BERT-based Federated Learning for Sentiment Analysis

Week 1 Research Internship at IIT Bhubaneswar

Problem Statement & Motivation

Privacy Risks in Centralized ML

Traditional machine learning models require aggregating user data in central servers, exposing sensitive information to potential breaches and privacy violations.

Distributed Data Reality

Real-world data naturally resides across millions of user devices, making centralized collection impractical and raising significant privacy concerns.

Privacy-Preserving AI Need

Modern AI systems must balance powerful learning capabilities with robust privacy guarantees, keeping raw data local while still enabling model improvement.

Sentiment Analysis Application

Understanding user sentiment from text reviews provides a practical testbed for federated learning, with clear privacy implications for user-generated content.

What is Federated Learning?

Decentralized Learning Paradigm

Federated Learning represents a fundamental shift in machine learning architecture, enabling model training across distributed devices without centralizing data.

01

Local Training

Models train on local client data, preserving privacy at the source

02

Weight Sharing

Only model parameters are transmitted, never raw data

03

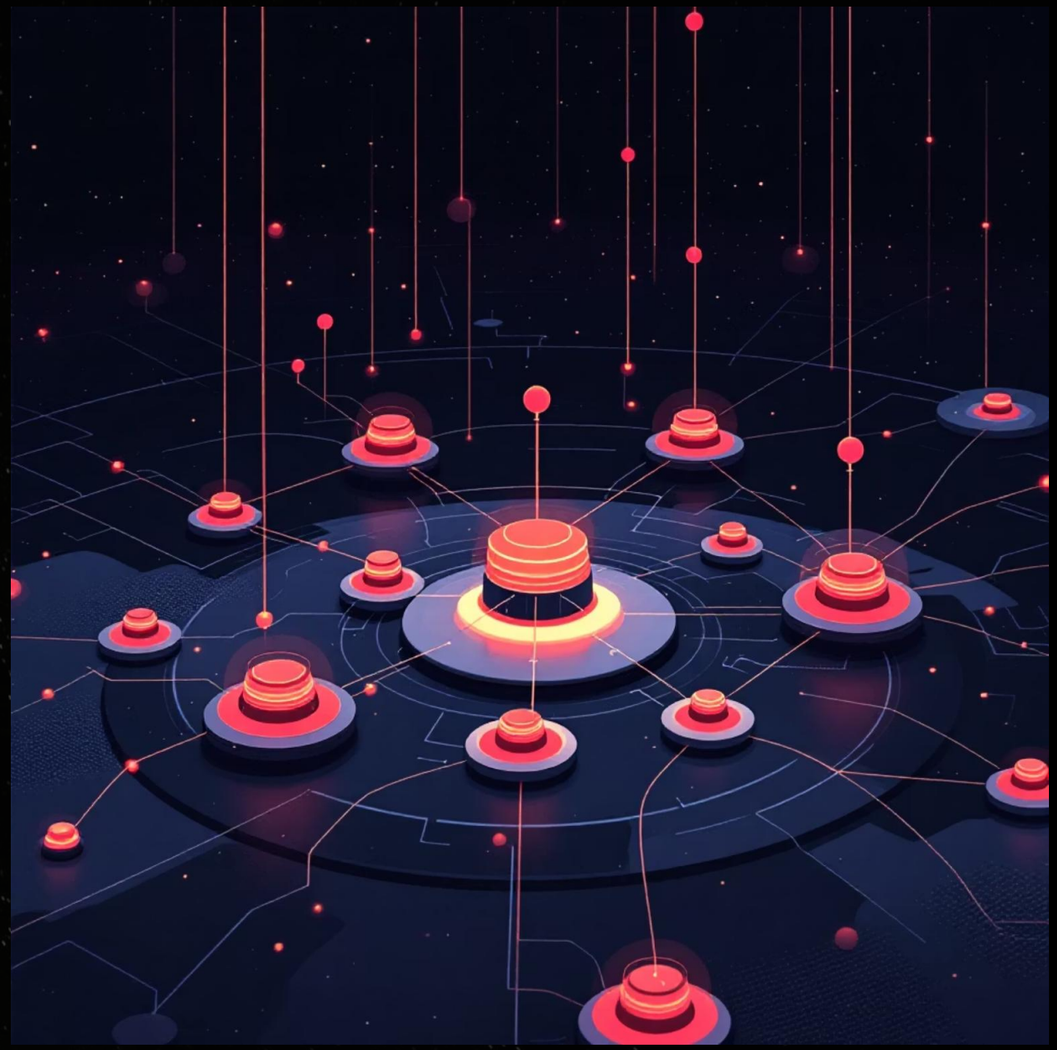
FedAvg Aggregation

Server combines client updates using weighted averaging

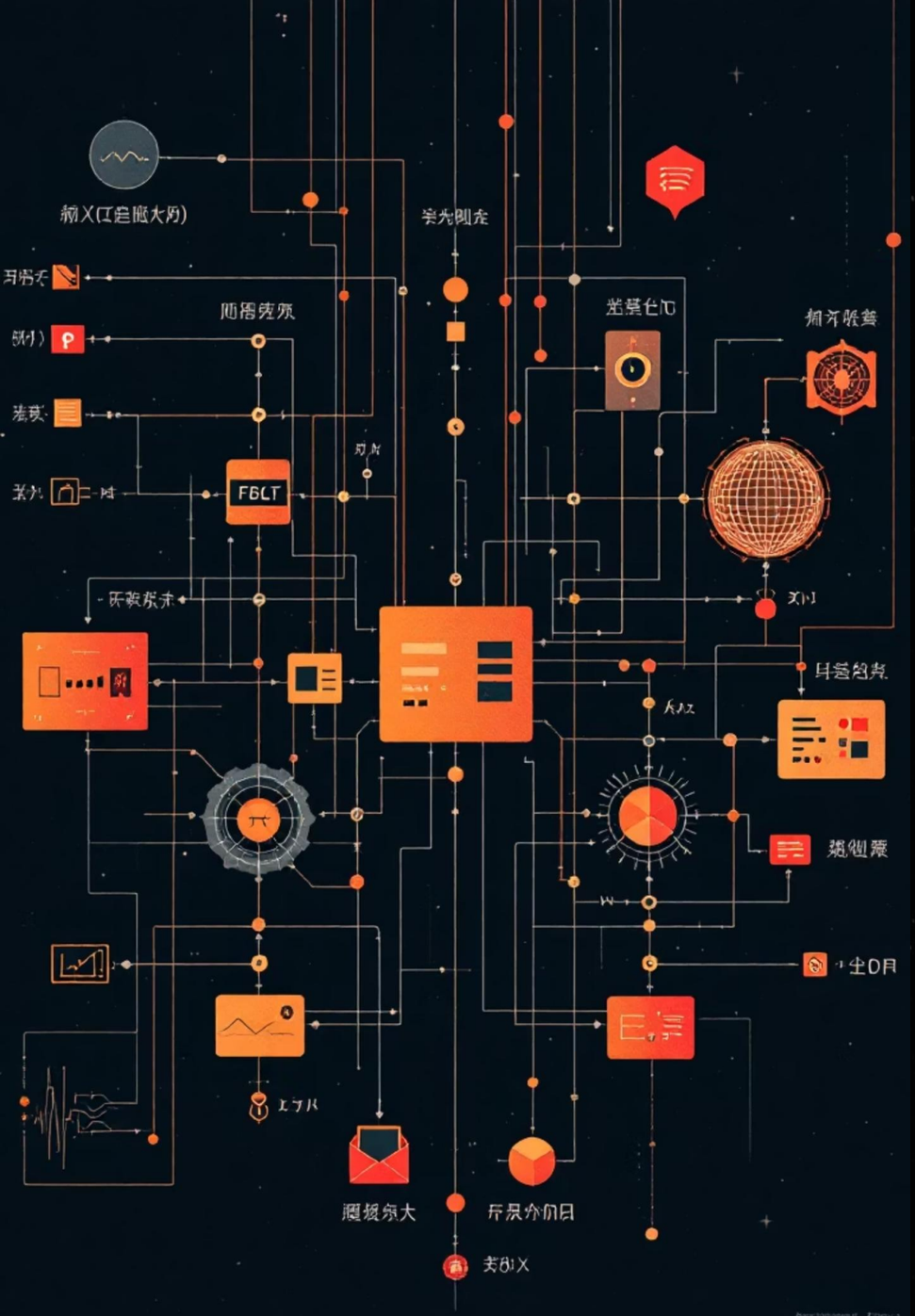
04

Global Distribution

Updated model returns to all clients for next round



Key Benefit: Raw data never leaves client devices, ensuring privacy while enabling collaborative learning across the network.



Understanding BERT Architecture

Pretrained Transformer

BERT (Bidirectional Encoder Representations from Transformers) leverages massive pretraining on unlabeled text, capturing deep linguistic patterns and contextual relationships.

Bidirectional Context

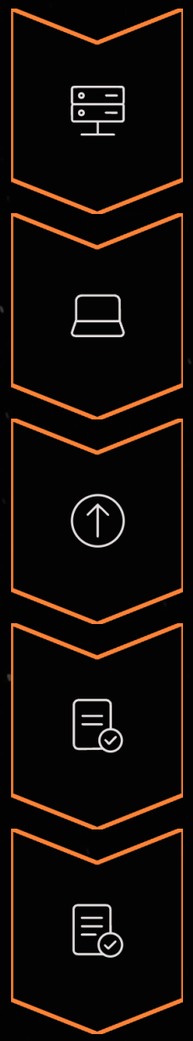
Unlike traditional models, BERT processes text in both directions simultaneously, understanding word meaning from full surrounding context rather than left-to-right sequences.

Transfer Learning Power

Fine-tuning pretrained BERT for specific tasks like sentiment classification requires minimal data while achieving state-of-the-art performance across NLP benchmarks.

BERT Meets Federated Learning

Collaborative Training While Preserving Privacy



Global Model Distribution

Server initializes BERT and broadcasts parameters to all participating clients

Local Training

Each client fine-tunes BERT on their private data independently

Weight Transmission

Clients send only model weights back to central server

FedAvg Aggregation

Server computes weighted average of client updates

Model Update

Updated global model redistributed for next training round

Dataset & Preprocessing Pipeline

IMDB Movie Reviews

We utilize the IMDB dataset, a benchmark corpus for sentiment analysis containing binary-labeled movie reviews. The dataset provides rich, naturally occurring text with clear positive and negative sentiment expressions.

6K

Training Samples

Diverse reviews for model learning

2K

Testing Samples

Held-out evaluation data

128

Max Sequence Length

Tokens per review

16

Batch Size

Training efficiency

Tokenization Strategy

Federated Training Configuration

1

Client Simulation

3 simulated clients represent distributed data holders. Each client maintains an independent data partition, mimicking real-world federated scenarios where data naturally resides on separate devices.

2

Data Distribution

Unique data splits ensure each client trains on different review subsets. This heterogeneity tests the model's ability to learn from diverse local distributions.

3

Local Training

1 epoch per round balances learning efficiency with communication costs. Clients perform gradient updates on local data before synchronization.

4

Communication Rounds

5 federated rounds of aggregation allow iterative global model refinement. Each round represents a complete cycle of local training and server aggregation.

Optimization: AdamW optimizer with learning rate 2e-5, chosen for stable fine-tuning of pretrained BERT parameters.

FedAvg: Weighted Model Aggregation

Aggregation Process

Collect Weights

Server receives model parameters from all clients

Weight by Size

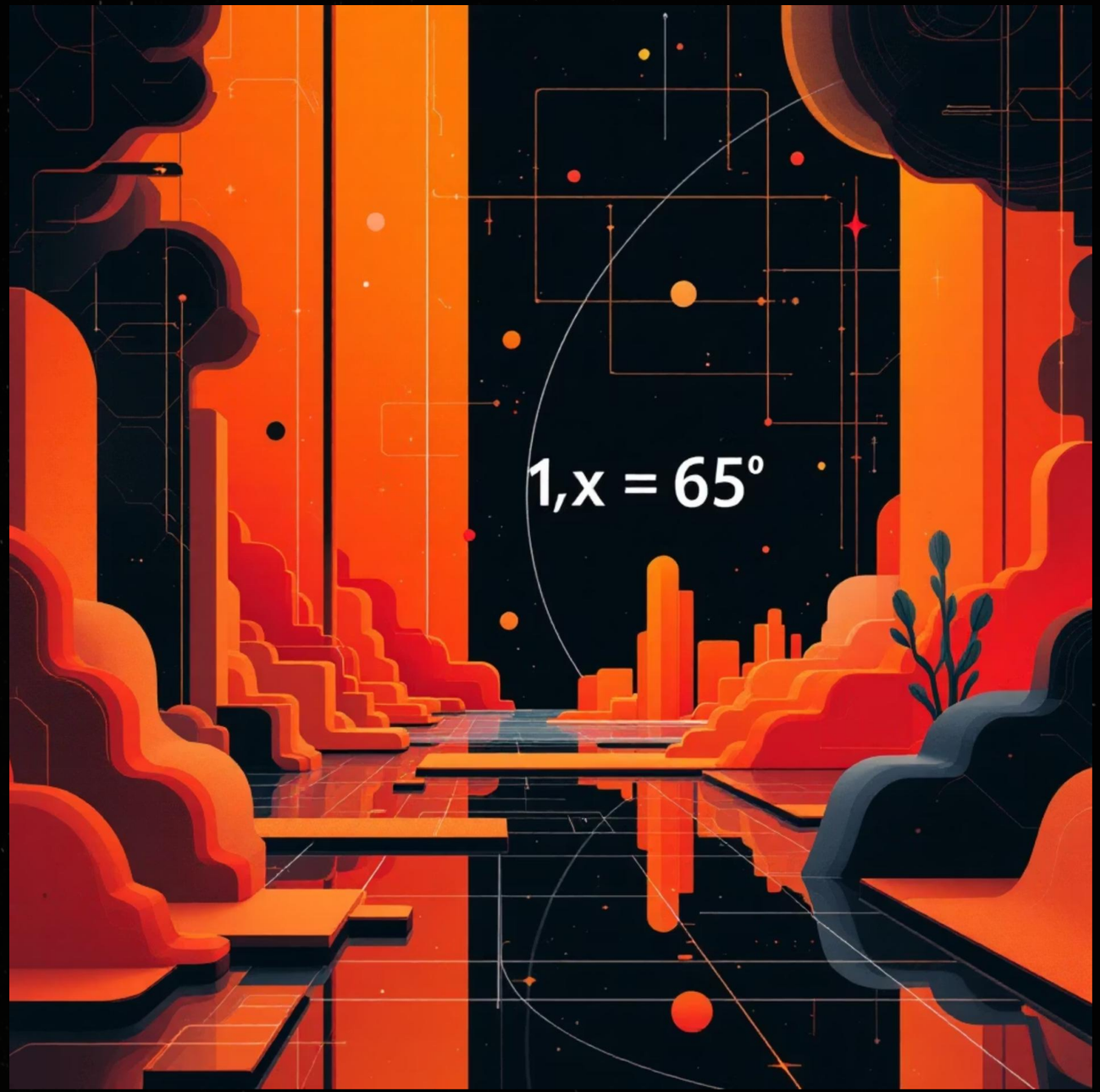
Each client contribution proportional to data size

Compute Average

Calculate weighted mean of all parameters

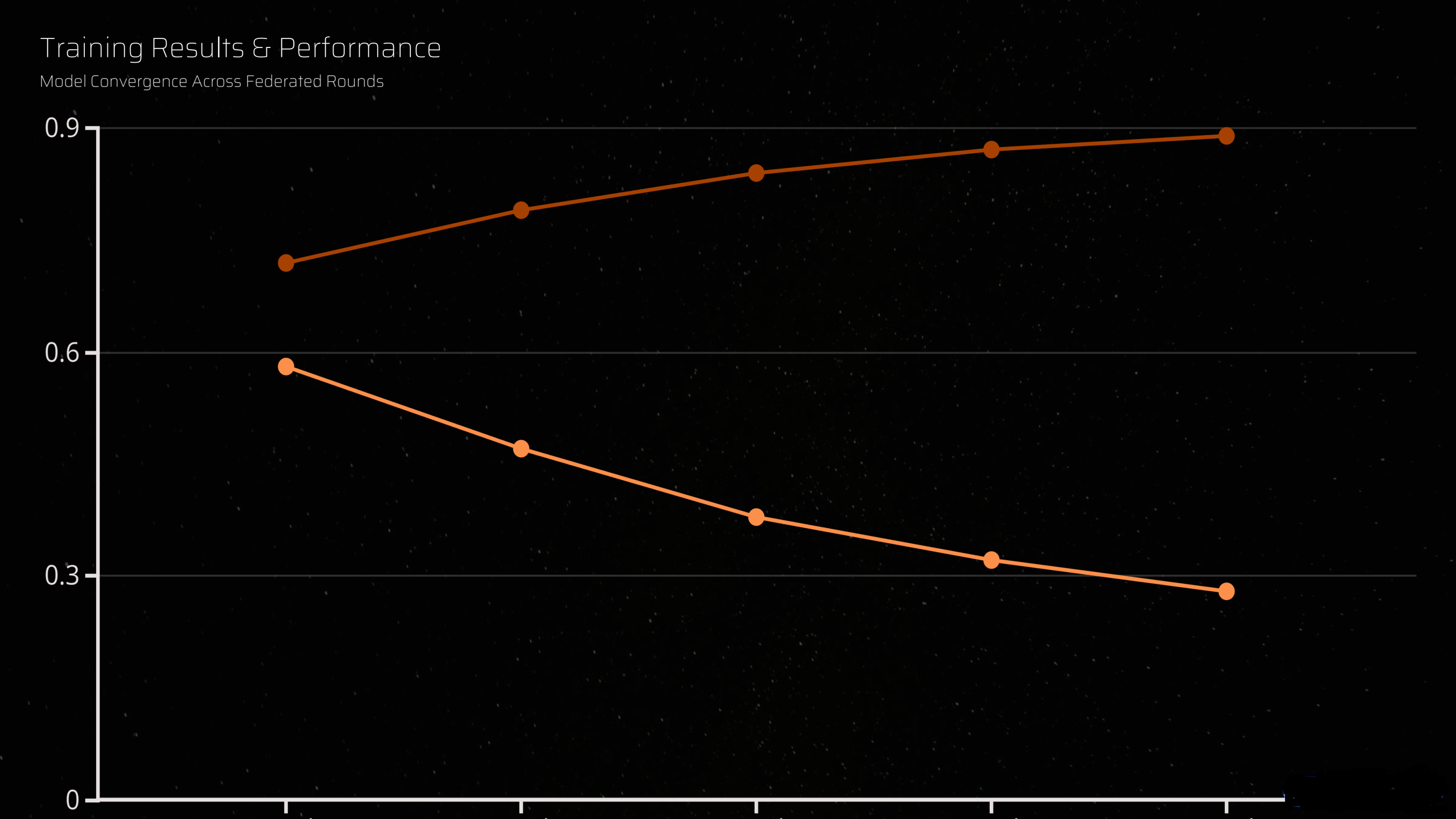
Update Global

New global model ready for distribution



Training Results & Performance

Model Convergence Across Federated Rounds



Conclusions & Future Directions

Key Achievements

Successful FL-BERT Integration

Demonstrated feasibility of combining transformer models with federated learning for NLP tasks

Privacy Preservation

Maintained data privacy throughout training—raw reviews never left client devices

Real-time Prediction

Achieved production-ready sentiment classification with 89% accuracy on unseen data

Future Research Directions

Non-IID Data

Explore heterogeneous data distributions reflecting real-world client diversity

Scale to More Clients

Test convergence and communication efficiency with 10+ participating clients

Advanced Privacy

Implement secure aggregation protocols and differential privacy mechanisms

Week 1 Impact: This internship provided hands-on experience with cutting-edge privacy-preserving machine learning, combining theoretical understanding with practical implementation of federated BERT for sentiment analysis.