

# Adversarial Noise Effects on Foundation Model Generalization in Synthetic Tabular Data

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the impact of varying levels of adversarial noise in synthetic tabular data on the generalization capability of foundation models, as measured by accuracy on out-of-distribution benchmarks. 12 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Robust Tabular Foundation Models. Research question: What is the impact of varying levels of adversarial noise in synthetic tabular data on the generalization capability of foundation models, as measured by accuracy on out-of-distribution benchmarks such as TabMNAR and TabCI?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

## 3 Results

13 papers retrieved. 12 claims extracted; 1 independently verified. Quality review score: 4.2/10.

## 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
Tabular foundation models (TFMs) rely on in-context learning (ICL) for classification and regression tasks with structured	×	0.11
TFMs can produce high-quality predictions on new datasets in milliseconds when GPU-accelerated.	×	0.07
Training TFMs relies on generating diverse synthetic datasets constructed from structural causal models (SCMs).	×	0.07
All current publicly available, competitive TFMs have been pretrained on datasets generated from a fixed prior distribution	×	0.06
Fixed priors in TFM training underrepresent certain regions of the parameter space, potentially degrading performance on	×	0.06
State-of-the-art TFMs lag behind tree-based methods on some benchmarks.	×	0.06
The proposed method formalizes adversarial training over the SCM parameter space to adapt to challenging regions such as	×	0.06
The proposed algorithm is named ROBUST TABULAR FOUNDATION MODELS (RTFM).	✓	0.18
Applying RTFM to TabPFN V2 with only 90k additional training datasets significantly improves its ranking on several real	×	0.12
The optimality gap is estimated by sampling a fixed number of generators and datasets, then computing the difference between	×	0.09
In the described implementation, estimating the optimality gap with $nds=20$ and $e=7$ takes a matter of seconds when parallel	×	0.02
The benchmark table includes SCM parameters such as feature count, sample size, activation functions (e.g., tanh, identity)	×	0.02

## References

- <http://arxiv.org/abs/2409.02136v2>
- <http://arxiv.org/abs/2504.20900v1>

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