

# Scalability and Accuracy Trade-offs in Llama3 and DeepSeek-R1 for Vulnerability Detection

Assignee Research

June 4, 2026

## Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How scalable are Llama3 and Deepseek R1 in detecting vulnerabilities in the Big-Vul dataset when trained with MultiVul’s multimodal approach, as measured by the throughput (samples per second) and. 11 claims were extracted from source literature; 4 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Introducing Representations of Facial Affect in Automated Multimodal Deception Detection. Research question: How scalable are Llama3 and Deepseek R1 in detecting vulnerabilities in the Big-Vul dataset when trained with MultiVul’s multimodal approach, as measured by the throughput (samples per second) and accuracy trade-offs when increasing the model size from 7B to 30B parameters?.

## 2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.8/10.

## 3 Results

16 papers retrieved. 11 claims extracted; 4 independently verified. Quality review score: 5.8/10.

## 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
The highest-performing automated approach achieved an AUC of 87.7% by using interpretable visual, vocal, and verbal feat	✓	0.24
The highest-performing automated approach used the following features: (1) head, face, and eye movements, (2) MFCC coeff	×	0.06
The primary metric for comparing the effectiveness of different approaches is AUC.	×	0.04
Unimodal SVMs were trained on the respective features of each modality to assess the effectiveness of individual modalit	×	0.10
Early fusion SVMs were trained on each of the 11 possible combinations of the 4 modalities.	×	0.04
Six non-generative ensemble approaches were experimented with, each with a different method of combining the decisions o	×	0.04
Deceptive and truthful speakers exhibited significant differences in mean, median, and standard deviation of facial vale	×	0.11
The work aims to inform and motivate the development of affect-aware approaches for automated deception detection in-the	×	0.12
The work presents a novel, automated approach for video-based deception detection that uses continuous representations o	✓	0.29
The work evaluates unimodal and multimodal SVM-based machine learning approaches to identify effective features, modalit	✓	0.25
The work provides a novel analysis of the computational relationship between facial affect and deception observed in thi	✓	0.17

## References

- <http://arxiv.org/abs/2403.09513v1>
- <http://arxiv.org/abs/2502.00425v2>
- <http://arxiv.org/abs/2008.13369v1>