RxR-Trained Agents with Tryout Controller Outperform Benchmarks in Path Efficiency

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the path efficiency of RxR-trained agents with the tryout controller compare to agents trained with other navigation benchmarks (e.g., ALFRED, Room-Across-Room) when evaluated on unseen. 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. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. 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: How does the path efficiency of RxR-trained agents with the tryout controller compare to agents trained with other navigation benchmarks (e.g., ALFRED, Room-Across-Room) when evaluated on unseen environments?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

10 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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
RxR contains 126K instructions covering 16.5K sampled guide paths and 126K human follower demonstration paths.	×	0.05
RxR includes instructions for three typologically diverse languages: English (en), Hindi (hi), and Telugu (te).	×	0.05
RxR's instructions are not translations: all instructions are created from scratch by native speakers.	×	0.02
RxR's English portion includes instructions by speakers in the USA (en-US) and India (en-IN).	×	0.03
RxR's dataset is available at https://github.com/google-research-datasets/RxR.	×	0.09
RxR addresses shortcomings of existing VLN datasets, including multilinguality, scale, fine-grained word grounding, and	×	0.09
RxR addresses path biases in R2R.	×	0.03
RxR's multilingual agent obtains across the board improvements when trained on both R2R and RxR datasets.	×	0.06
RxR's multilingual agent outperforms the monolingual agent in terms of NE, SR, SDTW, and NDTW metrics on the test set.	×	0.05
RxR's multilingual agent achieves an average NE of 11.9, SR of 21.3, SDTW of 17.1, and NDTW of 37.2 on the test set.	×	0.03
RxR's human performance achieves an average NE of 0.9, SR of 93.9, SDTW of 76.9, and NDTW of 79.5 on the test set.	×	0.02

References

- http://arxiv.org/abs/2010.07954v1
- http://arxiv.org/abs/2305.09349v1
- http://arxiv.org/abs/2203.08975v2