

# Optimal Model Scaling Factors for Federated LLM Code Generation Under IID and Non-IID Data

Assignee Research

May 31, 2026

## Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the optimal model scaling factor for code generation quality (measured by HumanEval pass@10) when training federated LLMs with partial client participation and IID versus non-IID data. The proliferation of edge devices has brought Federated Learning (FL) to the forefront as a promising paradigm for decentralized and collaborative model training while preserving the privacy of clients' data. However, FL struggles with a significant performance reduction and. 6 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Stable Diffusion-based Data Augmentation for Federated Learning with Non-IID Data. Research question: What is the optimal model scaling factor for code generation quality (measured by HumanEval pass@10) when training federated LLMs with partial client participation and IID versus non-IID data distributions?.

## 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 3.8/10.

### 3 Results

11 papers retrieved. 6 claims extracted; 0 independently verified. Quality review score: 3.8/10.

### 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
Gen-FedSD significantly enhances accuracy by at least 12% and 6% on CIFAR-10 and CIFAR-100 datasets, respectively, under	×	0.04
Under extreme data heterogeneity, Gen-FedSD exhibits accuracy improvements of at least 20% and 7% for the CIFAR-10 and C	×	0.04
Gen-FedSD reduces the model size from 567.44 MB to 22.69 MB for the Vanilla FedAvg method under Dir (0.5) setting.	×	0.05
Gen-FedSD reduces the model size from 1588.84 MB to 22.69 MB for the Vanilla FedProx method under Dir (0.5) setting.	×	0.04
Gen-FedSD reduces the model size from 907.91 MB to 22.69 MB for the Vanilla FedNova method under Dir (0.5) setting.	×	0.04
Gen-FedSD reduces the model size from 794.42 MB to 22.69 MB for the Vanilla Scaffold method under Dir (0.5) setting.	×	0.04

### References

- <http://arxiv.org/abs/2506.02887v2>
- <http://arxiv.org/abs/2512.20329v1>
- <http://arxiv.org/abs/2405.07925v1>