

Habitat-Sim Performance and Generalization in Embodied AI Benchmarks vs. PyBullet and MuJoCo

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does Habitat-Sim's photorealistic 3D simulation performance compare to other embodied AI simulators like PyBullet or MuJoCo in terms of FPS and agent training convergence when evaluated on the. 10 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Habitat: A Platform for Embodied AI Research. Research question: How does Habitat-Sim's photorealistic 3D simulation performance compare to other embodied AI simulators like PyBullet or MuJoCo in terms of FPS and agent training convergence when evaluated on the RoboStack benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.8/10.

3 Results

13 papers retrieved. 10 claims extracted; 2 independently verified. Quality review score: 4.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Learning outperforms SLAM if scaled to an order of magnitude more experience than previous investigations.	✓	0.23
Only agents with depth (D) sensors generalize well across datasets.	✓	0.17
The availability of large-scale 3D scene datasets and community interest in active vision tasks led to a recent surge of	×	0.06
Existing simulators exhibit several shortcomings, including tight coupling of task, simulation platform, and 3D dataset,	×	0.06
Most existing indoor simulators operate at relatively low frame rates (10-100 fps), becoming a bottleneck in training ag	×	0.06
Work built on top of any of the existing platforms is hard to reproduce independently from the platform, and thus hard t	×	0.03
The Blind agent experiences the highest number of collisions, while agents possessing depth sensors (Depth and RGBD) hav	×	0.06
The distribution over Euclidean distance between start and goal, distribution over geodesic distance along shortest path	×	0.01
We manually curated the full dataset of Gibson 3D textured meshes to select meshes that do not exhibit significant recon	×	0.03
A key issue that we tried to avoid is the presence of holes or cracks in floor surfaces, which is particularly problemat	×	0.02

References

- <http://arxiv.org/abs/2507.02910v1>
- <http://arxiv.org/abs/2410.09604v1>
- <http://arxiv.org/abs/1904.01201v2>