# Adversarial Robustness of AI Text Detectors Across Model Sizes in Low-Resource Languages

Assignee Research

June 9, 2026

**Abstract**

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the impact of model size (e.g., 7B vs. 13B vs. 30B parameters) on the adversarial robustness of AI-generated text detectors when tested on low-resource language benchmarks with code. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Are AI-Generated Text Detectors Robust to Adversarial Perturbations?. Research question: What is the impact of model size (e.g., 7B vs. 13B vs. 30B parameters) on the adversarial robustness of AI-generated text detectors when tested on low-resource language benchmarks with code perturbations?.

## 2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.3/10.

## 3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.3/10.

## 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/2404.17216v1
- http://arxiv.org/abs/2406.01179v2
- http://arxiv.org/abs/2603.17522v1