] Mesoscale explorer: Visual exploration of large-scale molecular models. (Alexander Rose, 2024) View paper
    - [46] Fast Mining of Spatial Frequent Wordset from Social Database (Lee Yongmi, 2019) View paper
    - [47] Supplementary Materials for â□□GeoExplorer: Active Geo-localization with Curiosity-Driven Explorationâ□□ (L Mi, n.d.) View paper
- Cognitive and Learning Sciences Perspectives (7 papers)
    - [24] Exploring a cognitive basis for learning spatial relationships with augmented reality (Brett E. Shelton, 2004) View paper
    - [25] Student-centered learning environments: Foundations, assumptions and design (SM Land, 2012) View paper
    - [28] A dual-space model of iteratively deepening exploratory learning (J Ohn R Ieman, 1996) View paper
    - [31] Game object model version II: a theoretical framework for educational game development (Alan Amory, 2007) View paper
    - [43] Differential effects of aging on spatial learning through exploratory navigation and map reading. (Naohide Yamamoto, 2012) View paper
    - [44] Learning-dependent evolution of spatial representations in large-scale virtual environments. (Michael J. Starrett, 2019) View paper
    - [45] Cross-Embodied Cognitive Morphologies. Proceedings 2022, 81, 10 (VallverdÃº, 2022) View paper

Narrative

Core task: Actively constructing spatial beliefs through exploration in foundation models. The field encompasses a diverse set of approaches that address how agents build, maintain, and reason about spatial knowledge in embodied settings. At the highest level, the taxonomy distinguishes between works focused on spatial belief construction and cognitive mapping (which develop explicit or implicit representations of environments), exploration strategies and active perception (which determine how agents gather spatial information), embodied navigation and object search (which apply spatial reasoning to goal-directed tasks), reinforcement learning and interactive decision-making (which learn policies through environmental interaction), benchmarks and evaluation frameworks (which standardize assessment), foundation models for robotics and embodied AI (which leverage large pretrained models for spatial tasks), domain-specific and applied spatial systems (which target particular application areas), and cognitive and learning sciences perspectives (which draw on human spatial cognition research). Representative works such as SpatialVLA[1] and Embodied-r[3] illustrate how foundation models are being adapted to spatial reasoning, while Voronav[2] and SSR-ZSON[8] exemplify navigation-centric approaches.

A particularly active line of work centers on how agents should balance exploration with exploitation when spatial knowledge is incomplete or uncertain, as seen in Explore Until Confident[6] and Adaptive World Models[13]. Another contrasting theme involves whether to rely on end-to-end learned representations versus structured symbolic or schema-based spatial models, a tension visible across Foundation Models Hypothesis Testing[7] and Spatial Schema Intuitions[4]. Beyond the View[0] sits within the theory-driven spatial belief frameworks cluster, emphasizing principled mechanisms for belief updating during exploration. Compared to Spatial

Schema Intuitions[4], which examines cognitive primitives for spatial understanding, and Adaptive World Models[13], which focuses on dynamic model adaptation, Beyond the View[0] appears to prioritize the active construction process itself—how agents iteratively refine spatial hypotheses by strategically choosing where to look next. This positioning highlights ongoing questions about the interplay between model architecture, exploration policy, and the granularity of spatial representations needed for robust embodied intelligence.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Exploring Spatial Schema Intuitions in Large Language and Vision Models

Authors: Wicke, Philipp | **Year/Venue**: 2024 | **URL**: View paper

Abstract

N/A

Relationship Analysis

Both papers belong to the Theory-Driven Spatial Belief Frameworks category, focusing on evaluating spatial reasoning capabilities in foundation models through structured benchmarks. The candidate paper explores spatial schema intuitions in LLMs and VLMs, likely examining how these models understand fundamental spatial concepts and relationships, which overlaps with the original paper's focus on spatial belief construction and cognitive mapping. However, the original paper introduces the Theory of Space (ToS) framework emphasizing active exploration and dynamic belief updating through agent-environment interaction, while the candidate paper appears to focus on probing existing spatial schema knowledge without the active exploration component.

---

2. Assessing adaptive world models in machines with novel games

Authors: Ying, Lance, Collins, Katherine M., Lance Ying, et al. (38 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Human intelligence exhibits a remarkable capacity for rapid adaptation and effective problem-solving in novel and unfamiliar contexts. We argue that this profound adaptability is fundamentally linked to the efficient construction and refinement of internal representations of the environment, commonly referred to as world models, and we refer to this adaptation mechanism as world model induction. However, current understanding and evaluation of world models in artificial intelligence (AI) remains...

Relationship Analysis

Both papers belong to the Theory-Driven Spatial Belief Frameworks category, proposing theoretical frameworks for evaluating spatial belief construction capabilities in foundation models. While the original paper introduces Theory of Space (ToS) to evaluate active spatial belief construction through exploration in multi-room environments with direct cognitive map probing, the candidate paper proposes a broader framework for assessing adaptive world models through novel games, focusing on hierarchical Bayesian world model induction across diverse game structures rather than specifically spatial environments. The key difference is that the original paper targets spatial belief construction with explicit cognitive mapping in navigation contexts, whereas the candidate paper addresses general world model adaptation across varied game domains with latent mechanics discovery.

Contributions Analysis

Overall novelty summary. The paper introduces Theory of Space (ToS), a framework for evaluating how foundation models actively construct spatial beliefs through exploration, analogous to Theory of Mind for mental state modeling. It resides in the 'Theory-Driven Spatial Belief Frameworks' leaf, which contains only three papers total, including this work and two siblings. This represents a notably sparse research direction within the broader taxonomy of 48 papers across the field, suggesting the paper targets a relatively underexplored conceptual niche focused on principled frameworks for spatial belief construction rather than task-specific navigation or perception methods.

The taxonomy reveals that neighboring research directions are substantially more populated: 'Memory-Augmented Spatial Reasoning' (3 papers), 'Foundation Model-Guided Exploration' (4 papers), and 'Zero-Shot Object Navigation' (4 papers) all address related but distinct aspects of spatial intelligence. The sibling papers in the same leaf—Spatial Schema Intuitions and Adaptive World Models—examine cognitive primitives and dynamic model adaptation respectively, whereas ToS emphasizes the active exploration process itself. The taxonomy's scope and exclude notes clarify that ToS belongs here because it proposes a theoretical evaluation framework rather than applying existing methods to specific tasks, distinguishing it from application-oriented categories.

Among 30 candidates examined through semantic search, the contribution-level analysis shows varied novelty profiles. The ToS framework itself (10 candidates examined, 0 refutable) and the comprehensive benchmark (10 candidates examined, 0 refutable) appear to have limited direct prior work within the search scope. However, the direct probing mechanism for internal spatial beliefs (10 candidates examined, 1 refutable) shows at least one candidate providing overlapping prior work. This suggests that while the overarching framework may be relatively novel, specific technical components like belief probing have some precedent in the examined literature, though the limited search scale means substantial related work could exist beyond these 30 candidates.

Given the sparse taxonomy leaf and limited search scope, the work appears to occupy a conceptual space with relatively few direct competitors among the examined papers. The framework's emphasis on curiosity-driven exploration and explicit cognitive map probing distinguishes it from task-oriented navigation benchmarks, though the analysis acknowledges it covers only top-30 semantic matches and does not claim exhaustive coverage of all potentially relevant spatial reasoning literature.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Theory of Space (ToS) framework

Description: The authors propose ToS as a conceptual framework for evaluating an agent's ability to actively construct, update, and utilize an internal spatial belief from partial observations. Unlike Theory of Mind, which models hidden mental states of others, ToS models the uncertain, unobserved structure of physical space through curiosity-driven exploration.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Seeing is Believing: Belief-Space Planning with Foundation Models as Uncertainty Estimators

URL: View paper

Brief Assessment

Seeing is Believing[10] focuses on belief-space planning for mobile manipulation under partial observability using three-valued predicates and information-gathering actions. The original paper's ToS framework concerns curiosity-driven exploration to construct spatial beliefs about physical environments, which is a distinct conceptual contribution not addressed by the candidate's technical approach to uncertainty estimation in robotic tasks.

---

2. Cognitive mapping and planning for visual navigation

URL: View paper

Brief Assessment

Cognitive Mapping Planning[71] focuses on learning navigation policies through joint mapping and planning in simulated environments, not on evaluating an agent's ability to construct spatial beliefs from partial observations through curiosity-driven exploration as a conceptual framework.

---

3. Behavior determines the hippocampal spatial mapping of a multisensory environment.

URL: View paper

Brief Assessment

Hippocampal Spatial Mapping[77] investigates neural mechanisms of spatial mapping in mice navigating multisensory environments, focusing on hippocampal activity patterns. This is fundamentally different from ToS, which is a computational framework for evaluating foundation models' ability to construct spatial beliefs through active exploration in simulated environments.

---

4. Exploration patterns shape cognitive map learning

URL: View paper

Brief Assessment

Exploration Patterns Cognitive[72] studies how exploration patterns affect cognitive map learning in humans, focusing on spatial integration and accuracy. It does not propose a computational framework for evaluating foundation models' ability to construct spatial beliefs from partial observations through active exploration.

---

5. CogniPlan: Uncertainty-Guided Path Planning with Conditional Generative Layout Prediction

URL: View paper

Brief Assessment

CogniPlan[69] focuses on path planning using conditional generative layout prediction for mobile robots in exploration and navigation tasks, not on evaluating an agent's ability to construct spatial beliefs from partial observations through curiosity-driven exploration as a general framework.

---

6. An exploration of embodied visual exploration

URL: View paper

Brief Assessment

Embodied Visual Exploration[70] focuses on evaluating different exploration paradigms (curiosity, novelty, coverage, reconstruction) in 3D environments for task-agnostic information gathering, not on modeling spatial belief construction from partial observations as a cognitive framework analogous to Theory of Mind.

---

7. Qualitative spatial representations from task-oriented perception and exploratory behaviors

URL: View paper

Brief Assessment

Qualitative Spatial Representations[76] focuses on qualitative spatial representations from task-oriented perception in robotics, not on evaluating foundation models' ability to construct spatial beliefs through curiosity-driven exploration or probing internal cognitive maps.

---

8. The Spread of Beliefs in Partially Modularized Communities

URL: View paper

Brief Assessment

Beliefs Modularized Communities[73] focuses on belief propagation across social networks with modular community structures, not spatial belief construction from partial observations through physical exploration.

---

9. Exploratory spatial analysis

URL: View paper

Brief Assessment

Exploratory Spatial Analysis[75] is a statistical/geographic methodology for analyzing spatial patterns in data, not a framework for embodied agents constructing spatial beliefs through active exploration. The candidate addresses fundamentally different research questions than the original paper's ToS framework.

---

10. COVID-19 in Toronto: A Spatial Exploratory Analysis

URL: View paper

Brief Assessment

COVID Toronto Spatial[74] focuses on spatial epidemiology and disease spread patterns in geographic regions, not on agent-based spatial belief construction from partial observations through active exploration.

Contribution 2: Comprehensive benchmark for active spatial belief construction

Description: The authors develop a benchmark that evaluates agents through active exploration in procedurally generated multi-room environments. The benchmark includes both text-based and vision-based modalities, scripted proxy agents for disentangling exploration from reasoning, and a suite of spatial cognition tasks covering route and survey knowledge.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. 3DLLM-Mem: Long-Term Spatial-Temporal Memory for Embodied 3D Large Language Model

URL: View paper

Brief Assessment

3DLLM-Mem[57] focuses on long-term memory management for embodied tasks in 3D environments with interactive objects, rather than evaluating active exploration strategies for spatial belief construction through curiosity-driven exploration in procedurally generated multi-room environments.

---

2. Openeqa: Embodied question answering in the era of foundation models

URL: View paper

Brief Assessment

OpenEQA[49] focuses on embodied question answering in real-world environments with open-vocabulary questions, not on active spatial belief construction through procedurally generated multi-room environments with spatial cognition tasks covering route and survey knowledge.

---

3. EscapeCraft: A 3D Room Escape Environment for Benchmarking Complex Multimodal Reasoning Ability

URL: View paper

Brief Assessment

EscapeCraft[56] focuses on room escape tasks requiring multimodal reasoning and prop utilization in 3D environments, rather than evaluating active spatial belief construction through curiosity-driven exploration in multi-room layouts with cognitive map probing as the original paper does.

---

4. Noveld: A simple yet effective exploration criterion

URL: View paper

Brief Assessment

Noveld[53] focuses on exploration strategies in reinforcement learning environments (Minigrid, Nethack, Atari) using intrinsic reward mechanisms, not on benchmarking active spatial belief construction or spatial cognition tasks as described in the original paper.

---

5. SELM: From Efficient Autonomous Exploration to Long-term Monitoring in Semantic Level

URL: View paper

Brief Assessment

SELM[51] focuses on autonomous exploration and long-term monitoring in multi-room environments using semantic scene graphs and POMDP-based patrol strategies, not on benchmarking spatial cognition tasks or evaluating foundation models' ability to construct spatial beliefs through active exploration.

---

6. VisEscape: A Benchmark for Evaluating Exploration-driven Decision-making in Virtual Escape Rooms

URL: View paper

Brief Assessment

VisEscape[58] focuses on escape room puzzle-solving with discrete action spaces and visual quizzes, not on active spatial belief construction through curiosity-driven exploration in multi-room environments as evaluated by the original paper's framework.

---

7. SnapMem: Snapshot-based 3D Scene Memory for Embodied Exploration and Reasoning

URL: View paper

Brief Assessment

SnapMem[55] focuses on snapshot-based 3D scene memory for embodied exploration using frontier-based methods and object co-visibility clustering, not on benchmarking active spatial belief construction with procedurally generated multi-room environments and spatial cognition tasks covering route and survey knowledge as in the original paper.

---

8. CityEQA: A Hierarchical LLM Agent on Embodied Question Answering Benchmark in City Space

URL: View paper

Brief Assessment

CityEQA[54] focuses on embodied question answering in city spaces with task-driven exploration (answering specific questions), whereas the original paper evaluates curiosity-driven exploration for building general spatial beliefs without specific task goals. The benchmarks serve fundamentally different purposes.

---

9. Soundspaces: Audio-visual navigation in 3d environments

URL: View paper

Brief Assessment

Soundspaces[50] focuses on audio-visual navigation in 3D environments with sound-emitting targets, not on active spatial belief construction through curiosity-driven exploration in multi-room environments with spatial cognition tasks covering route and survey knowledge.

---

10. Robocas: A benchmark for robotic manipulation in complex object arrangement scenarios

URL: View paper

Brief Assessment

Robocas[52] focuses on robotic manipulation in complex object arrangement scenarios with physical grasping and obstacle clearance, not on active exploration for spatial belief construction in multi-room environments.

Contribution 3: Direct probing mechanism for internal spatial beliefs

Description: The authors introduce a method to explicitly probe the agent's internal spatial representation by requiring it to externalize its cognitive map at any exploration step. This allows measurement of not only task performance but also the quality, consistency, and evolution of the underlying spatial model itself, moving beyond black-box evaluation.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Spatial mental modeling from limited views

URL: View paper

Prior Art Analysis

Spatial Mental Modeling[63] demonstrates prior work that explicitly probes internal spatial representations by requiring agents to externalize cognitive maps. The candidate paper introduces a mechanism where models are prompted to 'explicitly present their cognitive map at each step' and measures 'the quality, consistency, and evolution of the underlying spatial model itself.' This directly parallels the original paper's contribution of probing internal spatial beliefs through externalized cognitive maps, showing that similar methodology existed before the original work.

Evidence

Evidence 1 - **Rationale**: While the candidate paper focuses on multi-view spatial reasoning rather than active exploration, it similarly evaluates internal spatial representations through explicit probing mechanisms, demonstrating that the methodology of externalizing and measuring cognitive maps was already established in prior work. - **Original**: explicit probing of the internal spatial belief. behavioral success such as whether the agent find the chair cannot directly reveal the quality of agent's internal model. we require the agent to

explicitly represent its spatial belief by probing its cognitive map at any point of exploration. - **Candidate**: we introduce m ind cube , featuring 21, 154 questions and 3, 268 images, organized into 976 multi-view groups through various types of viewpoint transformations (i.e., r otation , among , a round in figure 2). we annotate questions with a focus on objects that are not visible in the current query vi...

Evidence 2 - **Rationale**: Both papers establish metrics to evaluate the quality of externalized cognitive maps against ground truth, showing that the methodology of measuring spatial belief accuracy through map comparison was already present in the candidate work. - **Original**: we evaluate probed belief using three complementary metrics. positional accuracy (pos.acc) is the average euclidean distance between predicted and true object coordinates (lower is better). directional accuracy (dir.acc) is the average absolute difference (in degrees) between pairwise object directi... - **Candidate**: beyond evaluating task performance using qa accuracy, we also introduce two welldefined graph metrics for generated cognitive maps: (1) overall similarity, a weighted score combining directional and facing consistency; and (2) isomorphic rate, measuring whether all pairwise object relations match th...

---

2. Probing mental representations of space through sketch mapping: a scoping review

URL: View paper

Brief Assessment

Probing Mental Representations[68] focuses on human cognitive mapping through sketch maps in real-world spatial cognition research, not on computational agents' internal representations in reinforcement learning frameworks.

---

3. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond

URL: View paper

Brief Assessment

GeoAI[65] focuses on spatially explicit AI techniques for geographic knowledge discovery and does not address probing internal spatial representations or cognitive maps in agents. The candidate paper's content is insufficient to evaluate overlap with the original paper's contribution on externalizing agent beliefs.

---

4. VResin: Externalizing spatial memory into 3D sketch maps

URL: View paper

Brief Assessment

VResin[62] focuses on externalizing human spatial memory through 3D sketching interfaces in VR for post-task recall, not on probing AI agents' internal spatial representations during active exploration as in the original paper.

---

5. From reactive to cognitive: brain-inspired spatial intelligence for embodied agents

URL: View paper

Brief Assessment

Reactive to Cognitive[61] does not probe internal spatial beliefs during exploration. Instead, it constructs structured spatial memory (landmark memory and cognitive maps) that is retrieved for navigation tasks, but does not externalize or measure the agent's cognitive map at each exploration step as the original paper does.

---

6. Probing for consciousness in machines

URL: View paper

Brief Assessment

Probing Consciousness Machines[64] focuses on probing neural activations to detect spatial position encoding in reinforcement learning agents playing video games, not on externalizing cognitive maps during active exploration as in the original paper.

---

7. Thinking in space: How multimodal large language models see, remember, and recall spaces

URL: View paper

Brief Assessment

Thinking in Space[59] focuses on probing cognitive maps in video-based spatial understanding tasks, while the original paper addresses active exploration in interactive environments with explicit belief externalization at each exploration step. The modalities and evaluation paradigms differ fundamentally.

---

8. Mind meets space: Rethinking agentic spatial intelligence from a neuroscience-inspired perspective

URL: View paper

Brief Assessment

Mind Meets Space[67] focuses on neuroscience-inspired frameworks for spatial intelligence but does not demonstrate prior work on explicitly probing agents' internal cognitive maps at exploration steps. The candidate emphasizes bio-inspired architectures rather than evaluation methodologies for externalizing spatial beliefs.

---

9. Learning place cells and remapping by decoding the cognitive map

URL: View paper

Brief Assessment

Learning Place Cells[60] focuses on decoding spatial representations from neural network units to reconstruct position, not on probing an agent's internal cognitive map during active exploration as in the original paper.

---

10. From geometry to behavior: An introduction to spatial cognition

URL: View paper

Brief Assessment

Geometry to Behavior[66] discusses cognitive maps in the context of animal research and spatial cognition theory, but does not present a computational mechanism for probing foundation models' internal representations during active exploration as the original paper does.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] What Lies Beyond the View? Actively Constructing Spatial Beliefs in Foundation Models View paper
- [1] SpatialVLA: Exploring Spatial Representations for Visual-Language-Action Model View paper
- [2] Voronav: Voronoi-based zero-shot object navigation with large language model View paper

- [3] Embodied-r: Collaborative framework for activating embodied spatial reasoning in foundation models via reinforcement learning View paper
- [4] Exploring Spatial Schema Intuitions in Large Language and Vision Models View paper
- [5] Pre-trained language models for interactive decision-making View paper
- [6] Explore until confident: Efficient exploration for embodied question answering View paper
- [7] Can foundation models actively gather information in interactive environments to test hypotheses? View paper
- [8] SSR-ZSON: Zero-Shot Object Navigation via Spatial-Semantic Relations within a Hierarchical Exploration Framework View paper
- [9] LLM-Guided Zero-Shot Visual Object Navigation with Building Semantic Map View paper
- [10] Seeing is Believing: Belief-Space Planning with Foundation Models as Uncertainty Estimators View paper
- [11] Et-plan-bench: Embodied task-level planning benchmark towards spatial-temporal cognition with foundation models View paper
- [12] Wmnav: Integrating vision-language models into world models for object goal navigation View paper
- [13] Assessing adaptive world models in machines with novel games View paper
- [14] Foundation Models for Robotic Tasks: Survey, Challenges and Future Directions View paper
- [15] Sensei: Semantic exploration guided by foundation models to learn versatile world models View paper
- [16] Towards Autonomous UAV Visual Object Search in City Space: Benchmark and Agentic Methodology View paper
- [17] Spatial Representations for Learning Robotic Mobile Manipulation View paper
- [18] STMA: A spatio-temporal memory agent for long-horizon embodied task planning View paper
- [19] BeliefMapNav: 3D Voxel-Based Belief Map for Zero-Shot Object Navigation View paper
- [20] Semantic Exploration from Language Abstractions and Pretrained Representations View paper
- [21] Seal: Self-supervised embodied active learning using exploration and 3d consistency View paper
- [22] Research on an autonomous exploration strategy for hydroelectric generator inspection robots driven by large language models View paper
- [23] Mesoscale explorer: Visual exploration of large-scale molecular models. View paper
- [24] Exploring a cognitive basis for learning spatial relationships with augmented reality View paper
- [25] Student-centered learning environments: Foundations, assumptions and design View paper
- [26] Actial: Activate Spatial Reasoning Ability of Multimodal Large Language Models View paper
- [27] IndustryNav: Exploring Spatial Reasoning of Embodied Agents in Dynamic Industrial Navigation View paper
- [28] A dual-space model of iteratively deepening exploratory learning View paper
- [29] Stairway to Autonomy: Hierarchical Decision-Making for LLM-Guided Planning, Bandit-Driven Exploration, and Multi-Agent Navigation View paper
- [30] Vision to Geometry: 3D Spatial Memory for Sequential Embodied MLLM Reasoning and Exploration View paper
- [31] Game object model version II: a theoretical framework for educational game development View paper
- [32] ASTRA: Autonomous Spatial-Temporal Red-teaming for AI Software Assistants View paper
- [33] Thinking on Maps: How Foundation Model Agents Explore, Remember, and Reason Map Environments View paper
- [34] Foundation Models for Semantic Novelty in Reinforcement Learning View paper
- [35] PIGEON: VLM-Driven Object Navigation via Points of Interest Selection View paper
- [36] From Prompts to Paths: Large Language Models for Zero-Shot Planning in Unmanned Ground Vehicle Simulation View paper
- [37] SpatialReasoner: Active Perception for Large-Scale 3D Scene Understanding View paper
- [38] ImagineNav++: Prompting Vision-Language Models as Embodied Navigator through Scene Imagination View paper
- [39] Multi-Floor Zero-Shot Object Navigation Policy View paper
- [40] Active-O3: Empowering Multimodal Large Language Models with Active Perception via GRPO View paper
- [41] The Foundation Model Path to Open-World Robots View paper
- [42] From Prompts to Paths: Large Language Models for Zero-Shot Planning and Simulation View paper
- [43] Differential effects of aging on spatial learning through exploratory navigation and map reading. View paper
- [44] Learning-dependent evolution of spatial representations in large-scale virtual environments. View paper
- [45] Cross-Embodied Cognitive Morphologies. Proceedings 2022, 81, 10 View paper
- [46] Fast Mining of Spatial Frequent Wordset from Social Database View paper
- [47] Supplementary Materials for â□□GeoExplorer: Active Geo-localization with Curiosity-Driven Explorationâ□□ View paper
- [48] Language-Guided Semantic Affordance Exploration for Efficient Reinforcement Learning View paper
- [49] Openeqa: Embodied question answering in the era of foundation models View paper
- [50] Soundspaces: Audio-visual navigation in 3d environments View paper
- [51] SELM: From Efficient Autonomous Exploration to Long-term Monitoring in Semantic Level View paper
- [52] Robocas: A benchmark for robotic manipulation in complex object arrangement scenarios View paper
- [53] Noveld: A simple yet effective exploration criterion View paper
- [54] CityEQA: A Hierarchical LLM Agent on Embodied Question Answering Benchmark in City Space View paper
- [55] SnapMem: Snapshot-based 3D Scene Memory for Embodied Exploration and Reasoning View paper
- [56] EscapeCraft: A 3D Room Escape Environment for Benchmarking Complex Multimodal Reasoning Ability View paper
- [57] 3DLLM-Mem: Long-Term Spatial-Temporal Memory for Embodied 3D Large Language Model View paper
- [58] VisEscape: A Benchmark for Evaluating Exploration-driven Decision-making in Virtual Escape Rooms View paper
- [59] Thinking in space: How multimodal large language models see, remember, and recall spaces View paper
- [60] Learning place cells and remapping by decoding the cognitive map View paper
- [61] From reactive to cognitive: brain-inspired spatial intelligence for embodied agents View paper
- [62] VResin: Externalizing spatial memory into 3D sketch maps View paper
- [63] Spatial mental modeling from limited views View paper
- [64] Probing for consciousness in machines View paper
- [65] GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond View paper
- [66] From geometry to behavior: An introduction to spatial cognition View paper
- [67] Mind meets space: Rethinking agentic spatial intelligence from a neuroscience-inspired perspective View paper
- [68] Probing mental representations of space through sketch mapping: a scoping review View paper
- [69] CogniPlan: Uncertainty-Guided Path Planning with Conditional Generative Layout Prediction View paper
- [70] An exploration of embodied visual exploration View paper
- [71] Cognitive mapping and planning for visual navigation View paper

- [72] Exploration patterns shape cognitive map learning View paper
- [73] The Spread of Beliefs in Partially Modularized Communities View paper
- [74] COVID-19 in Toronto: A Spatial Exploratory Analysis View paper
- [75] Exploratory spatial analysis View paper
- [76] Qualitative spatial representations from task-oriented perception and exploratory behaviors View paper
- [77] Behavior determines the hippocampal spatial mapping of a multisensory environment. View paper