

RxR and R2R Trained Agents on ALFRED: A Comparative Performance Analysis

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the relative performance of RxR-trained agents versus R2R-trained agents on the ALFRED benchmark for task and language grounding in realistic indoor environments. We introduce Room-Across-Room (RxR), a new Vision-and-Language Navigation (VLN) dataset. RxR is multilingual (English, Hindi, and Telugu) and larger (more paths and instructions) than other VLN datasets. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Room-Across-Room: Multilingual Vision-and-Language Navigation with Dense Spatiotemporal Grounding. Research question: What is the relative performance of RxR-trained agents versus R2R-trained agents on the ALFRED benchmark for task and language grounding in realistic indoor environments?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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.

References

- <http://arxiv.org/abs/2010.07954v1>
- <http://arxiv.org/abs/2506.00739v4>
- <http://arxiv.org/abs/2507.04047v2>