What is the impact of diffusion-based tabular data augmentation on zero-shot performance of LLMs on the SuperG

Assignee Research

June 10, 2026

Abstract

The exponential growth of Large Language Models (LLMs) continues to highlight the need for efficient strategies to meet ever-expanding computational and data demands. This survey provides a comprehensive analysis of two complementary paradigms: Knowledge Distillation (KD) and Dataset Distillation (DD), both aimed at compressing LLMs while preserving their advanced reasoning capabilities and linguistic diversity. We first examine key methodologies in KD, such as task-specific alignment, rationale-based training, and multi-teacher frameworks, alongside DD techniques that synthesize compact, high

1 Introduction

This paper examines: Knowledge distillation and dataset distillation of large language models: emerging trends, challenges, and future directions. Research question: What is the impact of diffusion-based tabular data augmentation on zero-shot performance of LLMs on the SuperGLUE benchmark when compared to CTGAN-generated data?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.3/10.

3 Results

11 papers retrieved. 8 claims extracted; 8 independently verified. Quality review score: 8.3/10.

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 exponential growth of Large Language Models (LLMs) continues to highlight the need for efficient strategies to meet	✓	0.34
Knowledge Distillation (KD) and Dataset Distillation (DD) are two complementary paradigms aimed at compressing LLMs whil	✓	0.38
Key methodologies in KD include task-specific alignment, rationale-based training, and multi-teacher frameworks.	✓	0.26
Dataset Distillation (DD) techniques synthesize compact, high-impact datasets through optimization-based gradient matchi	✓	0.37
Integrating KD and DD can produce more effective and scalable compression strategies for LLMs.	✓	0.25
Distillation techniques address persistent challenges in model scalability, architectural heterogeneity, and the preserv	✓	0.28
Applications of distillation techniques span domains such as healthcare and education, enabling efficient deployment wit	✓	0.19
Open challenges remain in preserving emergent reasoning and linguistic diversity, enabling efficient adaptation to conti	✓	0.41

References

- `https://doi.org/10.1007/s10462-025-11423-3`
- `https://doi.org/10.1016/j.jiixd.2024.01.002`
- `https://doi.org/10.48550/arxiv.2308.05384`