ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Generating Adversarial Networks to Simulate Authentic Cross-Cultural English Conversation Scenes for Foreign Trade English Teaching Applications

Xiaofeng Wang[1,*]

[1] Xi'an University of Architecture and Technology Huaqing College, Xi'an, Shaanxi, 710043, China

SUMMARY: *With the rise of generative adversarial network technology, generative adversarial network brings new research direction for cross-cultural English conversation scenarios, which helps the development of foreign trade English teaching in colleges and universities. In this paper, we introduce the "encoding-decoding" structure as a cross-cultural English response generator on the basis of the generator, and the encoder and decoder G are both composed of GRU units. The word vector approximation layer is generated by multiplying the word probability distribution output from the cross-cultural English response generator by the corresponding word vector. A discriminator based on convolutional neural network is used to prompt the cross-cultural English reply generator to produce results closer to the real data, and finally the work of constructing the Generative Adversarial Network-based Conversation Reply Generation Model (GAN-AEL) is completed and the loss function for model training is set. With the support of the corresponding development tools, the Generative Adversarial Network-based Conversation Response Generator model is successfully integrated into cross-cultural English conversation scenarios, and the supporting teaching system is finally designed and analyzed. Before using the FTES, the students' English writing scores were in the range of 6~15, and after using the FTES for one semester, the scores increased to 21~29, with a difference of 7~21, and the rest of the English reading scores, English speaking scores, and English listening scores were the same, which comprehensively confirmed the practical application effectiveness of FTES based on the cross-cultural English conversation scenario, aiming at the design and analysis of the supporting teaching system. It comprehensively confirms the effectiveness of the foreign trade English teaching system based on cross-cultural English dialog scenes, and aims to boost the development of foreign trade English teaching in colleges and universities.*

KEYWORDS: *generative adversarial network; GAN-AEL; foreign trade English teaching system; English conversation scene*

1 Introduction

Accompanied by the extensive use of Internet information technology, cross-border e-commerce, international business in the total value of exports of foreign trade in the transaction volume continues to rise, the talent side has generated a great demand for the cultivation of composite foreign language to foreign trade personnel put forward a new test and requirements [1-3]. Due to the diversified background of customers in foreign trade cooperation, cultural

*submrine24@163.com
https://doi.org/10.65102/is2026247 Published: 30 April, 2026

differences can easily lead to cross-cultural communication barriers. And cross-cultural communication ability is one of the necessary abilities of an excellent foreign language communication international talents, the creation of cross-cultural language environment is conducive to help students in colleges and universities to tune up the self-knowledge ability as well as enhance cultural self-confidence [4, 5]. Enhance students' cross-cultural communication ability, and make students improve their cultural literacy at the same time to obtain a sense of self-identity and cultural identity. Currently, the teaching of foreign trade English is often carried out by traditional teacher explanation and basic role-playing, which lacks the simulation of cross-cultural communication scenarios, making it difficult for students to understand the real foreign trade conversation scenarios, resulting in insufficient cross-cultural communication skills [6-8].

Generative Adversarial Network (GAN) is an important breakthrough in the field of machine learning, GAN is through two mutually antagonistic neural networks - generator and discriminator - constantly playing, and finally achieve the effect of generating highly realistic data. In English language teaching, GAN has a large potential for application. Literature [9] utilizes GAN to construct an English cultural communication material generation model, and its generated material shows high image realism and high user clicking and participation, which verifies the effective potential of GAN in creating high realism and diverse English cultural content. Literature [10] combines GAN with reinforcement learning to construct a generative model that can handle discrete texts, which has the potential to improve the quality of English translation teaching. Literature [11] integrates gated recurrent units, long and short-term memory networks, and GAN to propose an automatic expert comment generation method for English speaking assessment, which can effectively generate speaking comments with both depth and humanization features. Literature [12] formulated a GAN-based content optimization method for academic English, using Transformer as a generator, combining data enhancement and strategy gradient training to effectively correct grammatical errors. Literature [13] used recursive loop GAN with gazelle optimization for English language teaching and learning assessment through learnable edge collaborative filtering preprocessing with recursive loop GAN prediction. Literature [14] explored the application of GAN fused with virtual reality and augmented reality technologies in constructing virtual simulation educational scenarios to promote the comprehensive development of morality, intelligence, physical fitness, and aesthetics, demonstrating the potential in revolutionizing the future education model. Literature [15] created a personalized English learning path optimization algorithm with the help of GAN, which can effectively improve learning performance, efficiency and satisfaction by about 20% and 15% through the adversarial training of generators and discriminators and dynamic adaptation of learner characteristics. It can be seen that GAN technology is able to generate English text, evaluate teaching, construct virtual simulation educational scenarios, and realize personalized education.

In this paper, we add the “encoding-decoding” structure to the generator in the generative adversarial network to form the cross-cultural English response generator. In order to deal with the non-conductivity problem caused by the discrete output of the cross-cultural English response generator, we propose to multiply the word probability distribution of the generator's output by the corresponding word vector to produce the approximation of the generated result, which is then set as the word vector approximation layer. In order to make the cross-cultural English reply generator produce results that better fit the actual situation, a discriminator based on convolutional neural network is used, and finally a dialogue reply generation model based on generative adversarial network (GAN-AEL) is constructed, which contains three main modules based on the likelihood probability maximization generator pre-training, word vector estimation layer, and a discriminator based on convolutional neural network. In the development of software support, to complete the construction of cross-cultural English dialogue scene, in order to make it better applied to the teaching of foreign trade English in colleges and universities, a foreign trade English teaching system based on cross-cultural English dialogue scene is designed, and the system is explored and analyzed from multiple dimensions.

2 The Application of Cross-cultural English Dialogue Scenes in Teaching English for Foreign Trade

2.1 Adversarial network-based model for dialog response generation

In order to realize the generative adversarial network to simulate real cross-cultural English conversation scenarios, this subsection takes word vectors as an entry point to illustrate the semantic consistency between the discrete sampling process of the text and the word vector estimation process, and then proposes a generative adversarial network variant for the task of generating replies to cross-cultural English conversations. The core of the approach is the introduction of a word vector estimation layer to replace the discrete sampling-based decoding process, so that the whole model is continuous, microscopic, and Adversarial Network-based Dialogue Response Generation Model (GAN-AEL) which consists of three main modules based on the likelihood probability maximization generator pre-training, word vector estimation layer, and the convolutional neural network-based discriminator.

2.1.1 Generator based on maximizing likelihood probability

The "encoding-decoding" structure was introduced on the basis of the generator as a cross-cultural English response generator. The "encoding-decoding" model structure is shown in Figure 1. The encoders and decoders G are both composed of GRU units, and their main function is to generate the corresponding response $r=\left\{w_{r,1}, w_{r,2}, \cdots, w_{r,K}\right\}$ based on the input message $q=\left\{w_{q,1}, w_{q,2}, \cdots, w_{q,J}\right\}$. For the "encoding-decoding" structure, its training objective is to maximize the conditional probability $p\left(r|q\right)$ during the generation process for the given message-response pair (q,r). Specifically, the encoder first encodes the sequence of words q into a vector q_v. Then, the generator G estimates the probability of each word appearing in the response r successively according to q_v. This process is as follows:

$$p\left(r|q\right)=\prod_{t=1}^{K} p\left(w_{r,t} | q_v, w_{r,1}, \cdots, w_{r,t-1}\right) \tag{1}$$

In general, the logarithmic form of Eq. (1) above is often used as the objective function to facilitate derivation and efficient computation. Thus the objective function of the generator follows the following equation (2):

$$\frac{1}{|\mathcal{D}|} \sum_{(q,r)\in\mathcal{D}} \sum_{t=1}^{K} \log p\left(w_{r,t} | q_v, w_{r,1}, \cdots, w_{r,t-1}\right) \tag{2}$$

Note that Eq. (2) is used as a loss function, thus enabling pre-training of the generator.

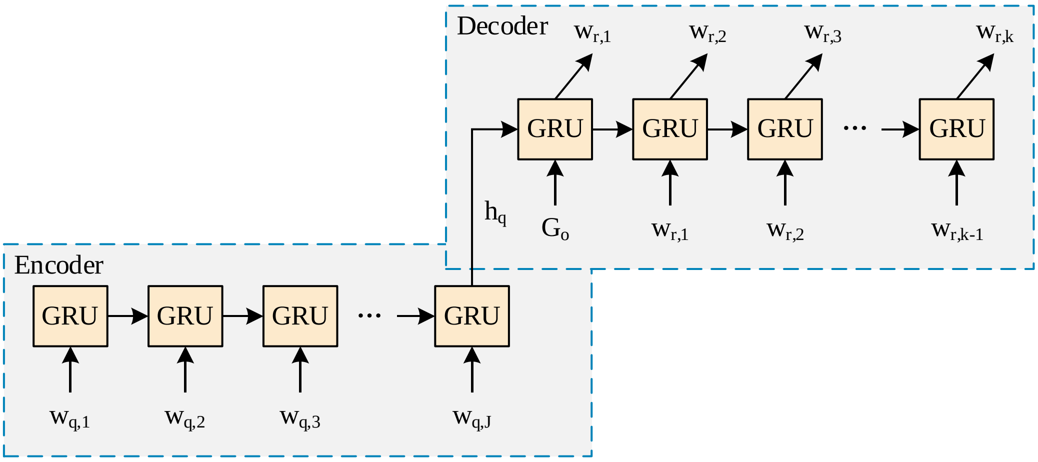


Figure 1: Encoder-decoding model structure

2.1.2 Word vector estimation layer

The structure of the word vector approximation layer is shown in Fig. 2. The word vector approximation takes the output of the generator h_i and the random noise z_i as inputs, and the main reason for adding noise to this approximation layer is that h_i can represent the implied semantics of the words at the current time step, and the overall semantics is very close to the original semantics after adding a small amount of noise, but this kind of noise brings the model a greater chance to choose more words and increases the diversity of the generated responses. diversity of the generated responses. Then, the projection layer from the implied vectors to the word distributions maps $h_i + z_i$ to the corresponding word distributions w_i , and the word vectors of the words corresponding to h_i are estimated by the product of the word distributions to the word vector matrices. The word vector approximation is calculated as in Equation (3):

$$e_{\hat{w}_i} = \sum_{j=1}^{V} e_j \cdot Soft\max\left(W_p\left(h_i + z_i\right) + b_p\right)_j \tag{3}$$

where W_p and b_p are the weights and bias of the implied vector to the projection layer of the word distribution, respectively. According to Eq. (3) an approximate representation of the word vectors can be obtained, which is directly input to the generator to get the word vector estimates for the next time step. h_i is then the implicit vector representation of w_i input by the generator. Namely:

$$h_i = g\left(h_{i-1}, e_{\hat{w}_{i-1}}\right) \tag{4}$$

where $g(\cdot)$ classical forward process of the gating unit. The above process is repeated until a preset sequence length is output. The entire sequence of $e_{\hat{w}_i}$ is then used as input to the discriminator.

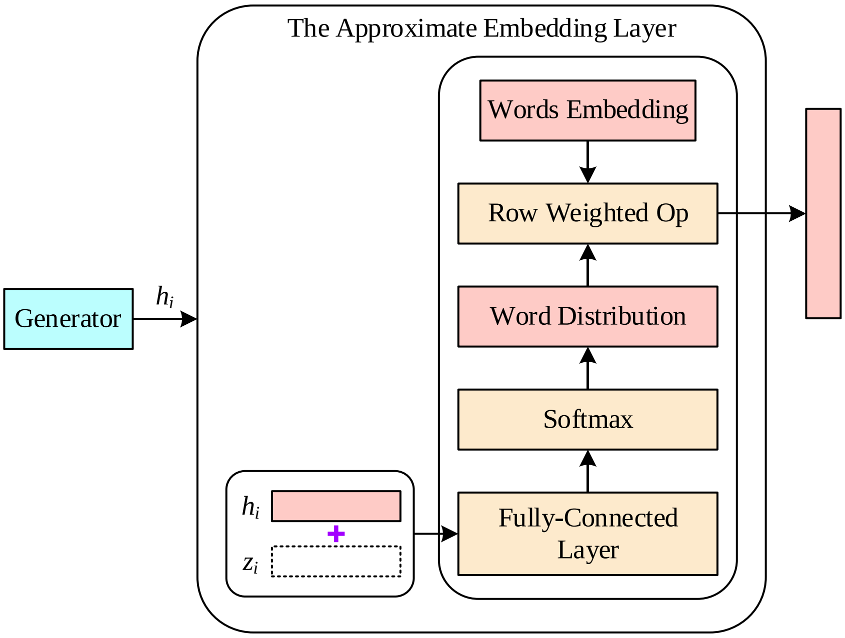


Figure 2: The structure of the word vector approximation layer

2.1.3 Convolutional neural network based discriminator

The discriminator, as an important component of the generative adversarial network, serves to measure the distance between the generated results and the true distribution, thus guiding the generator to produce results that are closer to the real data, and a discriminator based on the convolutional neural network CNN is also used in this subsection.

The discriminator has three inputs: a sequence of word vectors $V_q=\left\{e_{w_{q,1}},e_{w_{q,2}},\cdots,e_{w_{q,i}},\cdots,e_{w_{q,k}}\right\}$ for the current message q, a sequence of word vectors $V_r=\left\{e_{w_{r,1}},e_{w_{r,2}},\cdots,e_{w_{r,i}},\cdots,e_{w_{r,k}}\right\}$ for the true replies r, and a sequence of word vectors $V_{\hat{r}}=\left\{e_{\hat{w}_1},e_{\hat{w}_2},\cdots,e_{\hat{w}_i},\cdots,e_{\hat{w}_k}\right\}$ for the generation $\hat{r}$. where $V_{\hat{r}}$ is a sequence of word vectors that is directly estimated through the word vector estimation layer based on the probability distribution of the words output by the generator. For ease of computation, all word vector sequences are padded or truncated to a fixed length k.

Based on the input word vector sequences V_r, $V_{\hat{r}}$ and V_q, convolutional neural networks $CNN-r$ and $CNN-q$ are used to encode them into vectors $A_r, A_{\hat{r}}$ and A_q respectively. $CNN-r$ and $CNN-q$ mainly abstract and extract the semantic information from the word vector sequences, and their last layer is a Max-pooling pooling layer that compresses the sequence features into a vector to represent the semantic information of the sentence. Then, A_r and $A_{\hat{r}}$ are concatenated with A_q to achieve interaction with the current message through a fully connected network, thereby obtaining the conditional representation of the reply $A_{r,q}$ and $A_{\hat{r},q}$. Finally, a multi-layer neural network is used to determine the probability that r and $\hat{r}$ are the real reply based on the conditional representations $A_{r,q}$ and $A_{\hat{r},q}$, and the process is as follows:

$$D(r|q) = f(A_{r,q}) \tag{5}$$

$$D(\hat{r}|q) = f(A_{\hat{r},q}) \tag{6}$$

where $f(\cdot)$ is the discriminant function based on a multilayer neural network, where the activation function of the last layer of the network of the discriminator is a sigmoid function, and therefore it can represent the probability that r is a true response under q conditions.

For the discriminator distinguishing the true response r and generating the result $\hat{r}$ is a binary classification problem, so according to $D(r|q)$ and $D(\hat{r}|q)$, the loss function of the discriminator D is shown in Equation (7):

$$D_{loss} = -\log D(r|q) - \log(1 - D(\hat{r}|q)) \tag{7}$$

Learn the parameters in discriminator D by minimizing (7).

2.1.4 Mathematical modeling

In the above theoretical support, complete the modeling work of dialogue reply generation model based on adversarial network, the dialogue reply generation model based on adversarial network is shown in Fig. 3, the dialogue reply generation model based on adversarial network (GAN-AEL) which mainly contains three modules:

(1) Reply generator G: The generator G adopts a gated recurrent unit (GRU)-based encoding-decoding structure, in which the encoding structure first maps the input message into a real-valued vector through the word vector quantization and sequence representation process, and then generates a reply through the decoding process conditioned on this vector.

(2) Word Vector Approximation Layer (AEL): it mainly estimates the word vectors based on the word distribution output by the generator and then passes the vectors to the discriminator. At the same time, it can also pass the feedback given by the discriminator to the generator. That is, the word vector approximation layer AEL is a bridge between the generator and the discriminator, which ensures that the generator can be guided during the confrontation process.

(3) Discriminator D: The discriminator is placed on top of the word vector approximation layer to estimate the difference between the generated response and the true response. The convolutional neural network CNN abstracts the semantic representation of its sentences based on the word vectors, and then determines whether the responses are real or generated results through a multilayer neural network. Finally, the result of its judgment is used as a feedback to adjust the generator parameters and guide its generated result to be closer to the real reply.

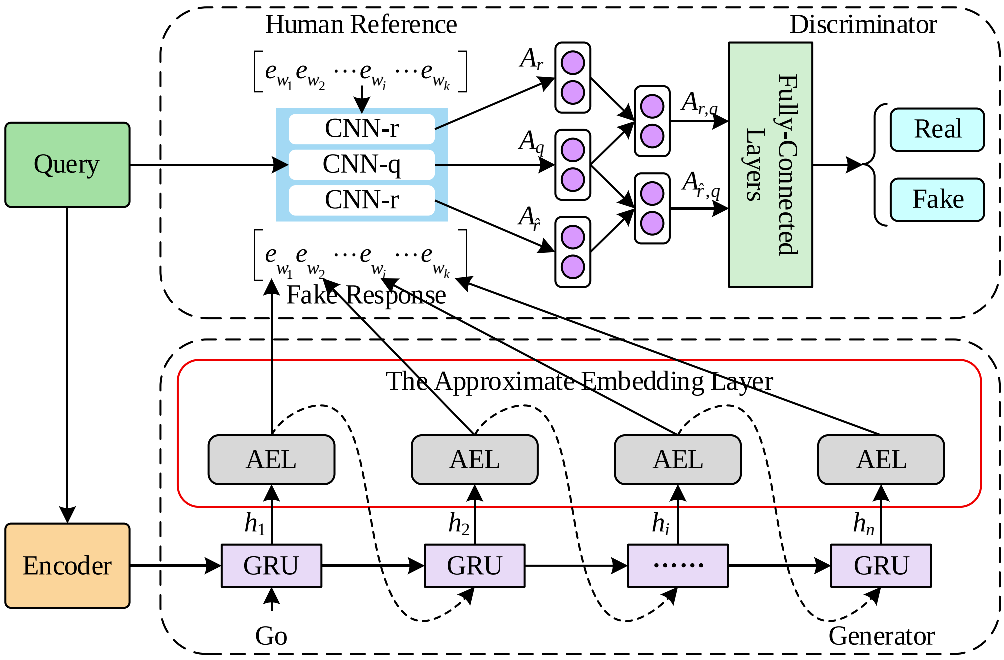


Figure 3: The Framework of GAN for Response Generation

2.1.5 Model training

In the original generative adversarial network framework, the loss function of the generator is defined as $-D_{loss}$. Since $D(r|q)$ in $-D_{loss}$ is probability estimated by a sigmoid function, it suffers from gradient saturation. , in the reply generation task, the generator needs to generate a sequence, and it is very unfavorable for the training of the generator if the feedback back from the discriminator is too little. To solve the above problem, the L_2-parameter between $A_{r,q}$ and $A_{\hat{r},q}$ is used as the loss of the generator, so the loss function of the generator in the adversarial network is as in Equation (8):

$$G_{loss} = \left\| A_{r,q} - A_{\hat{r},q} \right\| \tag{8}$$

This objective function does not have the problem of gradient saturation, and the difference between the conditional representations $A_{r,q}$ and $A_{\hat{r},q}$ can also fully reflect the gap between the true responses and the generated results under the condition of q, which is in line with the basic theory of adversarial networks. According to the loss function (8), the gradient calculation process of the generator is as in Equation (9):

$$\nabla_{g_{D,G}(\theta_G)} = \frac{\partial G_{loss}}{\partial V_{\hat{r}}} \frac{\partial V_{\hat{r}}}{\partial \theta_G} = \frac{\partial G_{loss}}{\partial V_{\hat{r}}} \frac{\partial V_{\hat{r}}}{\partial G} \frac{\partial G}{\partial \theta_G} \tag{9}$$

where θ_G denotes the parameters to be trained in the generator G, $V_{\hat{r}}$ denotes the approximated word vectors output by the word vector approximation layer AEL, and $g_{D,G}(\cdot)$ denotes the entire forward process of the adversarial network. According to the above gradient solving process (9), the feedback information G_{loss} from the discriminator D can be effectively passed to the generator G directly through the word vector approximation layer.

2.2 Intercultural English Dialogue Scene Construction

After the construction of the dialogue response generation model based on the adversarial network, this subsection of this paper identifies OpenGL as the development tool for the virtual cross-cultural English dialogue scene, combines the knowledge of spatial geometry topology and 3D Stdio MAX, completes the geometric modeling of the cross-cultural English dialogue scene, saves it as a 3ds file, and then imports it into the OpenGL software, renders the lighting and materials, and finally introduces the above constructed dialogue response generation model based on the adversarial network into the cross-cultural English dialogue scene to fit the actual situation. Then import it into OpenGL software and render the lighting and materials to make the cross-cultural English conversation scene more suitable for the actual situation, and finally introduce the above constructed adversarial network-based conversation reply generation model into the cross-cultural English conversation scene.

2.2.1 Development tools

OpenGL, a 3D graphics interface released by GI, has been recognized as the standard for high-performance graphics and interactive visual processing. It supports computer graphics algorithms such as surface modeling, graphical transformations, lighting, materials, textures, pixel manipulation, fusion, fogging, etc. Compared to technologies such as VRML, QuickTimeVR, and MUD/MOD, OpenGL has the advantages of hardware platform independence, powerful modeling capabilities, scene realism, and interactivity. Compared with technologies such as VRML, QuickTimeVR and MUD/MOD, OpenGL has the advantages of hardware platform independence, powerful modeling functions, scene realism and interactivity. Therefore, this paper proposes to use OpenGL to complete the development of virtual cross-cultural English conversation scene.

2.2.2 Geometric modeling

In the creation of the virtual cross-cultural English dialogue scene, we should first adopt different creation methods for objects with different characteristics according to the spatial geometric topology of the scene objects. For regular shape objects, you can directly use OpenGL's basic drawing functions such as points, lines and surfaces. For regular surface objects, you can use quadratic surface drawing functions such as sphere and cylinder. For non-regular surface objects, a series of discrete points, such as Bezier surfaces and NURBS surfaces, can be used to draw the surface using OpenGL's NURBS surface drawing functions. In order to improve the efficiency of scene modeling, complex objects can be modeled using 3D Stdio MAX, saved as 3ds files, and then imported into the cross-cultural English conversation scene created by OpenGL.

2.2.3 Light and materials

Objects drawn in the scene should not only have precise geometry, but also a realistic visual appearance. The appearance of an object depends not only on the object, but also on the interaction of scene lighting and surface materials. Therefore, in order to generate a realistic scene with the computer, the lighting and surface materials of the scene must be set up further. One of the start lighting is to call gEnable to determine the scene to use lighting and material effects, lighting is set to call the month gILight () to set the different light components and location, start color retrace is to call gEnabl so that the material's reflective properties and gColor located in the same color, the material attributes by the gColoMaterial () to set the surface of a certain light components of the Reflection properties.

2.2.4 Interactive dialogues

In order to realize the interactive dialogue in the scene, embedding the adversarial network-based dialogue response generation model in the scene drawing helps to establish a unified and reasonable coordination mechanism between different characters, plots, topics and environments in the dialogue scene. This approach, which is consistent with real logic, emotional display and individual needs, can guide students to deeply feel the real-life scenarios of English application on the basis of the complete knowledge system and promote the improvement of language proficiency.

2.3 Design of Foreign Trade English Teaching System

In order to make the above cross-cultural English conversation scenarios better applied to the current foreign trade English teaching in colleges and universities, a foreign trade English teaching system based on cross-cultural English conversation scenarios is designed from the user requirements and functional modules, aiming to improve the quality of foreign trade English teaching in colleges and universities.

2.3.1 User needs analysis

Research on user needs is the first prerequisite for the development of foreign trade English teaching system, through a reasonable demand analysis, in order to make the development of the product more application value and market value, you need to pay attention to the current situation of the development of the field of research in a timely manner, and carry out research to accurately grasp the direction of software development.

Requirements analysis as shown in Figure 4, Y model requirements analysis model contains four basic points: user needs, user goals, product features and Maslow needs. The first point, user needs, is both a starting point and an appearance, analyzing user behavior, demand points and applicable scenarios, and digging deep into elements such as Who, What and Where to get a viewpoint. The second point is that user goals are used to explore the deeper needs of users, and to understand the nature of user needs as well as the ultimate design goals through sociological and psychological means. The third point, the feasibility of the solution is specifically assessed by proposing solutions, choosing among Which and How that require input, and prioritizing against the input-output ratio. The fourth point, Maslow's hierarchy of needs, including physiological, safety, self-actualization, social and respect needs, the needs representation is advanced from top to bottom, when the bottom needs are satisfied, the value of self-needs is continuously improved, that is, the process of pursuing advanced needs.

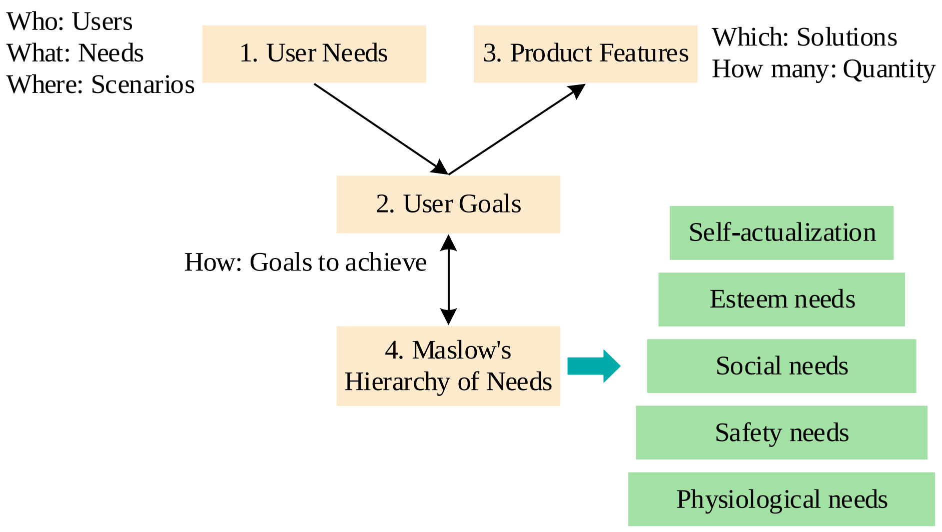


Figure 4: Demand analysis

2.3.2 System Functional Design

The functional design of the system in this paper is centered on the training of teaching English in foreign trade, and the functional design is carried out in the cross-cultural English dialogue scene, and the teaching system of English in foreign trade contains four modules, which are the module of English dialogue scene, the module of reading and writing words, and the module of following sentences and the module of multiple rounds of dialogues. Users enter the foreign trade English teaching system and experience in the cross-cultural English conversation scene. Users can record the conversation by pressing the start button, and release it to end the conversation. The voice input by the user is automatically recognized, converted from speech to text, and then returned to the dialog content through the adversarial network-based dialog response generation model, with a view to improving the user's English speaking ability. The system switches scenes at any time according to the user's needs. When the user sends a command that needs to go to the restaurant in the hotel scene, the system jumps to the corresponding scene, which is visually a virtual character leading the user into the restaurant scene, increasing the real interaction experience. Users can start a task-based multi-round dialog based on a specific scene around the cross-cultural English scene.

3 In-depth probing analysis

3.1 Model Exploration Analysis

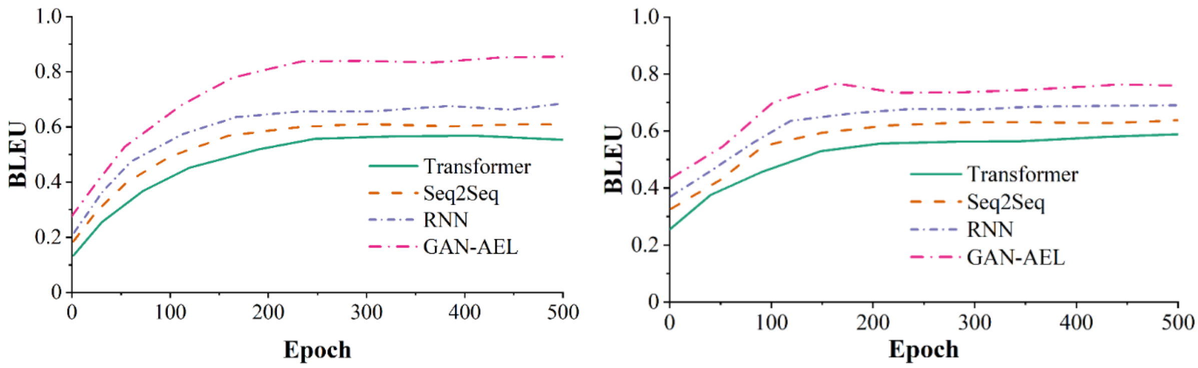
FutureBeeAI English General Conversational Text Dataset and Nemotron-CC dataset are used as the model exploration and analysis dataset, and BLEU and translation accuracy are used as the evaluation indexes, under the theoretical guidance of the evaluation indexes, the adversarial network-based conversation reply generation model is carried out. In-depth inquiry analysis.

3.1.1 Exploratory environment

In order to analyze the actual running effect of the Adversarial Network-based Conversation Response Generation Model (GAN-AEL), the study uses FutureBeeAI English General Conversational Text Dataset and Nemotron-CC dataset to train the model, in which FutureBee-AI English General Conversational Text Dataset dataset contains 10,000 English conversational text data messages covering a variety of domains such as music, health, life, and food, and the Nemotron-CC dataset contains 6.3 trillion English languages, including 1.9 trillion synthetically generated data. The study judged the translation and dialog effects using the Bilingual Evaluation Index (BLEU) metric, which indicates the similarity of the translations, with higher levels indicating better dialog effects. The study selects 2000 dialog text data in the dataset for training, and the number of iterations of the network model are all set to 500.

3.1.2 Analysis of results

The study compares the BLEU of Recurrent Neural Network (RNN), Transformer model and Sequence to Sequence (Seq2Seq) model, and the comparison of the test results of the models in different datasets is shown in Fig. 5, where (a) ~ (b) are the FutureBeeAI English General Conversational Text Dataset dataset and Nemotron-CC dataset, respectively. On the FutureBeeAI English General Conversational Text Dataset dataset, the BLEU value of the model increases with the number of iterations, and tends to a relatively stable state after reaching a certain value, in which the value of GAN-AEL is the highest, with the highest value of 0.821, and the BLEU value of RNN is the lowest The highest value is only 0.563, which is reduced by 0.258 compared to the research using the network.The dataset Nemotron-CC is the same, which shows that GAN-AEL can be more effective in cross-cultural English conversation.

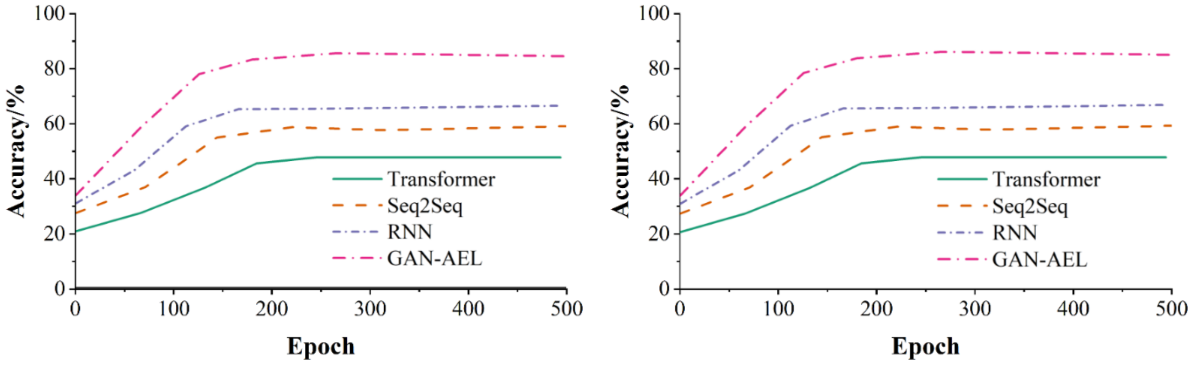


(a)FutureBeeAI English General Conversational Text Dataset

(b)Nemotron-CC Dataset

Figure 5: Comparison of test results of different dataset models

The results of the dialog translation accuracy of different models are shown in Figure 6, in FutureBeeAI English General Conversational Text Dataset dataset, the accuracy of different models will increase with the number of iterations, its accuracy will fluctuate and change the trend, in which the GAN-AEL's accuracy performance is higher, its The accuracy rate of GAN-AEL is higher, with a maximum of 89.52%, while the accuracy rate of RNN is relatively low with a maximum value of only 57.64%, which is 31.88% lower compared to the research using the network, and the same is true for the Nemotron-CC dataset. This shows that GAN-AEL has the highest translation accuracy in different datasets tested English translation is better.

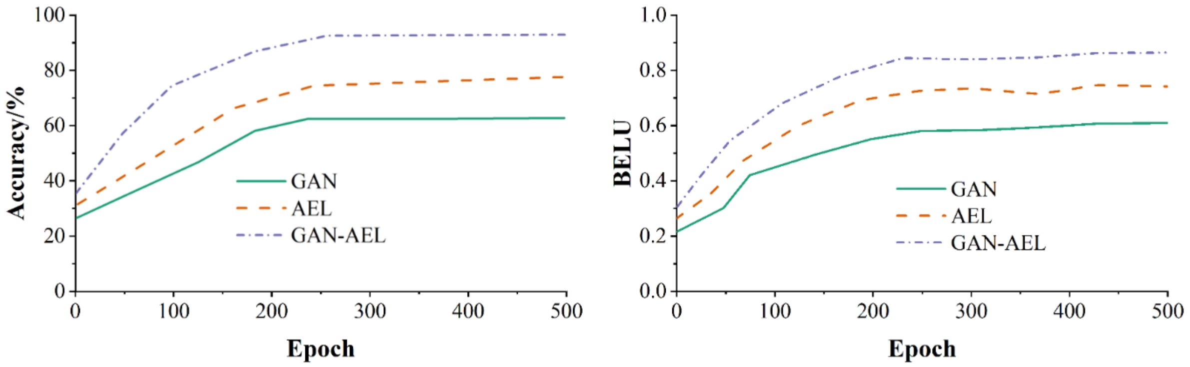


(a)FutureBeeAI English General Conversational Text Dataset

(b)Nemotron-CC Dataset

Figure 6: Comparison of dialogue translation accuracy rates among different models

Ablation experiments are conducted to demonstrate the facilitating effect of the word vector approximation layer (AEL) on the generative adversarial network (GAN), using the dataset Future BeeAI English General Conversational Text Dataset, and the results of the ablation experiments are shown in Fig. 7, where (a) ~ (b) are the translation accuracy, BLEU values, respectively. . The highest translation accuracy of GAN before adding the word vector approximation layer (AEL) is only 72.57%, which is 18.95% lower than that of the improved model, and the translation accuracy of word vector approximation layer (AEL) is 60.62%, which is 30.90% lower than that of GAN-AEL. The highest BLEU value of the word vector approximation layer (AEL) is only 0.512, which is 0.309 lower compared to the improved GANs, and the highest BLEU value of the GAN is only 0.705, which is 0.116 lower compared to the GAN-AEL.It can be seen that there is a significant improvement in the performance of the adversarial network-based dialog reply generation model after adding the word vector approximation layer (AEL).



(a)Translation accuracy rate

(b)BLEU

Figure 7: Results of the ablation experiment

The analysis of the effect of cross-cultural English dialogues is shown in Table 1, after using GAN-AEL, not only will the questions posed be translated into English, but also be able to carry out dialogues on their own, such as “Have you eaten yet?” Conducting a conversation can autonomously recognize the translation and can recognize the question to conduct a conversation. While in GAN, AEL, although it can realize autonomous translation of the content, but the effect of autonomous dialogue is poor, some of the answers are not perfect, comprehensively verified, based on the adversarial network of dialogue response generation model application effect.

Table 1: Analysis of the Effect of Cross-cultural English Conversations

Model	Single dialogue	Multi-round dialogue
GAN-AEL	Have you eaten? -No, I don’t need to eat.	What are the specialisations of the university?-English
	Suggest a film for me. -Shawshank’ s Redemption	What are the language specialisations?- English, Japanese, etc
GAN	Have you eaten?	What are the specialisations of the university?
	Suggest a film for me.	What are the language specialisations?
AEL	Have you eaten?	What are the specialisations of the university?
	Recommend a film	What are the language specialisations? - English.

3.2 Exploration and Analysis of Foreign Trade English Teaching System

After verifying the adversarial network-based dialog response generation model (GAN-AEL), with the support of the development software, it is introduced into the cross-cultural English dialog scene, and then a foreign trade English teaching system based on the cross-cultural English dialog scene is designed, and the practical value of the system is confirmed from the aspects of performance test and application effect, aiming at improving the level and quality of foreign trade English teaching in colleges and universities. The purpose is to improve the level and quality of foreign trade English teaching in colleges and universities.

3.2.1 System performance testing

Performance testing is technically more complex than functional testing. In the old days, performance testing was only an optional part of the project testing process. However, with the continuous development of testing technology, performance testing has gradually begun to be emphasized and made independent. In order to better highlight the priority of the foreign trade English teaching system based on cross-cultural English conversation scenes, the traditional teaching system is used as a reference. The English conversation scene module is tested as shown in Table 2; the reading and writing words module is tested as shown in Table 3; the following sentence module is tested as shown in Table 4; and the multiple rounds of conversation module is tested as shown in Table 5. The system is set to concurrently concurrently concurrently with 50 users, and then it is gradually incremented to 100 users, 1,500 users, and all the way up to 1,000 users, and it is observed that the system's response time is whether or not it meets the user's acceptable time. According to the table, it can be seen that with the increase of the number of concurrent users, the response time of the foreign trade English teaching system based on cross-cultural English dialog scenes, and the traditional teaching system also show a monotonically increasing trend. Taking the English conversation scene module as an example, when the number of concurrent users is 50, the response time gap between the foreign trade English teaching system and the traditional teaching system is 1ms, and as the number of concurrent users grows to 650, the response time gap between the foreign trade English teaching system and the traditional teaching system reaches 19ms, which fully verifies that the foreign trade English teaching system based on the cross-cultural English conversation scene has a high priority, and the same is true for the remaining three modules. The same is true for the remaining three modules, which are based on the data in the table.

Table 2: Test of the English dialogue scenario module

User	Response time/ms		User	Response time/ms	
	Foreign trade English teaching system	Traditional teaching system		Foreign trade English teaching system	Traditional teaching system
50	6	7	550	59	68
100	9	25	600	62	70
150	22	42	650	69	88
200	26	34	700	72	86
250	32	44	750	76	79
300	42	50	800	79	94
350	47	54	850	81	94
400	50	56	900	86	99
450	52	58	950	89	91
500	56	63	1000	93	102

Table 3: Word reading and writing module test

User	Response time/ms		User	Response time/ms	
	Foreign trade English teaching system	Traditional teaching system		Foreign trade English teaching system	Traditional teaching system
50	6	14	550	59	61
100	9	19	600	62	69
150	22	27	650	69	75
200	26	30	700	72	79
250	32	40	750	76	80
300	42	42	800	79	82
350	47	47	850	81	83
400	50	50	900	86	96
450	52	59	950	89	102
500	56	60	1000	93	107

Table 4: Test of the sentence shadowing module

User	Response time/ms		User	Response time/ms	
	Foreign trade English teaching system	Traditional teaching system		Foreign trade English teaching system	Traditional teaching system
50	7	20	550	60	67
100	16	25	600	61	67
150	17	27	650	65	74
200	20	30	700	65	75
250	21	36	750	66	78
300	26	37	800	70	82
350	40	44	850	84	92
400	50	54	900	85	92
450	51	56	950	86	95
500	60	61	1000	88	101

Table 5: Multi-round dialogue module testing

User	Response time/ms		User	Response time/ms	
	Foreign trade English teaching system	Traditional teaching system		Foreign trade English teaching system	Traditional teaching system
50	3	6	550	41	51
100	4	6	600	52	57
150	14	24	650	55	59
200	19	24	700	57	63
250	24	27	750	58	72
300	25	39	800	60	80
350	32	42	850	65	81
400	32	44	900	75	89
450	33	45	950	86	92
500	38	50	1000	95	103

3.2.2 Effectiveness of system applications

Through the performance test of the foreign trade English teaching system, the response time of each functional module of the system is within the acceptable range, and the priority of the foreign trade English teaching system based on cross-cultural English conversation scenarios is also verified. In order to make the research results more convincing, the system application effect analysis is carried out, in which the most intuitive application effect of the teaching system is the students' English scores, and the students' English scores can be subdivided into English writing scores (0~30 points), English reading scores (0~30 points), English speaking scores (0~30 points), English listening scores (0~30 points), which can be obtained from the foreign trade English test papers. These scores can be obtained through the foreign trade English test paper. 10 students are randomly selected from the university students as the system experience users, and the experience time is one semester. The analysis of English writing scores is shown in Table 6; English reading scores are shown in Table 7; English speaking scores are shown in Table 8; and English listening scores are shown in Table 9. Through the data performance in the table, it can be seen that, for example, in the case of English writing performance, before the 10 students did not experience the foreign trade English teaching system, their scores were distributed in the range of 6~15 points, after experiencing the foreign trade English teaching system, their scores were distributed in the range of 21~29 points, and their writing performance was improved in the range of 7~21 points, and the rest of the English reading scores, English speaking scores, and English listening scores were the same, and the all-around It verifies the validation of the foreign trade English teaching system based on cross-cultural English conversation scenarios, and also provides theoretical references for the research on the application of foreign trade English teaching by generating adversarial networks to simulate real cross-cultural English conversation scenarios.

Table 6: Analysis of English writing performance

Student	Before	After	Difference
1	6	23	17
2	13	27	14
3	7	28	21
4	14	29	15
5	13	21	8
6	9	24	15
7	15	26	11
8	6	24	18
9	8	24	16
10	14	21	7

Table 7: Analysis of English Reading Scores

Student	Before	After	Difference
1	15	24	9
2	13	21	8
3	11	23	12
4	13	26	13
5	11	28	17
6	8	27	19
7	10	22	12
8	12	29	17
9	8	26	18
10	11	26	15

Table 8: Analysis of Oral English scores

Student	Before	After	Difference
1	14	21	7
2	7	26	19
3	12	24	12
4	13	23	10
5	9	27	18
6	11	24	13
7	10	29	19
8	11	21	10
9	13	29	16
10	7	23	16

Table 9: Analysis of English Listening Scores

Student	Before	After	Difference
1	12	23	11
2	10	29	19
3	14	20	6
4	11	25	14
5	5	22	17
6	12	20	8
7	11	28	17
8	14	28	14
9	14	30	16
10	11	29	18

4 Conclusion

As an international common language, English plays a crucial role in cross-cultural communication, but at present the language barrier is still one of the main factors hindering effective communication. For this reason, this paper designs a foreign trade English teaching system based on cross-cultural English conversation scenarios with the support of generative adversarial network and system development tools, and conducts in-depth investigation and analysis of the system.

(1) Take the English conversation scene module as an example, when the number of concurrency of the system is 50 users, the response time difference between the foreign trade English teaching system and the traditional teaching system is 1ms, and when the number of concurrency grows to 650 users, the response time difference between the two reaches 19ms, which indicates that compared to the traditional teaching system, the performance of the foreign trade English teaching system based on the cross-cultural English conversation scene is more outstanding, which will bring more quality experience to the users.

(2) Before 10 students used the foreign trade English teaching system, their scores ranged from 6 to 15, and after one semester of the foreign trade English teaching system, their writing scores increased from 7 to 21, which confirms the practical application effect of the foreign trade English teaching system based on the cross-cultural English dialogue scenario.

About the Author

Xiaofeng Wang (b. September 4, 1987), Female, native of Hengyang, Hunan, graduated from China University of Geosciences (Wuhan) in 2012 with a Master’s degree in Foreign Linguistics and Applied Linguistics, currently works at College of Arts and Sciences in Xi’an University of Architecture and Technology Huaqing College as a Lecturer with the main research direction focusing on College English Education and English in Foreign Trade.

References

[1] Qian, P. (2020). New thinking on the teaching mode of international economy and trade under the demand of innovative international talents. Int. J. Comput. Eng, 5(3).

[2] Roshid, M. M., & Kankaanranta, A. (2025). English communication skills in international business: Industry expectations versus university preparation. Business and Professional Communication Quarterly, 88(1), 100-125.

[3] Azizovich, A. N. M., & Murtoza o’g’li, D. S. (2025). ENGLISH AS A TOOL FOR INTERNATIONAL BUSINESS AND TRADE. Shokh Articles Library, 1(2).

[4] Wang, D., Fan, D., Freeman, S., & Zhu, C. J. (2017). Exploring cross-cultural skills for expatriate managers from Chinese multinationals: Congruence and contextualization. Asia Pacific Journal of Management, 34(1), 123-146.

[5] Pop, A. M., & Sim, M. A. (2022). Cross-cultural communication in multinational companies. The Annals of the University of Oradea. Economic Sciences, 31(31), 1.

[6] Amirovich, K. S., Zamirovich, G. E., Lutpullayevna, A. F., Nabiyevna, I. Z., & Saliyevna, S. D. (2021). Cross-cultural context of English language teaching. Annals of the

Romanian Society for Cell Biology, 25(2), 3737-3750.

[7] Xie, Q. (2022). Applying problem-based approach in business English courses for China's English majors. Sage Open, 12(2), 21582440221093349.

[8] Liu, Y. (2023). Effective Teaching Research on Cross-Cultural Understanding in English at Vocational Schools from the Perspective of Core Competencies. Frontiers in Educational Research, 6(28).

[9] Wen, Y., & Meng, J. (2024, December). Diversified Generation and Evaluation of English Cultural Communication Materials Based on Generative Adversarial Networks (GANs). In 2024 IEEE 4th International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA) (Vol. 4, pp. 636-640). IEEE.

[10] Zhang, L. (2022). An assisted teaching method of college English translation using generative adversarial network. Mobile Information Systems, 2022(1), 5408309.

[11] Zhang, L., Shu, C., Guo, J., Zhang, H., Xie, C., & Liu, Q. (2020). Generative adversarial network-based neural audio caption model for oral evaluation. Electronics, 9(3), 424.

[12] Gao, H. (2025). Optimization Method for Academic English Content Based on Generative Adversarial Networks and Data Augmentation. Computers and Education: Artificial Intelligence, 100492.

[13] Adversarial, R. G. (2024). Assessment and Prediction of University English Teaching Effect Using Progressive Recurrent Generative Adversarial Network. J. Electrical Systems, 20(3s), 2335-2346.

[14] Song, D. (2025, July). Construction and Effect Evaluation Of Virtual Simulation Education Scenarios for the Five-Education development driven by Generative Adversarial Network (GAN). In 2025 International Conference on Education Reform, Ideology and Politics (ERIP 2025) (pp. 386-395). Atlantis Press.

[15] Li, H., & Liu, Y. (2025, June). Research on Optimization Algorithm of Generative AI in Design of Personalized Learning Path for English. In 2025 IEEE 3rd International Conference on Image Processing and Computer Applications (ICIPCA) (pp. 1109-1114). IEEE.