Adversarial Obfuscation Effects on Llama3 and Codestral Vulnerability Detection

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the vulnerability detection accuracy of Llama3 and Codestral degrade under adversarial code obfuscation techniques compared to standard Big-Vul samples. 12 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Deep learning modelling techniques: current progress, applications, advantages, and challenges. Research question: How does the vulnerability detection accuracy of Llama3 and Codestral degrade under adversarial code obfuscation techniques compared to standard Big-Vul samples?.

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 6.7/10.

3 Results

10 papers retrieved. 12 claims extracted; 7 independently verified. Quality review score: 6.7/10.

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
Deep learning utilizes two or more levels of nonlinear feature transformation via representation learning.	✓	0.17
Articles surveying deep learning architectures encompassing the full scope of the field are limited.	✓	0.17
Many deep learning models exhibit highly domain-specific efficiency.	✓	0.19
Many deep learning models can be trained by two or more methods.	×	0.10
Training deep learning models can be very time-consuming.	×	0.14
Training deep learning models can be expensive.	×	0.08
Training deep learning models requires huge samples for better accuracy.	✓	0.17
Deep learning is susceptible to deception and misclassification.	×	0.13
Deep learning models tend to get stuck on local minima during training.	×	0.12
Deep learning has led to results in the healthcare, education, security, commercial, industrial, and government sectors.	✓	0.21
Convolutional neural networks (CNN), generative adversarial networks (GAN), recurrent neural networks (RNN), recursive n	✓	0.29
The potential of deep learning models other than CNN, GAN, RNN, recursive neural networks, and autoencoders remains wide	✓	0.21

References

- https://doi.org/10.1038/s42256-020-0186-1

- https://doi.org/10.48550/arxiv.2210.11416
- https://doi.org/10.1007/s10462-023-10466-8