

Llama3 and DeepSeek-R1 Robustness on Adversarially Perturbed Code Generation Benchmarks

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the difference in robustness scores between Llama3 and Deepseek R1 when evaluated on adversarially perturbed code generation benchmarks. 9 claims were extracted from source literature; 8 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. Research question: What is the difference in robustness scores between Llama3 and Deepseek R1 when evaluated on adversarially perturbed code generation benchmarks?.

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 7.7/10.

3 Results

13 papers retrieved. 9 claims extracted; 8 independently verified. Quality review score: 7.7/10.

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
AI Agents are modular systems driven and enabled by LLMs and LIMs for task-specific automation.	✓	0.23
Generative AI is a precursor providing the foundation for AI Agents.	✓	0.20
AI Agents advance through tool integration, prompt engineering, and reasoning enhancements.	✓	0.20
Agentic AI systems represent a paradigm shift marked by multi-agent collaboration, dynamic task decomposition, persistence	✓	0.34
AI Agents are used in application domains such as customer support, scheduling, and data summarization.	✓	0.21
Agentic AI is deployed in research automation, robotic coordination, and medical decision support.	✓	0.24
Challenges in AI Agents include hallucination and brittleness.	×	0.13
Challenges in Agentic AI include emergent behavior and coordination failure.	✓	0.19
Solutions proposed for challenges in AI Agents and Agentic AI include ReAct loops, retrieval-augmented generation (RAG),	✓	0.29

References

- <https://doi.org/10.3390/bdcc9120320>
- <https://doi.org/10.48550/arxiv.2412.19437>
- <https://doi.org/10.70777/si.v2i3.15161>