

# Synthetic-to-Real Data Ratio Effects on TabNet Fine-Tuning Performance in CausalMixFT

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: What is the impact of varying the ratio of synthetic-to-real data in CausalMixFT on the fine-tuning performance of TabNet when evaluated on the Adult Census dataset. 14 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Fine-T2I: An Open, Large-Scale, and Diverse Dataset for High-Quality T2I Fine-Tuning. Research question: What is the impact of varying the ratio of synthetic-to-real data in CausalMixFT on the fine-tuning performance of TabNet when evaluated on the Adult Census dataset?.

## 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 4.4/10.

## 3 Results

8 papers retrieved. 14 claims extracted; 1 independently verified. Quality review score: 4.4/10.

## 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
Fine-tuning LlamaGen on the Fine-T2I dataset achieves an 80.7% win rate for visual quality compared to a counterpart wit	×	0.13
Fine-tuning LlamaGen on the Fine-T2I dataset achieves a 65.3% win rate for text-image alignment compared to a counterpart	✓	0.17
GenEval benchmark scores improve for both LlamaGen and SD-XL models after fine-tuning on the Fine-T2I dataset.	×	0.12
In a human evaluation comparing LlamaGen fine-tuned on Fine-T2I, BLIP3o-60k, and T2I-2M, the Fine-T2I model achieved a 4	×	0.12
In a human evaluation comparing LlamaGen fine-tuned on Fine-T2I, BLIP3o-60k, and T2I-2M, the Fine-T2I model achieved a 3	×	0.09
In the human evaluation shown in Figure 8, the BLIP3o-60k fine-tuned model achieved a 29.5% win rate for Text-Image Alig	×	0.10
In the human evaluation shown in Figure 8, the T2I-2M fine-tuned model achieved a 21.3% win rate for Visual Quality.	×	0.07
The Fine-T2I dataset contains 37.9% prompts categorized as 'Text'.	×	0.08
The Fine-T2I dataset contains 27.8% prompts categorized as 'Nature'.	×	0.06
The Fine-T2I dataset contains 17.4% prompts categorized as 'Rendering'.	×	0.06
The Fine-T2I dataset contains 10.6% prompts categorized as 'Activities'.	×	0.07
The Fine-T2I dataset contains 6.3% prompts categorized as 'Portrait'.	×	0.07
36.9% of the tasks in the Fine-T2I dataset involve multiple colors.	×	0.04
Enhanced prompts in the Fine-T2I dataset have a longer average length than original prompts, as shown in the log-scaled	×	0.02

## References

- <http://arxiv.org/abs/2411.15497v3>

- <http://arxiv.org/abs/2005.07866v1>
- <http://arxiv.org/abs/2602.09439v1>