

# Synthetic Data Generation Methods and Their Impact on Tabular Foundation Model Reasoning

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does the choice of synthetic data generation method (e.g., GAN-based vs. rule-based) influence the reasoning capabilities of tabular foundation models when evaluated on explainability benchmarks. 11 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Generative adversarial networks vs large language models: a comparative study on synthetic tabular data generation. Research question: How does the choice of synthetic data generation method (e.g., GAN-based vs. rule-based) influence the reasoning capabilities of tabular foundation models when evaluated on explainability benchmarks like TabExplain?.

## 2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

## 3 Results

9 papers retrieved. 11 claims extracted; 2 independently verified. Quality review score: 4.7/10.

## 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
Among the four LLMs tested using the Iris dataset prompt, only GPT-4o was capable of generating tabular data in a zero-s	✓	0.17
Other LLMs were unable to output the datasets due to computational limits, output size restrictions, and/or inability to	×	0.02
The Claude 3.5 Sonnet model was capable of generating relevant data (in csv format) with additional guiding prompts, but	×	0.05
Three datasets were used in the present analysis: Iris, Fish Measurements, and Real Estate Valuation.	✓	0.15
Two of the datasets were sourced from the University of California, Irvine Machine Learning Repository, and a third from	×	0.03
The datasets were pre-processed to ensure completeness and to remove non-numeric columns.	×	0.04
No outlier removal was performed.	×	0.00
Each dataset’s univariate (mean, standard deviation, range) and bivariate (Pearson’s product moment correlation) statist	×	0.04
A plain-language prompt was designed to instruct the generation of synthetic data associated with each of the datasets i	×	0.15
No pre-training or fine-tuning was performed and the RWD was not provided to the LLMs.	×	0.11
The prompts associated with each dataset described the desired format, sample size, means, standard deviation.	×	0.02

## References

- <http://arxiv.org/abs/2502.14523v1>
- <http://arxiv.org/abs/2104.09630v2>
- <http://arxiv.org/abs/2512.03307v1>