Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Advanced Multimodal Federated Learning: A Privacy-Focused Approach to Social Network Systems

Guide: Dr. Kriti Srivastava

Candidates

Aryan Rajpurkar: 60009220144

Aaditya Malani: 60009220192

Advait Sankhe: 60009220024

Krish Jain : 60009210111

Date:
October 25, 2024

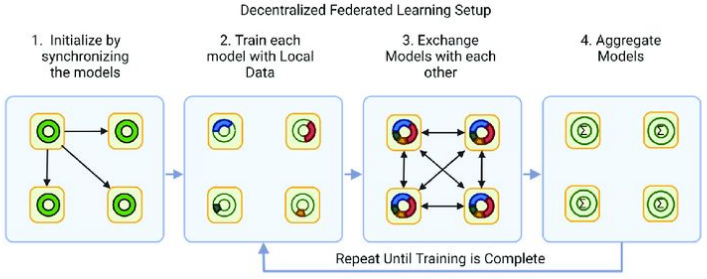
Introduction

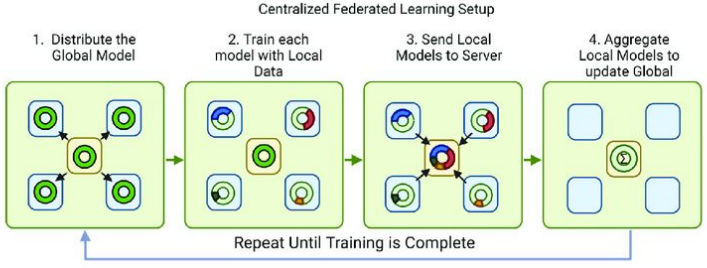
Need for Privacy-Preserving Recommendations (NEED)

- **Data Privacy Concerns:** Rising incidents of data breaches and stringent regulations (e.g., GDPR, CCPA) necessitate secure data handling.
- **User Trust:** Enhancing user confidence by safeguarding personal interactions and content.

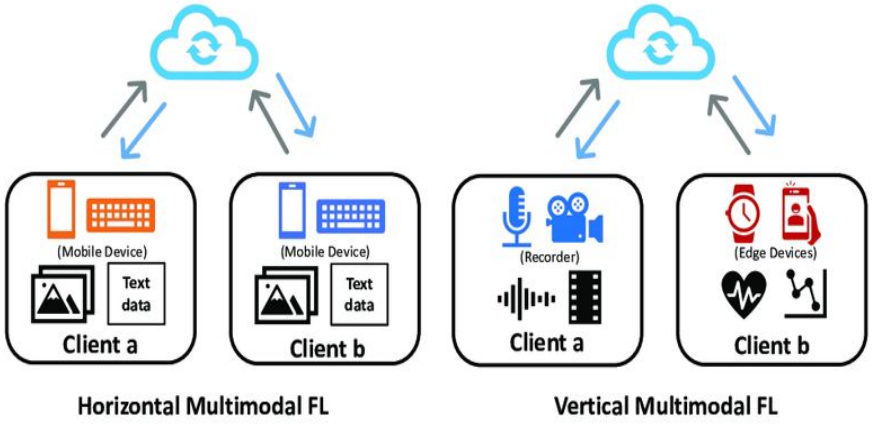
Impact of Federated Multimodal Learning in Social Media (USE CASE)

- **Personalized Experience:** Integrates diverse data types (images, text, interactions) for more accurate and relevant recommendations.
- **Enhanced Privacy:** Keeps user data localized on devices, minimizing exposure and potential misuse.
- **Scalability:** Supports large-scale deployments across millions of users without centralized data bottlenecks.





Centralized vs Decentralized FL Model System



Reference

Horizontal multimodal federated learning involving two clients. Both hold image and text data.

The vertical multimodal federated learning example includes two clients with exclusive modalities. Client a has audio and video data, while client b holds heat rate and acceleration sensor data.

Sr. No.	Name of the Paper	Conference/ Journal	Algorithm	Gap Analysis
1	Federated Contrastive Learning and Visual Transformers for Personal Recommendation “Uses clustering approach like our motive”	Cognitive Computation (2024)	• Federated learning • Graph NN (to represent users) • Transformers • k-means algorithms. • Angular Contrastive Loss • KGAT • KAUR (for recommendation model) • FLT-PR (useful for our use case)	• Focuses primarily on consumer electronics and lacks integration of **multimodal data (image and text fusion)** essential for social media platforms. • The clustering approach used does not address **real-time dynamic user behavior** prevalent on social media. • **Privacy mechanisms are limited to a trusted authority**, without exploring stronger privacy-preserving techniques like differential privacy in multimodal federated learning.
2	Adaptive Federated Learning in Resource-Constrained Edge Computing Systems	IEEE 2019	Federated learning with adaptive gradient descent	• Primarily addresses **resource constraints** without solving **privacy challenges.** • Lacks focus on **multimodal data integration**

3	Federated learning enabled graph convolutional autoencoder and factorization machine for potential friendship prediction in social networks (The model recommends friends to users that best match their personality.)	Information Fusion (2024)	• Douban and FilmTrust social network datasets • BayDNN • GCAFM model.	●Focuses solely on **graph-based data**, which limits its ability to handle **multimodal information** ●**Does not integrate NLP** or image auto-tagging, which are essential for richer recommendation ●The model focuses on **friendship prediction**
4	DFedSN: Decentralized federated learning based on heterogeneous data in social networks BEST PAPER	Springer 2023 (WWW)	• Karate Dataset (Social Networks) • Decentralized federated learning • Algorithm defined in paper • AlexNet, Inception-Net and Mini-ResNet	• Although it addresses **data heterogeneity**, the approach focuses on **image-based classification** and lacks **multimodal feature fusion** • Does not explore advanced **neural network architectures**
5	A Federated Learning-Enabled Predictive Analysis to Forecast Stock Market Trends	Journal of Ambient Intelligence and Humanized Computing (2023)	• Federated learning • Random Forest • Support Vector Machines • Linear Regression	• Focuses on **numerical stock market data**, not suitable for **multimodal data fusion** • Does not tackle **privacy-preserving multimodal feature**

6	Continual Horizontal Federated Learning for Heterogeneous Data	IEEE International Joint Conference on Neural Network	FedAvg: Federated Averaging CFHL : built on FedAvg PNN (Progressive Neural Network)	Collaborative learning with client-specific data can lead to privacy vulnerabilities, especially when exchanging intermediate data.
7	Vertical Federated Learning: Concepts, Advances and Challenges	IEEE	VFLow (Unified framework for VFL) FedCVT	AutoML in VFL faces challenges due to encryption and collaborative training settings, especially for participants without local labels.

8	A Comprehensive Survey of Federated Transfer Learning: Challenges, Methods and Applications	Springer	FedProx FedAvg PNN HFTL	Existing FTL studies do not adequately address scenarios where both label and feature spaces are heterogeneous across participants, limiting FTL's utility in complex real-world applications
9	A Survey on Federated Learning Systems: Vision, Hype, and Reality for Data Privacy and Protection	arXiv	FedAvg decentralized algorithms SimFL.	• Lack of comprehensive federated learning systems for addressing both effectiveness and privacy concerns. • Challenges in scalability and non-IID data handling.
10	Models of Privacy and Disclosure on Social Networking Sites	MDPI	Structural Equation Models SNS Algorithm	Limited geographic scope in studies about privacy and disclosure behaviors. More research needed on SEM application for diverse SNS platforms.

11	Federated Hyperparameter Tuning: Challenges, Baselines, and Connections to Weight-Sharing	ACM	Successive Halving Algorithm (SHA) FedEx	The paper mentions that they do not address the time-dependency challenge in federated evaluation. Exploring approaches to optimize hyperparameter tuning while considering time-dependent changes in client data is a potential research area that can be explored.
12	FedBully: A Cross-Device Federated Approach for Privacy Enabled Cyber Bullying Detection using Sentence Encoders	Journal of Cyber Security and Mobility (2023)	• Federated learning • Secure aggregation • Sentence encoders • Dense networks • Natural Language Processing (NLP)	Focuses solely on text-based cyberbullying detection and lacks integration of other data types like images or videos for comprehensive cyberbullying analysis. Limited exploration of advanced NLP techniques beyond sentence encoders.
13	FS-Real: Towards Real-World Cross-Device Federated Learning	(Published by Alibaba Group	• Federated learning • Communication compression • Asynchronous aggregation, • Personalization	Focuses on heterogeneous device scenarios but lacks exploration into privacy-preserving methods like differential privacy in heterogeneous settings. Limited support for multimodal data.

14	TrustFed: A Framework for Fair and Trustworthy Cross-Device Federated Learning in IIoT	IEEE Transactions on Industrial Informatics (2021)	• Blockchain-enabled federated learning • Statistical outlier detection • Ethereum smart contracts	Focuses heavily on IIoT and security using blockchain for reputation and fairness. Lacks advanced multimodal feature fusion and doesn't address data heterogeneity or privacy mechanisms for multimodal data, especially in social networks.
15	Breaking the Centralized Barrier for Cross-Device Federated Learning	NeurIPS 2021	• MIME (Mimicking Centralized Stochastic Algorithms) • Momentum-based variance reduction (MVR) • Server-level optimizer	Primarily focuses on addressing client drift and optimizing communication in cross-device settings. Does not explore multimodal data integration or privacy mechanisms (e.g., differential privacy), which are crucial for applications in social networks
16	Fairness and accuracy in horizontal federated learning	Information Sciences 589 (2022)	• FedFa: Double momentum gradient • Information quantity-based weighting strategies	Focuses on statistical heterogeneity and communication overhead in horizontal federated learning. Does not explore multimodal feature fusion or advanced privacy mechanisms (e.g., differential privacy) that are essential for social media platforms

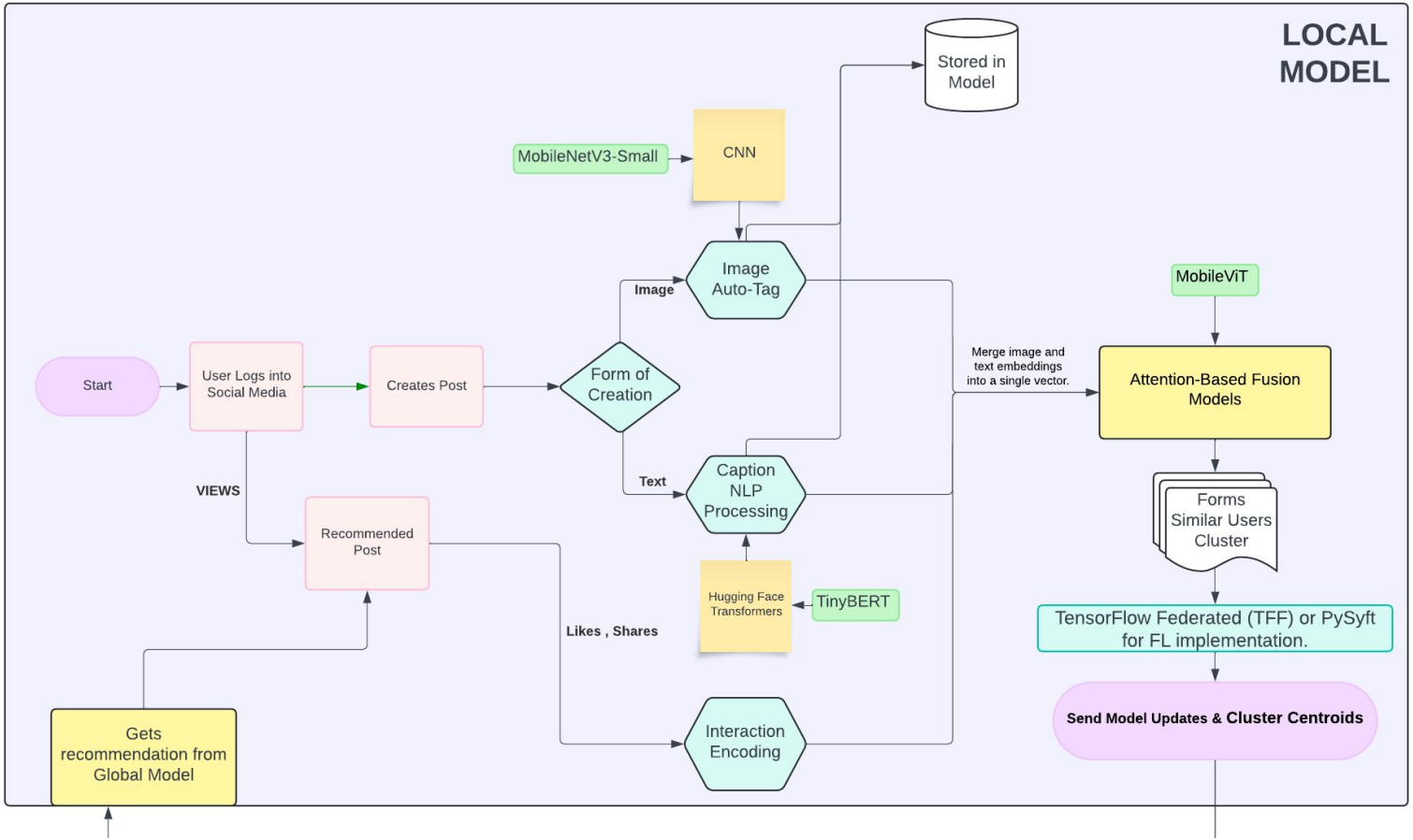
Gaps in existing survey	Gap Resolution
● Lack of **heterogeneous data**: Most papers focus on either single data types (text, images, or graphs) rather than using **multiple form of data** in federated learning.	● We will do that using all types of data **using Attention Layers and distributing weights depending on the context.**
● Incomplete **privacy mechanisms** for federated learning: Current studies lack robust privacy measures when dealing with multimodal data.	● We will introduce **differential privacy specific to multimodal data** in Federated Learning.
● No research has been done regarding benefits of distributed model learning in field of social networks.	● We will do it!

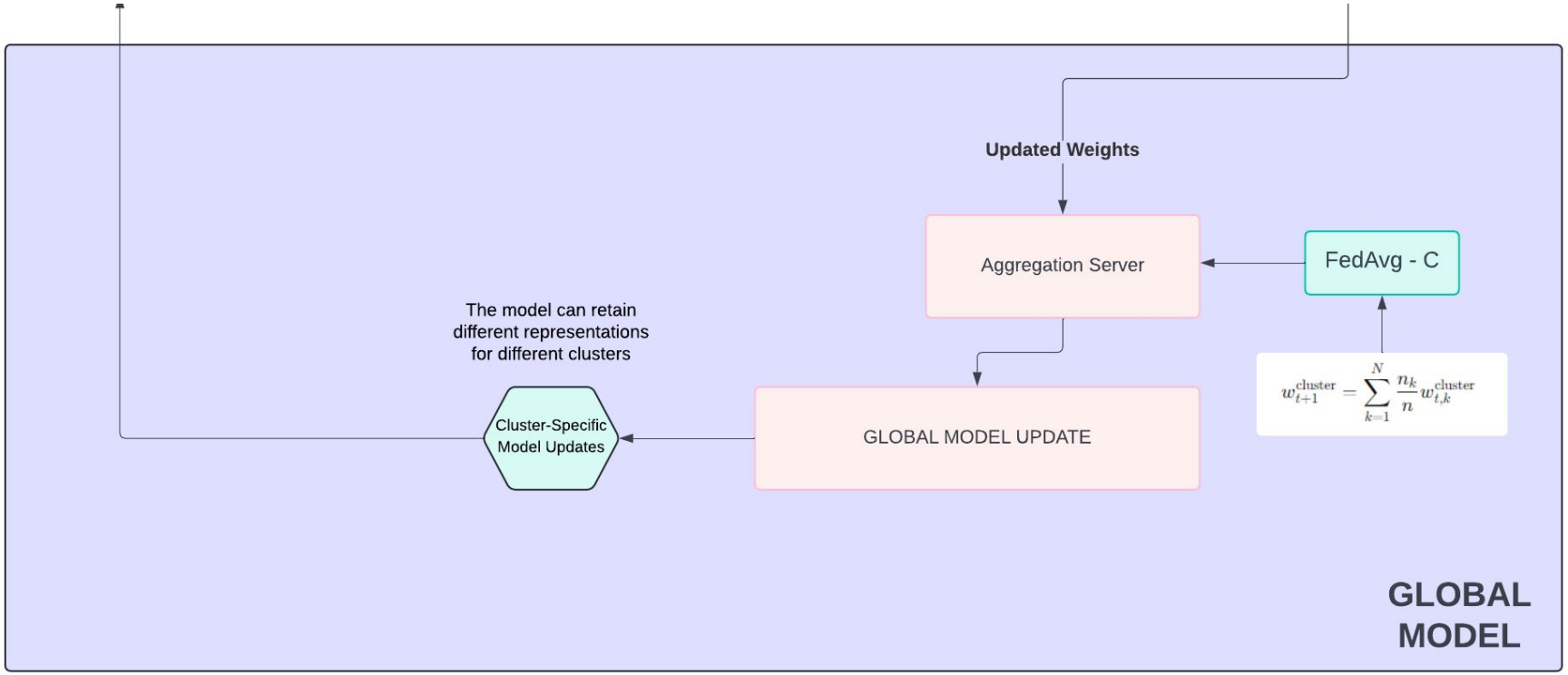
Problem Definition

This project aims to address the challenges of integrating **heterogeneous multimodal data** (images, text, and user interactions) and in a centralized **federated learning-based recommendation** for social media network systems.

By implementing Continual Horizontal Federated Learning (CHFL) alongside advanced multimodal feature fusion techniques using specialized neural networks

Proposed Design




Updated Weights
Aggregation Server
FedAvg - C
$w_{t+1}^{\text{cluster}} = \sum_{k=1}^{N} \frac{n_k}{n} w_{t,k}^{\text{cluster}}$
The model can retain different representations for different clusters
Cluster-Specific Model Updates
GLOBAL MODEL UPDATE
GLOBAL MODEL

Novelty

Integration of Heterogeneous Multimodal Data:

- Existing systems primarily focus on single-modal data types (e.g., text, images, or graphs), lacking a unified approach for multimodal integration.

Use of Attention Model with Clustering Approach

- Unlike traditional clustering methods that treat all features equally, the attention mechanism dynamically assigns **context-aware weights to different modalities** (e.g., images, text, interactions), emphasizing the most relevant data.
- This combination allows for the formation of personalized and adaptive clusters, reflecting real-time user behavior

Data Design

Local

```
{
  "post": {
    "post_id": "unique_post_identifier",
    "user_id": "unique_user_identifier",
    "image_embedding_vector": "vector_representation_here",
    "text_embedding_vector": "vector_representation_here",
    "clustering_metadata": {
      "user_preferences": ["travel", "photography"],
      "interaction_score": 0.75,
      "engagement_pattern": {
        "like_rate": 0.70,
        "share_rate": 0.20,
        "save_rate": 0.10
      }
    }
  }
}
```

Global

```
{
  "local_model_update": {
    "weights": "local_weight_parameters",
    "clustering_info": {
      "cluster_id": "cluster_12345",
      "centroid_vector": "centroid_representation_here"
    }
  }
}
```

Initial Dataset

Open Images Dataset - Google

- **Specialties**: Contains over 9 million images with rich annotations for object detection, image classification, and segmentation tasks. Ideal for training models that require diverse visual content and metadata.

Flickr API

- **Specialties**: Provides access to tagged images and extensive metadata by category keywords. Suitable for extracting real-world social media-like interactions, geolocation data, and user-generated tags.

YFCC100M - Yahoo Dataset (100M Images)

- **Specialties**: Comprehensive dataset with 100 million images and videos accompanied by extensive metadata

Plan for Semester VI

- Completion of a fully functional Social Media Platform supported by Federated Learning.
- Scaling of prototype to decentralization FL.
- Incorporate video analysis as well within local model.

Text Processing Architecture

Specification	TinyBERT
Model Size	10 MB
Parameters	~6 million
RAM Usage	30 MB to 50 MB
CPU Usage	~5% to 20%
Power	0.5W to 1W
Inference	Fast (optimized

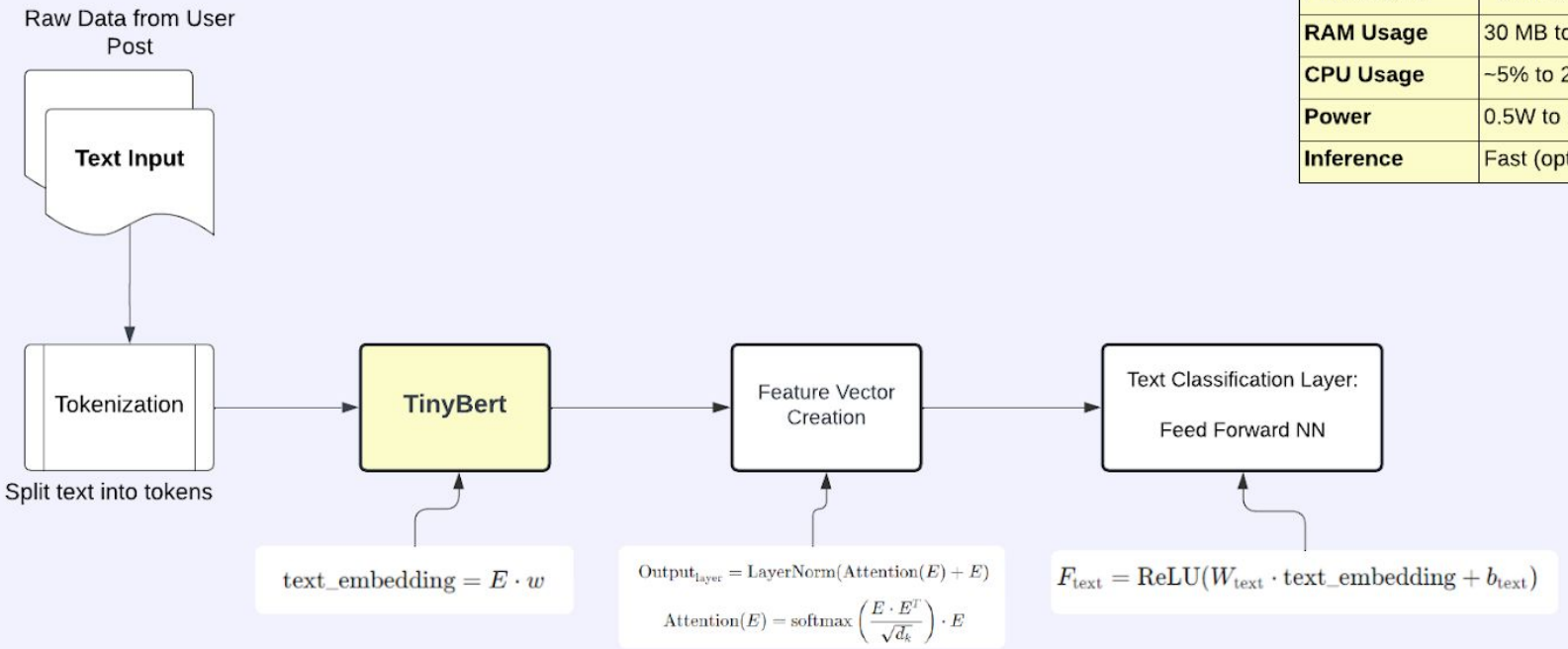


Image Processing Architecture

Specification	MobileNetV3-Small
Model Size	~4 MB
Parameters	~2.5 million
RAM Usage	40 MB to 70 MB
CPU Usage	5% to 20%
Power	0.5W to 1W
Inference Speed	Fast, optimized for low-power devices

User Input Image

Image Preprocessing

• Resize to 224 x 224 pixels

MobileVNetV3 Small

$$F = \text{ReLU}(W_{\text{cnn}} * I_{\text{norm}} + b)$$

Wcnn -Convolutional filter weights

Uses extracted features for image classification and tagging.

Auto Tagging Layer

$$\text{tags} = \text{softmax}(W_{\text{tag}} \cdot F + b_{\text{tag}})$$

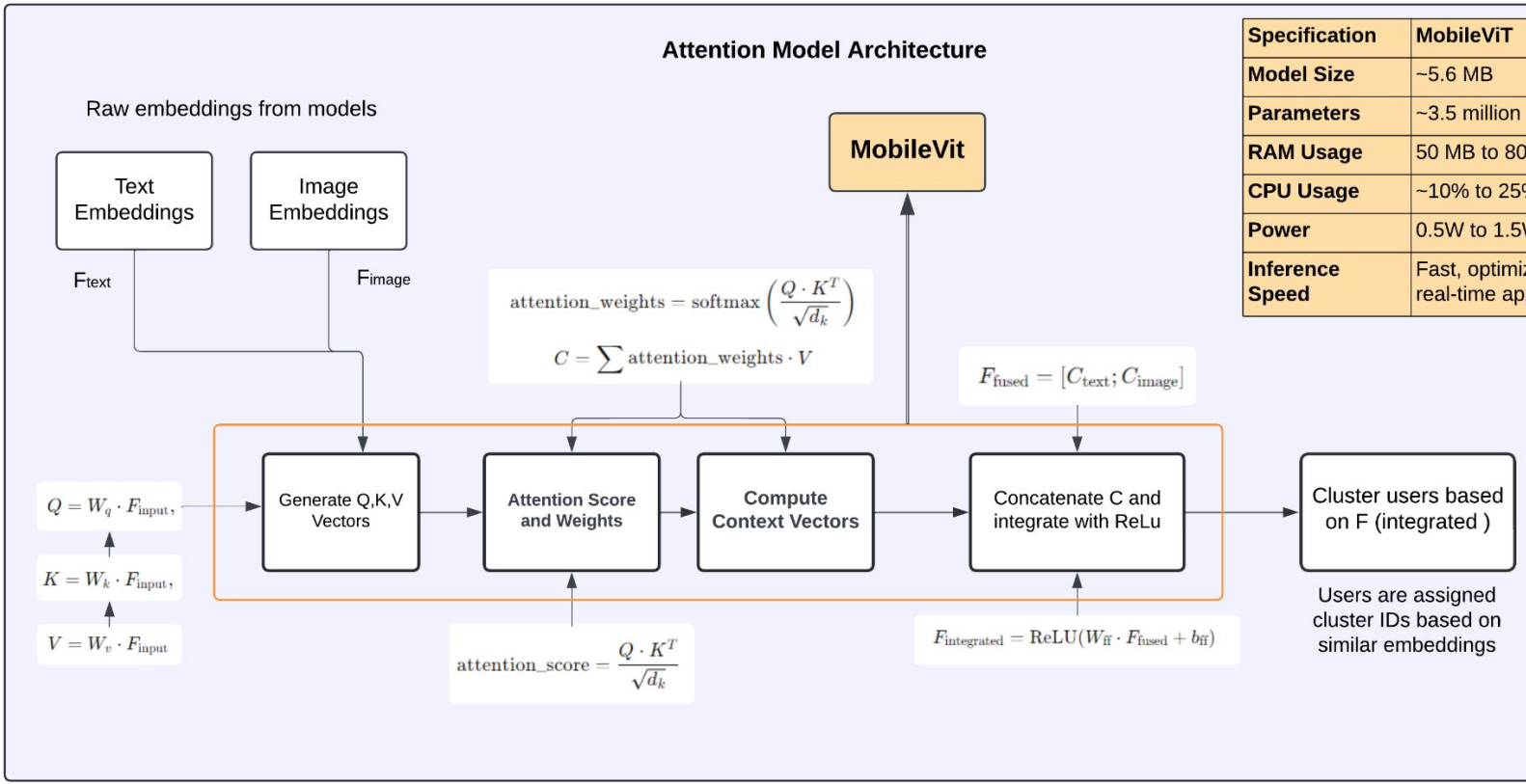
Feature Map flattened to 1D Vector

Image Embedding Creation

$$F_{\text{image}} = \text{Flatten}(F)$$

Fit for Integration with text embeddings

Attention Model Architecture



Specification	MobileViT
Model Size	~5.6 MB
Parameters	~3.5 million
RAM Usage	50 MB to 80 MB
CPU Usage	~10% to 25%
Power	0.5W to 1.5W
Inference Speed	Fast, optimized for real-time applications

References

[1] - Belhadi, A., Djenouri, Y., de Alcantara Andrade, F.A. *et al.* Federated Constrastive Learning and Visual Transformers for Personal Recommendation. *Cogn Comput* **16**, 2551–2565 (2024). https://doi.org/10.1007/s12559-024-10286-0

All papers referenced above :

https://drive.google.com/drive/folders/1UWNDpfv6YCDCtyxrTD16dy53btBP_NtH?usp=sharing