Multilingual Cultural Metaphor Transfer Driven by Federated Learning

Lei Shi
School of Foreign Studies, Shandong University of Finance and Economics, Jinan 250014, Shandong, China
Corresponding author: Lei Shi (e-mail: shileisdufe@163.com)

ABSTRACT Traditional federated learning faces significant challenges in multilingual cultural metaphor transfer because nonindependent and identically distributed language data make it difficult to balance local cultural heterogeneity with global semantic consistency. Privacy protection mechanisms may further amplify cultural cognitive bias and reduce transfer quality. To address these issues, this paper proposes a culturally aware personalized federated learning architecture. Based on Hofstede's cultural dimension theory, cultural feature vectors are constructed, cultural weight functions are defined, and a weighted aggregation strategy is introduced to account for variations in power distance and individualism across clients. Security is enhanced through differential privacy and secure aggregation mechanisms. Experimental results show that, compared with FedAvg, the proposed method improves cross-language metaphor recognition precision and F1-score by 7.4% and 7.1%, respectively. Cultural consistency reaches 0.82, transfer comprehensibility reaches 4.02 , and user matching degree reaches 0.79, while the minimum back-inference attack success rate is reduced to 0.5%. The results confirm that the proposed architecture enhances multilingual semantic transfer, cross-cultural consistency, and privacy-preserving collaborative learning.

INDEX TERMS multilingual culture, metaphor transfer, federated learning, cultural awareness, privacy protection

I. INTRODUCTION

WITH the acceleration of globalization and the development of artificial intelligence, cross-language cultural exchanges are becoming increasingly frequent, particularly in the industry, where global sourcing, design collaboration, and marketing across diverse consumer bases necessitate seamless and culturally nuanced communication [1,2]. How to achieve metaphor transfer between multilingual cultures while protecting data privacy has become an important challenge in the field of natural language processing and cultural communication [3,4]. Metaphor is the core carrier of language expression and cultural cognition. Its transfer not only involves the transformation of semantic level, but also concerns the deep mapping of cultural background, values and social experience. Cultural metaphor transfer is the process of converting metaphorical expressions in the source language into the target language in cross-language tasks, while retaining its semantic essence and adapting it to the cultural cognitive framework and expression habits of the target language. As an emerging distributed machine learning paradigm, federated learning has been widely used in multilingual tasks because of its advantages of supporting multi-party collaborative modeling and not sharing original data [5,6]. However, when it comes to the understanding and transfer of cultural metaphors, federated learning faces the high heterogeneity of multilingual cultural data, which makes it difficult for the model to unify the semantic space. The privacy protection mechanism may also aggravate cultural cognitive biases, resulting in distorted metaphor transfer results [7,8]. Therefore, there is an urgent need for a new federated learning architecture that takes into account both cultural difference modeling and privacy protection to promote the practical development of cross-language metaphor transfer. This paper focuses on the issues of semantic consistency and personalized modeling in multilingual cultural metaphor transfer. Existing studies have shown that Phani Krishna P [9] showed that bilinguals have considerable advantages in cognitive functions and are more capable of generating novel metaphors than monolinguals; Qiao R [10] explored the challenges of cross-language sentiment analysis, including the impact of language and cultural differences on sentiment recognition; Hong W [11] aimed to clarify the contribution of cognitive methods to translation research by critically reviewing recent studies on metaphor translation from a cognitive perspective; Sheng A [12] found that human translators are able to handle political and ideological factors more skillfully, while the working mechanism of the neural network machine translation system lacks the judgment, thinking ability, flexibility and subjectivity of human translators; Shima T [13] investigated

the impact of cultural background adjustment on metaphor effect, and the results showed that adjusting the cultural context is a factor in increasing the metaphor effect. The above research shows that metaphor transfer is not only a task at the language level, but also a deep mapping at the cultural cognitive level. It urgently needs a federated learning framework that can perceive and adapt to cultural characteristics to solve it. In recent years, many scholars have tried to improve the effect of cross-language metaphor transfer from different perspectives. Song Z [14] proposed a model that combines grammatical-aware local attention, a simple contrastive sentence embedding framework, and linguistic theory; Šajina R [15] used an encoder-only Transformer model for multi-task point-to-point learning for collaboration between agents that learn distinct natural language processing tasks; Ren J [16] proposed a framework based on graph neural networks to integrate cross-lingual word embeddings into the graph knowledge distillation process for detecting events in low-resource language data streams; Malan E [17] introduced federated learning with gradually frozen layers, which gradually reduced the part of model data sent back and forth, reducing the communication burden while maintaining the quality of training services; Wu X [18] proposed an optimized federated model with privacy protection and automatic learning rate adjustment functions. To address the challenges of cultural bias and heterogeneity in global AI models, particularly those affecting international collaboration and marketing, this paper proposes a novel culture-aware personalized federated learning architecture. This architecture aims to achieve efficient, secure, and culturally sensitive metaphor transfer, which is crucial for tasks like localizing fast-fashion trend predictions and ensuring accurate, culturally-appropriate product descriptions across diverse markets. Firstly, based on Hofstede's cultural dimension theory, this paper constructs a cultural feature vector and designs a cultural weight calculation function to introduce cultural sensitivity into the federation aggregation process. Secondly, combining contrastive learning and graph neural networks, this paper constructs a cross-language metaphor shared representation space, enhances the recognition ability of cross-language metaphor consistency through MSL (Multi-Similarity Loss), and uses GraphSAGE (Graph Sample and Aggregate) to achieve unified modeling of metaphor knowledge graphs. Thirdly, this paper designs a dynamic adapter module, which is inserted into the Transformer backbone network and only updates the adapter parameters to achieve efficient language adaptation while reducing communication overhead. Finally, this paper introduces differential privacy and security aggregation mechanism to inject controllable noise and encrypt it before uploading the gradient to ensure data privacy and prevent information leakage. This method innovatively introduces cultural dimension modeling into the federated learning framework, and improves the model's adaptability to multilingual cultural differences through a culturally aware weighted aggregation strategy; the design of a lightweight dynamic adapter significantly reduces communication costs; the metaphor shared space construction method that integrates contrastive learning and graph neural networks effectively improves cross-language semantic consistency and transfer effects; the overall average cultural consistency reaches 0.82, an increase of 17% compared with FedAvg, and realizes high-quality multilingual metaphor transfer under the premise of privacy protection.

II. ALGORITHM DESIGN

A. DESIGN OF CULTURAL-AWARE FEDERATED LEARNING ARCHITECTURE

This paper constructs a cultural-aware personalized federated learning architecture (CAP-FL), as shown in Figure 1. This architecture adopts the classic Client-Server distributed training structure. Each client holds metaphorical corpus in a single language and its corresponding cultural background, and the server is responsible for coordinating global model updates and maintaining a unified knowledge representation space [19,20]. Supposing there are N participating client sets $\mathcal{N} = \{1, 2, \ldots, N\}$ in the federated learning system, each client $i \in \mathcal{N}$ has a local dataset $D_i = \{(x_{i,j}, y_{i,j})\}_{j=1}^{|D_i|}$, $x_{i,j}$ is the input text, and $y_{i,j}$ is the corresponding metaphor label or metaphor expression. The local model parameter is recorded as θ_i, and the global model parameter is recorded as θ. In each round of federated training cycle t, the server broadcasts the current global model parameter θ^t to all clients. Each client performs local optimization on the model based on local data:

$$\theta_i^{t+1} = \arg\min_{\theta} \left(\frac{1}{|D_i|} \sum_{(x,y)\in D_i} l(x, y; \theta) + \lambda R(\theta) \right) \quad (1)$$

In formula (1), $l(\cdot)$ is the loss function, $R(\cdot)$ is the regularization term used to prevent overfitting, and λ is the regularization coefficient. The traditional federated average algorithm uses an equal-weight aggregation strategy to update the global model:

$$\theta^{t+1} = \sum_{i=1}^{N} \frac{|D_i|}{\sum_{j=1}^{N} |D_j|} \theta_i^{t+1} \quad (2)$$

In a multilingual cultural context, there are significant differences in the cultural dimensions carried by different languages. This paper introduces Hofstede's cultural dimension theory [21,22] and defines the cultural characteristic vector $c_i = [\mathrm{PDI}, \mathrm{IDV}, \mathrm{MAS}, \mathrm{UAI}, \mathrm{LTO}, \mathrm{IND}]^{\mathrm{T}}$ of the i-th client, which represents the six dimensions of power distance, individualism/collectivism, masculinity/femininity, uncertainty avoidance, long-term orientation, and indulgence/restraint. In order to measure the cultural influence of each client in federated learning, this paper proposes a cultural weight calculation function:

$$w_i = \alpha \cdot \frac{1}{Z} \exp\left(-\beta \cdot D(c_i, c)\right) \quad (3)$$

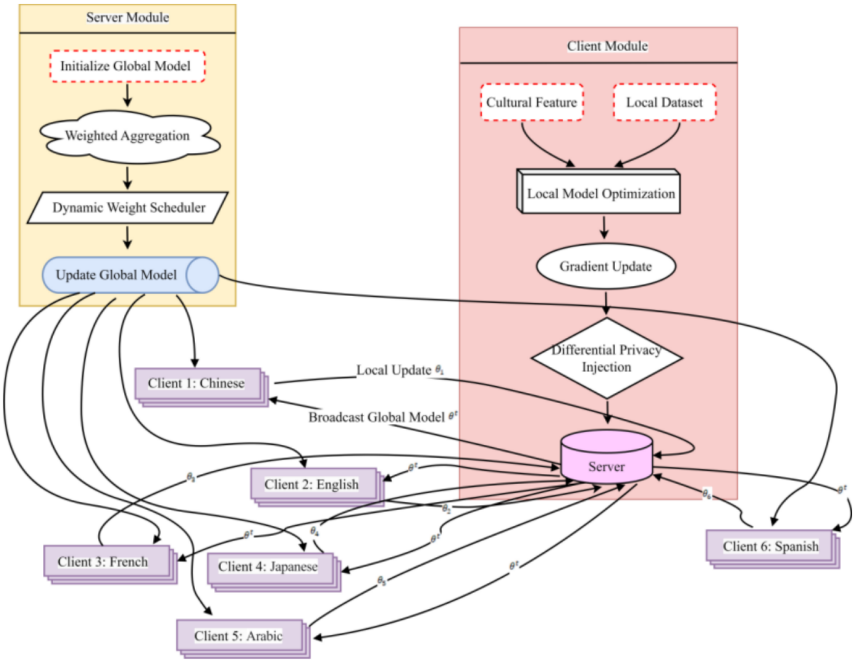

FIGURE 1. Architecture of culturally aware federated learning

In formula (3), D(·,·) is the Euclidean distance or cosine distance, c is the mean of all client culture vectors, Z is the normalization factor, and α and β are adjustable hyper-parameters used to control the influence of culture weight on the aggregation process. The server uses a weighted aggregation strategy to update the global model:

$$\theta^{t+1} = \sum_{i=1}^{N} w_i \theta_i^{t+1} \tag{4}$$

CAP-FL effectively integrates cultural dimension information without sharing the original data, making the model aggregation process culturally sensitive. The cultural feature vector constructed in this paper is quantified based on publicly available national/regional cultural dimension indices published by authoritative cross-cultural research institutions such as Hofstede Insights. Abstract cultural characteristics are transformed into six continuous numerical indicators. Each indicator typically ranges from 0 to 100, representing the relative strength of the culture in a specific dimension. We associate each language with representative cultural indices from its primary regions of use. Subsequently, to make it suitable for federated learning models, the original index values of all dimensions are standardized, resulting in a 6-dimensional, numerical cultural feature vector. This vector serves as metadata for the client and is uploaded to the server during each round of aggregation for calculating cultural weights. Figure 1 shows the overall process of the culture-aware federated learning architecture proposed in this paper. It adopts the classic Client-Server structure. Multiple language clients (Chinese, English, French, Japanese, Arabic, and Spanish) perform model optimization locally and upload gradient updates with differential privacy to the server. The server calculates the cultural weight based on the cultural feature vector of each client, uses a weighted aggregation strategy to update the global model, and optimizes the aggregation process through a dynamic weight scheduler; the updated model parameters are sent to each client to implement a new round of training. This architecture integrates cultural dimension modeling and personalized adaptation mechanism while ensuring that data does not leave the domain, effectively improving the semantic consistency and privacy security in multilingual cultural metaphor migration.

B. CONSTRUCTION OF METAPHOR SHARED REPRESENTATION SPACE

This paper proposes a method for constructing metaphor shared representation space based on the fusion of contrastive learning and graph neural network [23,24]. This method generates a unified semantic embedding space by modeling the common features and mapping relationships between cross-language metaphors, thereby improving the generalization ability and migration effect of the model between heterogeneous languages. Supposing the metaphor representation obtained after the input text $x \in X_{\text{multi}}$ is encoded by the local language model is $z = f(x;\theta_i) \in \mathbb{R}^d$, θ_i represents the model parameters of client i, and d is the embedding dimension. In order to enhance the model's ability to identify cross-language metaphor consistency, this

paper introduces MSL, which is in the following form:

$$L_{\mathrm{MSL}} = \frac{1}{|B|}\sum_{k=1}^{|B|}\left[\log\left(1+\sum_{j\in P_k} e^{-\lambda(\cos(z_k,z_j)-\epsilon)}\right) + \log\left(1+\sum_{m\in N_k} e^{\mu(\cos(z_k,z_m)-\epsilon)}\right)\right] \tag{5}$$

In formula (5), B is the current training batch, z_k is the anchor sample, P_k and N_k represent the positive and negative sample sets belonging to the same metaphor category and different metaphor categories as z_k, respectively, $\cos(\cdot,\cdot)$ is the cosine similarity, λ and μ are the control gradient scaling factors, and ϵ is the boundary threshold. On this basis, in order to further explore the correlation between metaphor structures between languages, this paper constructs a cross-language metaphor knowledge graph $G = (V, E)$, where the node set V represents metaphor entities and the edge set E represents the mapping relationship between corresponding metaphor entities in different languages. In order to achieve effective encoding of graph structure information, the GraphSAGE algorithm is used to learn the graph representation in an aggregated manner [25,26]. For each node $v \in V$, its neighborhood sampling set is recorded as $N(v)$, and its updated embedding representation is calculated as follows:

$$h_v^{(l+1)} = \sigma\left(W^{(l)} \cdot \mathrm{CONCAT}\left(h_v^{(l)}, \mathrm{MEAN}_{u\in N(v)} h_u^{(l)}\right)\right) \tag{6}$$

In formula (6), l is the number of layers of the graph neural network, $W^{(l)}$ is the learnable parameter matrix, CONCAT is the vector concatenation operation, MEAN represents the mean aggregation of neighbor node embeddings, and $\sigma(\cdot)$ is the nonlinear activation function. Finally, the metaphor embedding z output by the local model is mapped to the graph space, the mapping function is defined as $g(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d'}$, and the projection loss function is added:

$$L_{\mathrm{proj}} = \|z - g(h_v)\|_2^2 \tag{7}$$

This minimizes the Euclidean distance between the local language metaphor representation and the standard metaphor representation in the graph, ensuring that metaphorical expressions in different languages can converge to a unified shared space. This section introduces contrastive learning mechanism and graph neural network modeling to achieve effective extraction and unified modeling of common features of cross-language metaphors, providing a semantically consistent and clearly structured metaphor representation foundation for the subsequent federated learning aggregation process.

C. MULTI-LANGUAGE HETEROGENEOUS DATA ADAPTATION MODULE

This paper proposes a lightweight model adaptation mechanism based on dynamic adapter. This module is inserted between the layers of the Transformer backbone network and only updates the internal parameters of the adapter while keeping the weights of the backbone model frozen, achieving efficient and low-overhead language adaptation [27,28]. Supposing the intermediate representation of the input sequence $x \in X$ after passing through the Transformer encoder is $h_{\mathrm{in}} \in \mathbb{R}^{T\times d}$, T is the sequence length, and d is the hidden dimension. The traditional Transformer layer updates the representation through the self-attention mechanism and the feed-forward network (FFN) [29]:

$$h_{\mathrm{out}} = \mathrm{LayerNorm}(h_{\mathrm{in}} + \mathrm{SelfAtt}(h_{\mathrm{in}})) \tag{8}$$

$$h_{\mathrm{out}} = \mathrm{LayerNorm}(h_{\mathrm{out}} + \mathrm{FFN}(h_{\mathrm{out}})) \tag{9}$$

In order to introduce language adaptation capabilities, this paper inserts a two-layer fully connected adapter module between each multi-head attention layer and the FFN layer. Its structure is as follows:

$$a(h) = W_2 \cdot \sigma(W_1 \cdot h + b_1) + b_2 \tag{10}$$

In formula (10), W_1 and W_2 are the learnable parameter matrices for dimensionality reduction and dimensionality increase, b_1 and b_2 are bias terms, and $\sigma(\cdot)$ is a nonlinear activation function. The adapter output is added back to the original input and normalized:

$$h_{\mathrm{adapted}} = \mathrm{LayerNorm}(h + a(h)) \tag{11}$$

To further improve the adaptability of the adapter to different language feature distributions, this paper introduces language-specific scaling factor γ_l and translation factor β_l to form a dynamic language adaptation mechanism:

$$h_{\mathrm{lang}} = \gamma_l \odot h_{\mathrm{adapted}} + \beta_l \tag{12}$$

In formula (12), $\odot$ represents element-by-element multiplication, γ_l and β_l are obtained by the language identifier index $l \in \{1, 2, \ldots, L\}$, which is used to adjust the adapter output to match the feature distribution of the target language in the shared metaphor space. In order to reduce the communication cost, this paper adopts a parameter separation strategy. In the federated learning process, only the adapter parameter $\theta_{\mathrm{adapter}}$ is transmitted instead of the entire model parameter. Let the global model backbone parameter be fixed to θ_{base}, and the local client only updates the adapter parameter $\theta_{\mathrm{adapter}}^{(i)}$. The final aggregation process is defined as:

$$\theta_{\mathrm{adapter}}^{t+1} = \sum_{i=1}^{N} w_i \cdot \theta_{\mathrm{adapter}}^{(i),t+1} \tag{13}$$

w_i is the cultural perception weight of the i-th client. The dynamic adapter module proposed in this section achieves effective adaptation to multilingual heterogeneous data by fine-tuning local parameters without changing the backbone model, providing a unified structural and language-sensitive local model update path for subsequent federated aggregation.

D. CULTURAL SENSITIVITY MODELING AND WEIGHTED AGGREGATION

This paper proposes a dynamic weighted aggregation strategy based on cultural dimension information to make the model updating process culturally sensitive [30,31]. Assuming that the cultural feature vector of each client $i \in \mathcal{N}$ is $c_i \in \mathbb{R}^6$, and its dimensions have been extracted according to Hofstede's cultural dimension theory and standardized by mean normalization:

$$\hat{c}_i = \frac{c_i - \mu_c}{\sigma_c} \tag{14}$$

In formula (14), μ_c and σ_c are the global mean and standard deviation of all client culture vectors respectively. On this basis, the cultural influence weight function of the client is defined as follows:

$$w_i = \alpha \cdot \exp(-\beta \cdot \|\hat{c}_i - \bar{c}\|_W^2) \tag{15}$$

In formula (15), $\bar{c} = \frac{1}{N}\sum_{i=1}^{N} \hat{c}_i$ is the global cultural center point; $\|\cdot\|_W^2$ is the square of the Mahalanobis distance with the weight matrix $W \in \mathbb{R}^{6\times 6}$, which is used to emphasize the importance differences of different cultural dimensions; α and β are hyperparameters that control the smoothness of the weight distribution. The weighted aggregation strategy proposed in this study is calculated based on the complete six-dimensional Hofstede culture feature vector. In each round of federated training, after the server receives the model gradient update $\Delta\theta_i$ uploaded by each client, it uses the weighted average strategy to perform aggregation:

$$\Delta\theta_{\text{global}} = \sum_{i=1}^{N} \frac{w_i}{\sum_{j=1}^{N} w_j} \cdot \Delta\theta_i \tag{16}$$

This mechanism ensures that clients with stronger cultural influence occupy a higher proportion in the global model update. To further improve the adaptability of the aggregation strategy, this paper introduces an adaptive weight scheduler to dynamically adjust parameters α and β based on the historical aggregation error and the migration quality evaluation index fed back by the local adaptation module. Assuming the aggregate loss of round t is:

$$E_t = \frac{1}{|B|} \sum_{(x,y)\in B} l(x, y; \theta_t) \tag{17}$$

The scheduler updates the weight parameters according to the following rules:

$$\alpha^{(t+1)} = \alpha^{(t)} + \eta \cdot \nabla_\alpha E_t, \quad \beta^{(t+1)} = \beta^{(t)} + \eta \cdot \nabla_\beta E_t \tag{18}$$

In formula (18), η is the learning rate, $\nabla_\alpha E_t$ and $\nabla_\beta E_t$ are the estimated values of the gradient of the loss with respect to the weight parameters.

E. GRADIENT COMMUNICATION MECHANISM UNDER PRIVACY PROTECTION

This paper proposes a dual privacy protection mechanism that combines Differential Privacy (DP) and Secure Aggregation (SecAgg) [32,33]. This paper employs a differential privacy mechanism that adds Gaussian noise to the gradient of the local model on the client side, resulting in output perturbation. This mechanism injects controllable noise before the client uploads the local model gradient, and ensures that the server can only obtain the aggregated global updates through an encryption protocol, thus achieving a cross-language metaphor transfer training process with "data within the domain", as shown in Figure 2.

Assuming that the local model gradient output by the i-th client after the t-th round of training is updated to $\Delta\theta_i^{(t)} \in \mathbb{R}^d$. To meet the differential privacy requirements, adding a noise vector $\xi_i^{(t)} \sim \mathcal{N}(0, \sigma_i^2 I_d)$ that follows a Gaussian distribution, σ_i is the noise standard deviation, and I_d is a $d \times d$ unit matrix. The noise intensity is determined according to the following dynamic adjustment strategy:

$$\sigma_i = \gamma \cdot \left(\frac{|D_i|}{\sum_{j=1}^{N} |D_j|}\right)^{-\rho} \cdot (\|\hat{c}_i - \bar{c}\| + \epsilon) \tag{19}$$

In formula (19), $|D_i|$ represents the size of the local dataset of client i; $\hat{c}_i$ and $\bar{c}$ are the standardized cultural feature vector and the global cultural center point respectively; γ, ρ, and ϵ are the hyperparameters for controlling the noise scale; the noise decreases as the data scale increases, while the perturbation to clients with large cultural differences is enhanced to balance the conflict between privacy protection and model performance. The gradient update after perturbation is defined as:

$$\widetilde{\Delta\theta}_i^{(t)} = \Delta\theta_i^{(t)} + \xi_i^{(t)} \tag{20}$$

On this basis, a secure aggregation protocol is introduced to ensure that the server cannot directly access any single client's $\widetilde{\Delta\theta}_i^{(t)}$. Each client uses the shared encryption key sk to encrypt the local gradient, and performs an XOR operation with the random mask generated by other clients before uploading to form the final submitted encrypted gradient:

$$\Delta\theta_i^{\text{enc}} = \text{Enc}_{sk}\left(\widetilde{\Delta\theta}_i^{(t)} \oplus m_i^{(t)}\right) \tag{21}$$

$m_i^{(t)}$ is a random mask generated by negotiation among all clients, satisfying $\sum_{i=1}^{N} m_i^{(t)} = 0$ to ensure that the masks cancel each other out during the aggregation phase. After receiving all encrypted gradients, the server performs decryption and aggregation operations:

$$\begin{aligned} \Delta\theta_{\text{global}}^{(t)} &= \text{Dec}_{sk}\left(\sum_{i=1}^{N} \Delta\theta_i^{\text{enc}}\right) \\ &= \sum_{i=1}^{N}\left(\widetilde{\Delta\theta}_i^{(t)} \oplus m_i^{(t)}\right) = \sum_{i=1}^{N} \widetilde{\Delta\theta}_i^{(t)} \end{aligned} \tag{22}$$

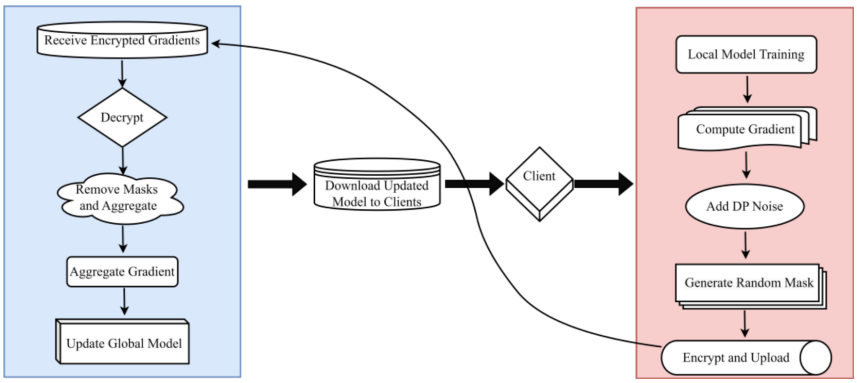

FIGURE 2. Gradient encryption and aggregation flow chart

Finally, the server uses the above aggregated gradients to update the global model parameters:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \cdot \Delta\theta^{(t)}_{\text{global}} \tag{23}$$

η is the learning rate. To avoid slow model convergence due to frequent noise injection, this paper designs a dynamic noise threshold mechanism to automatically adjust the noise injection intensity according to the historical training loss change trend:

$$\sigma_i^{(t+1)} = \begin{cases} \sigma_i^{(t)}, & E_t < \tau, \\ \sigma_i^{(t)} \cdot \kappa, & \text{otherwise} \end{cases} \tag{24}$$

E_t represents the training loss of the t-th round, τ is the set loss threshold, $\kappa \in (0, 1)$ is used to reduce the noise intensity to improve the convergence speed. The key hyperparameters in the CAP-FL architecture proposed in this paper are shown in Table 1. Figure 2 illustrates the privacy-preserving gradient communication mechanism, which uses differential privacy and secure aggregation for "data-in-domain" federated learning. After local training, the client injects differential privacy noise into the gradient, generates a random mask, and uploads the encrypted result. The server decrypts the aggregated gradients, removes the mask, and obtains the global model update. This update is then sent to clients for the next round. This process ensures efficient, secure cross-language metaphor transfer while guaranteeing user data privacy.

TABLE 1. Model hyperparameter settings

Parameter Name	Value / Range	Description
Learning Rate	0.001	Controls the step size of model updates
Cultural Weight Control Coefficients	$\alpha = 1.0, \beta = 0.5$	Adjusts the influence of cultural features during aggregation
Contrastive Learning Temperature	0.1	Regulates sensitivity to positive/negative sample distances
Graph Neural Network Layers	2	Number of GNN layers used for cross-lingual metaphor graph modeling
Adapter Bottleneck Dimension	64	Hidden layer size in dynamic adapter module
Differential Privacy Noise Scale	Gaussian noise, dynamically adjusted	Varies based on data size and cultural differences, with a base privacy budget of 1.0
DP Scaling Factors	$\gamma = 0.8, \rho = 0.5, \epsilon = 1.0$	Parameters controlling noise injection intensity
Secure Aggregation Mask Dimension	Same as model parameters	Ensures encrypted gradients remain aggregatable
Aggregation Frequency	Once per round	Controls communication efficiency
Maximum Training Rounds	40	Prevents overfitting and controls training duration

III. EXPERIMENT AND VERIFICATION

A. EXPERIMENTAL DESIGN

This paper constructs a multicultural metaphor corpus in six languages: Chinese (zh), English (en), French (fr), Japanese (ja), Arabic (ar), and Spanish (es). Each language is selected from news reports, literary works, social media comments, etc. to ensure the cultural diversity of the corpus and the authenticity of the language expression. All text samples are manually annotated, including metaphorical expression boundaries, ontology mapping relationships and their corresponding cultural background labels. The dataset is divided into training set, validation set and test set according to the ratio of 8:1:1, and a control group (FedAvg) and an experimental group (CAP-FL) are set up for comparative training under the same initialization conditions, as shown in Table 2.

B. EVALUATION OF CROSS-LINGUAL METAPHOR RECOGNITION PERFORMANCE

The cross-lingual generalization ability of the model is evaluated on the target language test set that is not involved in local training. Due to the uneven distribution of positive and negative samples in the metaphor recognition task, this paper uses precision, recall and their harmonic mean F1 Score (F1) as the core evaluation indicators. The F1 calculation formula is as follows:

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{25}$$

Precision measures the accuracy of the model in predicting metaphorical expressions, and Recall measures the coverage of actual metaphor samples. Figure 3 shows the metaphor recognition performance comparison between FedAvg and CAP-FL in six languages. The horizontal axis is the six languages (Chinese, English, French, Japanese, Arabic, and Spanish), and the vertical axis is the three indicators of F1 Score, precision, and recall. As can be seen from the figure, CAP-FL outperforms FedAvg in all languages and all indicators, with the largest F1 improvement reaching 9.1% (Japanese) and the smallest being 5.1% (English), with an average improvement of about 7.1%. In terms of precision, CAP-FL has the most significant im-

TABLE 2. Dataset statistics

Language	Total Samples	Training Set (80%)	Validation Set (10%)	Test Set (10%)	Hofstede Cultural Index
Chinese	10,000	8,000	1,000	1,000	65.2
English	12,000	9,600	1,200	1,200	73.5
French	9,000	7,200	900	900	68.4
Japanese	8,500	6,800	850	850	61.7
Arabic	7,000	5,600	700	700	54.9
Spanish	9,500	7,600	950	950	70.1

provement in Japanese (8.6%), and the largest improvement in recall (9.5%) in Japanese. This shows that the culture-aware federated learning architecture proposed in this paper has significant advantages in the cross-language metaphor transfer task. The improved performance stems from CAP-FL's effective modeling of multilingual cultural differences. Unlike FedAvg's equal-weight strategy, which suffers from cultural cognitive bias and semantic shift, CAP-FL uses Hofstede's cultural dimension theory and a weighted aggregation mechanism to enhance adaptation across cultural backgrounds. By combining dynamic adapters and contrastive learning, it constructs a unified metaphor representation space, significantly improving cross-language consistency. Crucially, the added complexity of differential privacy and secure aggregation did not hurt performance; instead, it mitigated data heterogeneity, confirming the method's success in balancing privacy with model efficacy.

C. ANALYSIS OF CULTURAL CONSISTENCY LOSS

To measure the semantic consistency of model outputs across different languages, the last layer of metaphor embedding vectors of each language model are extracted, and the cosine similarity matrix is calculated:

$$S_{ij} = \frac{z_i^\top z_j}{\|z_i\|\|z_j\|} \tag{26}$$

Then defining the cultural consistency loss function:

$$L_{\text{consist}} = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\log\left(1+e^{-\gamma(S_{ij}-\tau)}\right) \tag{27}$$

γ is the scaling factor, and τ is the target similarity threshold. This indicator is used to evaluate whether the model effectively aligns cultural metaphor expressions in different languages. Figure 4 shows a comparison of the similarity matrices of the two federated learning methods, FedAvg and CAP-FL, on the consistency of cross-language metaphor embeddings. The horizontal and vertical axes are the six target languages and source language labels (Chinese, English, French, Japanese, Arabic, and Spanish), and the color depth reflects the degree of semantic similarity of metaphorical expressions between different languages. As can be seen from the figure, the similarity of CAP-FL in most language pairs is significantly higher than that of FedAvg, with an overall average of 0.82. In the "Chinese-English" combination, FedAvg is 0.72, while CAP-FL increases to 0.85; in the "Chinese-Spanish" combination, FedAvg is 0.7, and CAP-FL reaches 0.92, showing stronger cross-language alignment capabilities. The overall similarity between Arabic and other languages is relatively low, indicating that its cultural uniqueness poses a greater challenge to the model. CAP-FL shows a higher concentration of similarities near the diagonal between languages, indicating that it has more advantages in modeling a unified semantic space.

D. METAPHOR TRANSFER COMPREHENSIBILITY EVALUATION

This paper invites native speakers of the target language to manually score the metaphorical expressions generated by the transfer. The scoring dimensions include naturalness, cultural fit and semantic clarity, with a full score of 5. This evaluation is used to verify whether the model meets the cultural adaptation requirements while maintaining semantic consistency. Figure 5 shows the comparison between FedAvg and CAP-FL in terms of cross-language metaphor transfer comprehensibility. The horizontal axis is six languages (Chinese, English, French, Japanese, Arabic, and Spanish), and the vertical axis contains the average scores of the three dimensions of naturalness, cultural fit, and semantic clarity. As can be seen from the figure, CAP-FL is significantly better than FedAvg in all languages and all dimensions. In terms of naturalness of Chinese, CAP-FL reaches 4.12, significantly higher than FedAvg's 3.68; in terms of cultural fit of Arabic, CAP-FL scores 3.68, while FedAvg is only 3.13; in terms of semantic clarity of English, CAP-FL reaches 4.31, significantly higher than FedAvg's 3.89. Overall, CAP-FL achieves an average score of 4.02 in six languages, 0.46 points higher than FedAvg's 3.56, indicating that it has significant advantages in improving the naturalness, cultural adaptability and semantic clarity of model-generated expressions.

E. MODEL COMMUNICATION EFFICIENCY EVALUATION

This paper records the client upload parameter amount and communication time in each round of federated training. The communication efficiency index is defined as:

$$\eta_t = \frac{\sum_{i=1}^{N}\|\Delta\theta_i^{(t)}\|_F}{T_t} \tag{28}$$

$\|\cdot\|_F$ represents the Frobenius norm, and T_t represents the total time consumed in the t-th round of communication. This indicator is used to analyze the impact of dynamic adapters and privacy mechanisms on communication overhead. Figure 6 shows the communication performance comparison between FedAvg and CAP-FL in the cross-language cultural metaphor transfer task. The horizontal axis is the

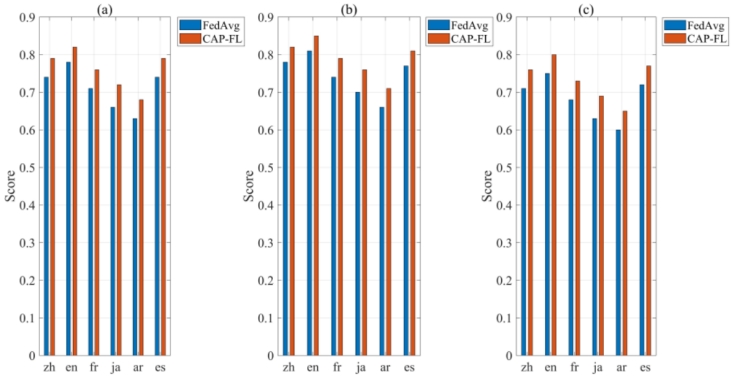

FIGURE 3. Cross-language metaphor recognition performance evaluation; Figure 3(a)F1 value; Figure 3(b)Precision; Figure 3(c):Recall

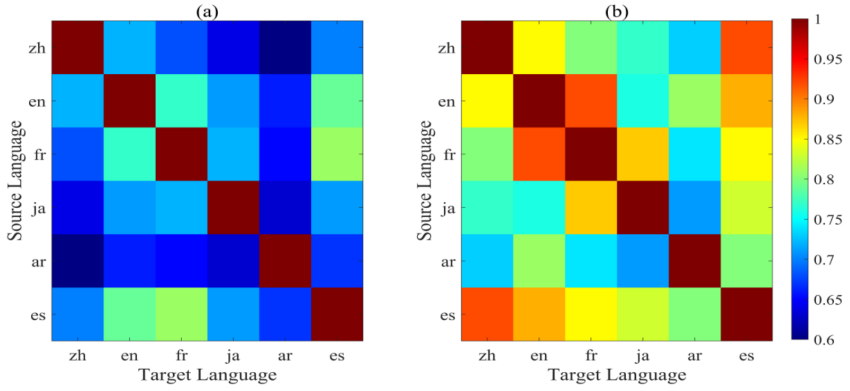

FIGURE 4. Metaphor vector similarity matrix; Figure 4(a) FedAvg group; Figure 4(b)CAP-FL group

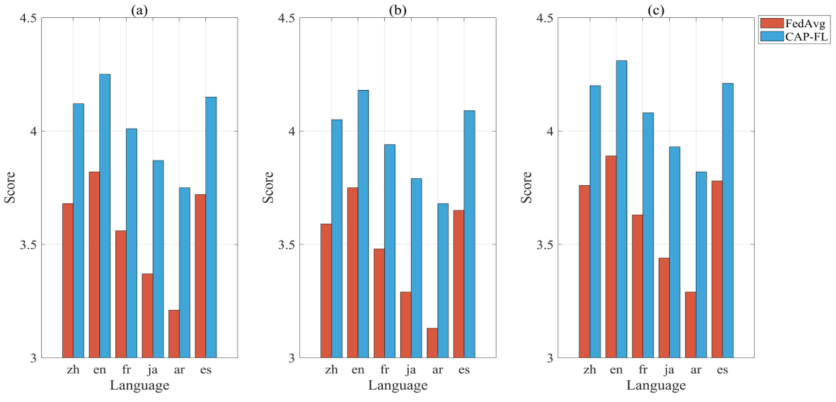

FIGURE 5. Metaphor transfer comprehensibility evaluation; Figure 5(a) Naturalness; Figure 5(b) Cultural fit; Figure 5(c) Semantic clarity

communication round, and the vertical axis is the uploaded parameter amount, communication time and communication efficiency. As can be seen from the figure, the average parameter upload volume per round of FedAvg is 20.4MB, while that of CAP-FL is only 5.1MB, a decrease of more than 75%; in terms of communication time, the average of FedAvg is 10.6 seconds, while that of CAP-FL is controlled within 3.1 seconds on average; in terms of communication efficiency, the communication efficiency of FedAvg is slightly higher than that of CAP-FL, and CAP-FL generally remains above 1.4 MB/s. Overall, CAP-FL significantly reduces communication overhead while ensuring privacy, demonstrating stronger scalability and practicality.

F. PRIVACY LEAKAGE RISK CONTROL ASSESSMENT

This paper simulates the process of attackers inferring sensitive client information through model updates, and defines

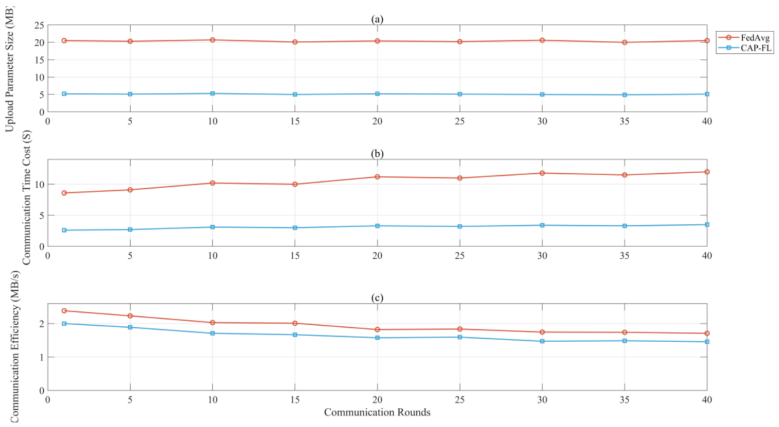

FIGURE 6. Communication efficiency evaluation; Figure 6(a)Upload parameter quantity; Figure 6(b)Communication time consumption; Figure 6(c)Communication efficiency

the success rate of the inferring attack as:

$$P_{\text{attack}} = \frac{1}{|D|} \sum_{x \in D} 1(\hat{x} = x) \tag{29}$$

$\hat{x}$ is the data sample reconstructed by the attack model, $\mathbf{1}(\cdot)$ is the indicator function, which verifies the protection effect of the differential privacy mechanism under different disturbance levels. Figure 7 shows the trend of the reconstruction success rate of FedAvg and CAP-FL methods when facing model back-stepping attacks as the noise intensity changes. The horizontal axis is the standard deviation of the noise injected by differential privacy, that is, the intensity of the disturbance injected into the model update, and the vertical axis is the success rate of the back-test attack. From the figure, it can see that under noise-free or low-noise conditions (σ=0.1), the inversion attack success rate of FedAvg is as high as 87%, while CAP-FL is slightly lower at 84%; as the noise intensity increases, the inversion attack success rates of both decrease, but under the same privacy budget, CAP-FL is always better than FedAvg, and CAP-FL drops to a minimum of 0.5%. When σ=0.3, the success rate of FedAvg's inverse attack drops to 68%, while CAP-FL has dropped to 60%; when σ=0.5, the success rate of CAP-FL's inverse attack is only 42%, while FedAvg remains at 51%. The overall trend shows that CAP-FL has stronger data leakage prevention capabilities under the same privacy protection mechanism.

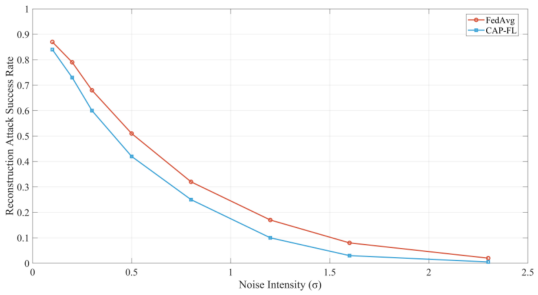

FIGURE 7. Privacy leakage risk control assessment

G. USER CULTURAL PREFERENCE MATCHING EVALUATION

This paper collects user feedback data on migration results and constructs a user-metaphor interaction matrix $R \in \mathbb{R}^{M \times K}$, M is the number of users, and K is the number of candidate metaphors. The matching degree between the recommended metaphor and the user's cultural background is calculated based on the collaborative filtering algorithm:

$$s_{mk} = \frac{\mathbf{u}_m \mathbf{v}_k}{\|\mathbf{u}_m\| \|\mathbf{v}_k\|} \tag{30}$$

$\mathbf{u}_m$ and $\mathbf{v}_k$ are the embedding vectors of user and metaphor respectively. Table 3 shows the cultural preference matching evaluation results of the CAP-FL method proposed in this paper on users of different languages, including the average matching degree and user satisfaction score. As can be seen from the table, English users have the highest average match (0.87) and a satisfaction score of 4.32; Arabic users have the lowest match (0.68) and a satisfaction score of 3.65, reflecting the adaptation challenges faced by the model in languages with large cultural differences. Overall, the average matching degree of users in each language reaches 0.79, and the average satisfaction score is higher than 4 points, indicating that the model can better capture the relationship between users' cultural background and metaphorical expressions, and has strong personalized recommendation capabilities.

TABLE 3. Statistics of user cultural preference matching evaluation results

Language	Average Matching Score	User Satisfaction Score (1–5)
Chinese (zh)	0.82 ± 0.13	4.15
English (en)	0.87 ± 0.10	4.32
French (fr)	0.80 ± 0.15	4.08
Japanese (ja)	0.75 ± 0.17	3.92
Arabic (ar)	0.68 ± 0.21	3.65
Spanish (es)	0.83 ± 0.12	4.21

H. ABLATION STUDY ON THE CULTURAL WEIGHTING MECHANISM

To systematically evaluate the contribution of the culture-aware aggregation mechanism to overall performance, this paper designs an ablation experiment. Under the same training and testing conditions, starting from the complete model, the culture-aware aggregation mechanism module is removed while the remaining configurations are fixed. Both models are trained and tested under identical conditions. To verify the core value of the cultural perception aggregation mechanism, this paper designed an ablation experiment. Table 4 clearly reveals the contribution of this module to the overall performance. When this core module was removed, the model performance showed a significant and comprehensive decline: the average F1 score for cross-linguistic metaphor recognition dropped sharply from 76.0% to 69.2%, a decrease of 6.8%. This directly proves the decisive role of cultural weights in improving the model's generalization ability and accuracy. At the same time, in terms of key indicators measuring the quality of cross-cultural semantic alignment, cultural consistency dropped from 0.82 to 0.77, and the user's rating of transfer comprehensibility also decreased from 4.02 to 3.87. These data collectively indicate that the cultural feature vectors and weighted aggregation strategy based on Hofstede theory are not optional additions, but core components for effectively coordinating semantic shifts between different cultural clients and achieving high-quality cross-linguistic metaphor transfer. Their absence will seriously weaken the model's performance in multilingual cultural scenarios.

TABLE 4. Ablation Study Results Comparison

Model	Cultural perception aggregation mechanism	Avg. F1 Score (%)	Cultural Consistency	Transfer Comprehensibility
Full Model	√	76	0.82	4.02
	—	69.2	0.77	3.87

IV. CONCLUSION

This paper proposes a novel Culture-Aware Personalized Federated Learning (CAP-FL) architecture specifically to address the challenges of semantic consistency and privacy protection in multilingual cultural metaphor transfer. This is highly relevant to the industry, where consistent meaning for terms related to style, quality, or sustainability must be maintained across diverse international design, manufacturing, and retail units while ensuring the confidentiality of local market data. The method incorporates Hofstede's cultural dimension theory by designing a cultural weight function to introduce cultural sensitivity into federated aggregation. It uses contrastive learning and Graph Neural Networks (GNN) to create a cross-language metaphor shared representation space, enhancing semantic alignment via Multi-Similarity Loss and GraphSAGE. A dynamic adapter module achieves efficient, lightweight language adaptation, while differential privacy and secure aggregation ensure data privacy. Experimental results show CAP-FL significantly outperforms FedAvg in cross-language metaphor recognition, with an average F1 improvement of 7.1%. It also effectively boosts cultural consistency (average 0.82), transfer comprehensibility (average 4.02), and user cultural preference matching (average 0.79). Future work will focus on expanding the model to include a wider range of languages and concentrate on finer-grained cultural cognitive modeling. For the industry, this will further improve adaptability in complex cross-cultural scenarios, ensuring highly localized AI systems that can accurately capture nuanced fashion trends, material preferences, and regulatory language across the world's most diverse manufacturing and consumer markets. The core innovation of the culture-aware personalized federated learning architecture proposed in this paper lies in its integration of macro-level, interpretable cultural theories into the model aggregation decision-making process. Techniques such as knowledge distillation and local fine-tuning typically aim at data-level adaptation, i.e., fitting the unique distribution of local data by adjusting model parameters. In contrast, the cultural weights introduced by CAP-FL serve as a priori, sociologically based guiding signals. They do not rely on implicit features generated after model training, but rather utilize Hofstede's cultural dimension—external knowledge independent of the training data—to proactively and purposefully coordinate the learning directions among different cultural clients during the aggregation phase. This design enables the model to fundamentally understand and compensate for semantic shifts caused by cultural differences, rather than merely passively adapting to their outcomes.

AUTHOR CONTRIBUTIONS

Lei Shi designed, collected and analyzed the data, and drafted the manuscript. Lei Shi conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Lei Shi participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, 2023 Research Project under the Shandong Province Education Science "14th Five-Year Plan": "A Study on the Talent Development Model for Innovation and Entrepreneurship in Digital Language Services Based on Industry-Education Integration" (2023YB051), Shandong University of Finance and Economics 2024 University-Level Teaching Reform Research Project: "Research on the Development and Application

of a Digital Language Service Resource Repository for Economic and Trade Cooperation under the Belt and Road Initiative" (jy202423) and 2023 Jinan Philosophy and Social Science Project "Promotion Strategies of Quancheng Stories in New Media of the Core Area of Silk-road Economic Belt" (JNSK23C43).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. H. Ahtif and N. Gandhi, "The role of language in cross cultural bonds," Journal of Asian multicultural research for social sciences study, vol. 3, no. 4, pp. 7-16, 2022, doi: 10.47616/jamrsss.v3i4.321.

[2] S. Cheng, "The Status and Prospect Analysis of Language Service Industry in the Era of Artificial Intelligence," The Frontiers of Society, Science and Technology, vol. 6, no. 10, 2024, doi: 10.25236/FSST.2024.061014.

[3] P. O'Shaughnessy and Y. X. Lin, "Privacy Protection Practice for Data Mining with Multiple Data Sources: An Example with Data Clustering," Mathematics, vol. 10, no. 24, pp. 4744, 2022, doi: 10.3390/math10244744.

[4] Y. Weng and S. Dong, "A Privacy-Preserving Multilingual Comparable Corpus Construction Method in Internet of Things," Mathematics, vol. 12, no. 4, pp. 598, 2024, doi: 10.3390/math12040598.

[5] F. Deng, C. Fu, Y. Qian, J. Yang, S. He, and H. Xu, "Federated learning based multi-task feature fusion framework for code expressive semantic extraction," Software: Practice and Experience, vol. 52, no. 8, pp. 1849-1866, 2022, doi: 10.1002/spe.3094.

[6] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, "A survey on federated learning: challenges and applications," International Journal of Machine Learning and Cybernetics, vol. 14, no. 2, pp. 513-535, 2023, doi: 10.1007/s13042-022-01647-y.

[7] S. Banabilah, M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh, "Federated learning review: Fundamentals, enabling technologies, and future applications," Information processing & management, vol. 59, no. 6, Art. no. 103061, 2022, doi: 10.1016/j.ipm.2022.103061.

[8] C. Quan, Z. Chen, K. Ren, and Luo Z.FedOcw: optimized federated learning for cross-lingual speech-based Parkinson's disease detection, "npj Digital Medicine," 2025; 8(1):1-16, doi: 10.1038/s41746-025-01763-3.

[9] P. Phani Krishna, S. Arulmozi, S. R. Male, and Mishra RK.Are Older Bilinguals' Better in Metaphor Generation?.Journal of Psycholinguistic Research, "2023; 52(4):1183-1204,", doi: 10.1007/s10936-022-09929-w.

[10] R. Qiao and X. Huang, "Application of Deep Learning in Cross-Lingual Sentiment Analysis for Natural Language Processing," Journal of Artificial Intelligence Practice, vol. 7, no. 1, pp. 1-6, 2024, doi: 10.23977/jaip.2024.070101.

[11] W. Hong and C. Rossi, "The cognitive turn in metaphor translation studies: A critical overview," Journal of Translation Studies, vol. 5, no. 2, pp. 83-115.https://hal.science/hal-03342406, 2021.

[12] A. Sheng and Y. Kong, "Neural machine translation and human translation: A political and ideological perspective," Babel, vol. 69, no. 4, pp. 483-498, 2023, doi: 10.1075/babel.00332.she.

[13] T. Shima, N. Tsuda, K. Hashiguchi, and Muto T.Effect of Adjusting Cultural Backgrounds on the Impact of Metaphors: A Preliminary Study, "International journal of psychology and psychological therapy," 2022; 22(1):45-63.

[14] Z. Song, S. Tian, L. Yu, ZhangX, and J. Liu, "Multi-task metaphor detection based on linguistic theory," Multimedia Tools and Applications, vol. 83, no. 24, pp. 64065-64078, 2024, doi: 10.1007/s11042-023-18063-1.

[15] R. Šajina, N. Tanković, and I. Ipšić, "Multi-task peer-to-peer learning using an encoder-only transformer model," Future generation computer systems, vol. 152, pp. 170-178, 2024, doi: 10.1016/j.future.2023.11.006.

[16] J. Ren, H. Peng, L. Jiang, HaoZ, J. Wu, S. Gao, et al., "Toward Cross-Lingual Social Event Detection with Hybrid Knowledge Distillation," ACM Transactions on Knowledge Discovery from Data, vol. 18, no. 9, pp. 1-36, 2024, doi: 10.1145/3689948.

[17] E. Malan, V. Peluso, A. Calimera, and Macii E.Communication-efficient federated learning with gradual layer freezing, "IEEE Embedded Systems Letters," 2022; 15(1):25-28, doi: 10.1109/LES.2022.3190682.

[18] X. Wu, Y. Zhang, M. Shi, et al., "An adaptive federated learning scheme with differential privacy preserving," Future Generation Computer Systems, vol. 127, pp. 362-372, 2022, doi: 10.1016/j.future.2021.09.015.

[19] A. Sharma and N. Marchang, "A review on client-server attacks and defenses in federated learning," Computers & Security, Art. no. 103801, 2024, doi: 10.1016/j.cose.2024.103801.

[20] J. Liu, J. Huang, Y. Zhou, X. Li, S. Ji, H. Xiong, et al., "From distributed machine learning to federated learning: A survey," Knowledge and Information Systems, vol. 64, no. 4, pp. 885-917, 2022, doi: 10.1007/s10115-022-01664-x.

[21] J. Jan, K. A. Alshare, and P. L. Lane, "Hofstede's cultural dimensions in technology acceptance models: a meta-analysis," Universal Access in the Information Society, vol. 23, no. 2, pp. 717-741, 2024, doi: 10.1007/s10209-022-00930-7.

[22] Y. Gao, "The application of Hofstede's cultural dimension theory to middle school English teaching," International Journal of Frontiers in Sociology, vol. 3, no. 5, pp. 68-71, 2021, doi: 10.25236/IJFS.2021.030518.

[23] H. Yang, "Multilingual Information Retrieval Using Graph Neural Networks: Practical Applications in English Translation," J. Electrical Systems, vol. 20, no. 6s, pp. 1729-1739, 2024.

[24] D. Wang, X. Feng, Z. Liu, and WangC.2M-NER: contrastive learning for multilingual and multimodal NER with language and modal fusion, "Applied Intelligence," 2024; 54(8):6252-6268.

[25] M. Ma, C. Zhang, Y. Li, J. Chen, and X. Wang, "Rumor detection model with weighted GraphSAGE focusing on node location," Scientific Reports, vol. 14, no. 1, Art. no. 27127, 2024, doi: 10.1038/s41598-024-76738-7.

[26] T. Zhang, C. Mai, Y. Chang, C. Chen, L. Shu, and Zheng Z.FedEgo: privacy-preserving personalized federated graph learning with ego-graphs, "ACM Transactions on Knowledge Discovery from Data," 2023; 18(2):1-27, doi: 10.1145/3624017.

[27] B. B. Al-onazi, M. A. Nauman, R. Jahangir, MalikMM, AlkhammashEH, and Elshewey AM.Transformer-based multilingual speech emotion recognition using data augmentation and feature fusion, "Applied Sciences," 2022; 12(18):9188, doi: 10.3390/app12189188.

[28] M. Z. Hossain and S. Goyal, "Advancements in Natural Language Processing: Leveraging Transformer Models for Multilingual Text Generation," Pacific Journal of Advanced Engineering Innovations, vol. 1, no. 1, pp. 4-12, 2024, doi: 10.70818/pjaei.2024.v01i01.02.

[29] H. I. Liu and W. L. Chen, "X-transformer: a machine translation model enhanced by the self-attention mechanism," Applied Sciences, vol. 12, no. 9, pp. 4502, 2022, doi: 10.3390/app12094502.

[30] E. Agyei, X. Zhang, A. B. Quaye, V. A. Odeh, and Arhin JR.Dynamic Aggregation and Augmentation for Low-Resource Machine Translation Using Federated Fine-Tuning of Pretrained Transformer Models, "Applied Sciences," 2025; 15(8):4494, doi: 10.3390/app15084494.

[31] N. Lin, M. Zeng, X. Liao, W. Liu, A. Yang, and D. Zhou, "Addressing class-imbalance challenges in cross-lingual aspect-based sentiment analysis: Dynamic weighted loss and anti-decoupling," Expert Systems with Applications, vol. 257, Art. no. 125059, 2024, doi: 10.1016/j.eswa.2024.125059.

[32] X. Xiao, Y. Zhang, H. Chen, W. Ren, J. Zhang, and J. Xu, "A differential privacy-based mechanism for preventing data leakage in large language model training," Academic Journal of Sociology and Management, vol. 3, no. 2, pp. 33-42, 2025, doi: 10.70393/616a736d.323732.

[33] T. M. Mengistu, T. Kim, and J. W. Lin, "A survey on heterogeneity taxonomy, security and privacy preservation in the integration of IoT, wireless sensor networks and federated learning," Sensors, vol. 24, no. 3, pp. 968, 2024, doi: 10.3390/s24030968.