# Multimodal GUI Agent Error Accumulation Across Task Complexity Levels in AndroidWorld

Assignee Research

June 6, 2026

**Abstract**

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does task complexity in multimodal GUI agents correlate with step-wise error accumulation when evaluated on the AndroidWorld benchmark. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: LaSM: Layer-wise Scaling Mechanism for Defending Pop-up Attack on GUI Agents. Research question: How does task complexity in multimodal GUI agents correlate with step-wise error accumulation when evaluated on the AndroidWorld benchmark?.

## 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. 0 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.

## References


- http://arxiv.org/abs/2406.10819v2
- http://arxiv.org/abs/2411.18279v12
- http://arxiv.org/abs/2507.10610v3