# Scaling Laws and Error Patterns in Multimodal Model Performance

Assignee Research

June 8, 2026

**Abstract**

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What is the effect of model size scaling on error patterns in multimodal models, as measured by task-specific accuracy and failure mode distributions. 11 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Scaling Law Analysis in Federated Learning: How to Select the Optimal Model Size?. Research question: What is the effect of model size scaling on error patterns in multimodal models, as measured by task-specific accuracy and failure mode distributions?.

## 2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.3/10.

## 3 Results

4 papers retrieved. 11 claims extracted; 3 independently verified. Quality review score: 5.3/10.

## 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 |
|---|---|---|
| Estimating the optimal model size in federated scenarios should depend on the average training compute across clients. | ✓ | 0.33 |
| The study empirically validates results with extensive training runs on different models, network settings, and datasets | × | 0.06 |
| Clients generally have fewer computational resources than the server in federated learning scenarios. | × | 0.09 |
| Most existing works on applying FL to large-scale models follow the intuition to reduce model size by tailoring architec | × | 0.12 |
| Few studies have explored the modified scaling behavior of language models in federated scenarios prior to this work. | × | 0.09 |
| Previous studies on federated scaling behavior primarily offered observational insights based on empirical evidence. | × | 0.05 |
| The authors model federated training as an SGD optimization problem over distributed data. | × | 0.09 |
| The authors derive an analytic solution for the optimal model size to quantify the impact on scaling. | ✓ | 0.22 |
| Stochastic gradient descent (SGD) is a widely used optimization method in machine learning. | × | 0.09 |
| The generalization performance of stochastic algorithms can be quantified using a PAC-Bayes upper bound. | ✓ | 0.17 |
| PAC-Bayes upper bounds have been applied to explore algorithm convergence. | × | 0.06 |

## References


- http://arxiv.org/abs/2603.21389v1
- http://arxiv.org/abs/2204.12724v2
- http://arxiv.org/abs/2511.12188v1