

Federated Data Heterogeneity and Few-Shot Code Generation Performance in LLMs

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does data heterogeneity in federated learning affect the few-shot code generation performance of LLMs (e.g., CodeLlama) on HumanEval, measured by pass@1 and pass@10 scores under varying degrees. Federated Learning (FL) enables decentralized training of machine learning models on distributed data while preserving privacy. However, in real-world FL settings, client data is often non-identically distributed and imbalanced, resulting in statistical data heterogeneity which. 5 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: FedDiverse: Tackling Data Heterogeneity in Federated Learning with Diversity-Driven Client Selection. Research question: How does data heterogeneity in federated learning affect the few-shot code generation performance of LLMs (e.g., CodeLlama) on HumanEval, measured by pass@1 and pass@10 scores under varying degrees of client dropout?.

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 3.5/10.

3 Results

9 papers retrieved. 5 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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
In standard FEDAVG, the parameter aggregation is performed by computing the weighted mean: $\theta_{t+1} = \sum_{k \in K} \frac{n_k}{n} \theta_{t+1}^k$.	×	0.03
Statistical data heterogeneity emerges when there is a subpopulation shift, i.e., when the representation of subpopulations	×	0.10
Class Imbalance (CI) occurs when the distribution of the target labels y is different between the training and test distributions	×	0.04
Attribute Imbalance (AI) occurs when the probability of occurrence of a certain attribute a in the training set is much	×	0.04
Spurious Correlation (SC) is a statistical dependency between attributes and target labels that does not hold in the test	×	0.03

References

- <http://arxiv.org/abs/2206.05507v1>
- <http://arxiv.org/abs/2504.11216v2>
- <http://arxiv.org/abs/2506.02887v2>