Clustering Technique Effects on Adversarial Robustness in AdaptToken Models

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the impact of different clustering techniques (e.g., DBSCAN, K-Means) on the adversarial robustness of AdaptToken-8B vs. AdaptToken-3B, as evaluated by accuracy under untargeted adversarial. 9 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Evaluating Cascading Impact of Attacks on Resilience of Industrial Control Systems: A Design-Centric Modeling Approach. Research question: What is the impact of different clustering techniques (e.g., DBSCAN, K-Means) on the adversarial robustness of AdaptToken-8B vs. AdaptToken-3B, as evaluated by accuracy under untargeted adversarial attacks on AdvGLUE benchmarks?.

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 3.5/10.

3 Results

16 papers retrieved. 9 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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 impact of attack was found to be greater when the initial state of the system was at the high vulnerable state as co	×	0.04
The time-to-critical-state increases downstream for processes that are connected sequentially.	×	0.06
The time-to-critical-state was found to be shorter when the initial state was at the low vulnerable state.	×	0.05
The time-to-critical-state were identical for pairs of processes which do not have tanks (i.e. P1-P2 and P4-P5).	×	0.05
There is an increase in the time-to-critical-state for processes with tanks.	×	0.09
The effect of attacks on P1 and P2 were instantaneous and flow was disrupted.	×	0.02
Due to the tank in P3, disruption of flow was only observed after the tank was depleted.	×	0.00
The impact of attack for command spoofing and spoofing of Tank 3 level were similar where the impact was largest for FIT	×	0.03
The spoofing of Tank 1 level resulted in a positive ratio due to the increase in tank level from the exploitation of con	×	0.04

References

- http://arxiv.org/abs/1905.11736v5
- http://arxiv.org/abs/1905.03156v2
- http://arxiv.org/abs/2307.02055v1