Novel Tabular Generative Metrics and Their Correlation with LLM Reasoning Performance

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: To what extent do novel tabular generative metrics correlate with downstream LLM reasoning performance when synthetic mixed-type data is used for instruction tuning. 13 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.9/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Evaluating Generative Models for Tabular Data: Novel Metrics and Benchmarking. Research question: To what extent do novel tabular generative metrics correlate with downstream LLM reasoning performance when synthetic mixed-type data is used for instruction tuning?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.9/10.

3 Results

8 papers retrieved. 13 claims extracted; 5 independently verified. Quality review score: 5.9/10.

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
FAED effectively captures generative modeling issues overlooked by existing metrics.	✓	0.28
FPCAD exhibits promising performance but requires further refinements to enhance its reliability.	✓	0.18
FAED successfully detects all synthesized problems in the experimental results on three standard network intrusion detec	×	0.14
FPCAD shows promising but improvable performance in detecting synthesized problems.	×	0.04
Existing metrics fail to identify key issues in generative modeling for tabular data.	×	0.14
FAED and FPCAD are novel and robust metrics for evaluating generative models in the tabular data domain.	✓	0.28
Quality Decrease, Mode Drop, and Mode Collapse issues were embedded into the datasets to simulate real-world generative	×	0.05
FAED, FPCAD, and RFIS are tailored for assessing generative models in tabular data, inspired by IS and FID from the imag	✓	0.20
Existing evaluation principles from the image domain can be effectively adapted for tabular data.	×	0.12
SDV Fidelity, Utility, TSTR, and TRTS exhibit limitations in detecting key generative modeling challenges.	×	0.13
TSTR is useful for detecting cases where synthetic data only partially represents real data.	×	0.07
TRTS assesses whether synthetic samples introduce patterns absent in real data.	×	0.04
FAED, FPCAD, and RFIS offer robust tools for assessing generative models in the tabular data domain.	✓	0.24

References

- http://arxiv.org/abs/2310.04793v2
- http://arxiv.org/abs/2504.20900v1

- http://arxiv.org/abs/2408.16440v2