

# Adversarial Fine-Tuning Effects on Cross-Language Vulnerability Detection in Llama3 and Codestral

Assignee Research

June 4, 2026

## Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does adversarial fine-tuning affect the cross-language vulnerability detection F1 scores of Llama3 compared to Codestral on C++ and Python codebases. 7 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Convolutional Neural Networks: A Survey. Research question: How does adversarial fine-tuning affect the cross-language vulnerability detection F1 scores of Llama3 compared to Codestral on C++ and Python codebases?.

## 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 8.2/10.

## 3 Results

10 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 8.2/10.

## 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
Convolutional neural networks (CNNs) are a subset of Artificial intelligence (AI).	✓	0.25
CNNs have been utilized to classify DNA sequences in the field of genomics.	✓	0.17
CNNs are used for tasks including image recognition, speech recognition, and natural language processing (NLP).	✓	0.27
The fundamentals of CNNs include layers, convolution operation (Conv_Op), feature maps (Feat_Maps), activation functions	✓	0.21
LeNet, AlexNet, VGG, ResNet, and Inception-Net are popular CNN architectures.	✓	0.21
TensorFlow, Keras, PyTorch, Caffe, and MXNet are existing platforms and libraries for CNNs.	✓	0.23
Recent developments in CNNs include attention mechanisms, capsule networks, transfer learning, and adversarial training.	✓	0.25

## References

- <https://doi.org/10.1109/tnsm.2020.2971776>
- <https://doi.org/10.3390/computers12080151>
- <https://doi.org/10.1007/s12559-023-10179-8>