ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

A Study of Deepseek Tool to Optimize Intercultural Communication and Emotional Expression Supported by Multicultural Contexts

Hanyan Yu[1] *and Yan Zhuang*[2,*]

[1] School of General Courses, Xiamen Ocean Vocational College, Xiamen, Fujian, 361102, China

[2] School of Liberal Arts Education and Art Media, Xiamen Institute of Technology, Xiamen, Fujian, 361021, China

SUMMARY: *Emotions come from the long-term evolution of daily language, DeepSeek-like generative artificial intelligence can gain insights into the different language habits and multicultural backgrounds implied in users' utterances, thus providing a new source of motivation and strong technical support for the continuation of the vitality of multicultural civilization. In this paper, we take advantage of the core technology of DeepSeek tool to calculate user's emotions with its large language model, and further propose a deep learning-based emotion dialog generation model. The emotional dialog generation model is combined with the bionic intelligent technology development of "bionic intelligence" to design the emotional expression of humanoid robots. Analyze the performance of the proposed emotional dialog generation model based on recurrent convolutional network on different indexes, and analyze the feasibility of DeepSeek accessing multi-channel apps for effective emotion recognition. The humanoid robot is subjected to experiments on multi-language emotion expression to analyze the possibility of its multi-language communication. DeepSeek is able to distinguish the emotion expression and emotion intention of different app users by accessing to multiple apps. The humanoid robot's multiple emotional expressions can be universally recognized and can be used as an innovative way of development in multicultural communication.*

KEYWORDS: *DeepSeek; Emotional dialog generation; Recurrent Convolutional Networks; Humanoid Robots*

1 Introduction

With the process of globalization, cross-cultural communication has become increasingly important [1]. When communicating with people from other cultures, understanding and using appropriate emotion expression skills can help us better establish and maintain interpersonal relationships [2, 3]. And in cross-cultural communication, the use of emotion expression skills is crucial for establishing and maintaining good interpersonal relationships [4]. By reflecting on self-expression of emotions, listening to and observing others' expressions, rationally using verbal and nonverbal expressions, and respecting and tolerating different habits of emotional expression, we can better understand and communicate, promote mutual respect and cooperation between cultures, and realize true cross-cultural communication [5-8].

*daisychng324@163.com
https://doi.org/10.65102/is20261139
Published: 10 June, 2026

With the development of artificial intelligence, a variety of intelligent tools are widely used in cross-cultural communication, among which the emergence of Deepseek has revolutionized the field [9]. DeepSeek is an emerging technology in the field of artificial intelligence, which has its own unique features and advantages [10, 11]. Unlike many traditional AI algorithms, DeepSeek leaves behind some common fixed patterns and works to understand and process data in a completely new way [12, 13]. Instead of simply following existing algorithmic frameworks, it builds its own logic for understanding various problems from the ground up [14]. This unique approach allows DeepSeek to show strong adaptability and innovation when facing complex and changing real-world problems [15, 16]. Due to DeepSeek's sharp semantic comprehension, sentiment analysis, semantic understanding and logical reasoning, it is able to deeply analyze the structure and meaning of every word and sentence in the text [17-19]. In cross-cultural communication, it can also accurately grasp some obscure emotional expressions through contextual correlation and semantic reasoning, so as to more accurately judge the emotional tendency of the text and optimize the emotional expression in communication [20-22].

In this paper, we propose an emotional dialog generation model based on deep learning by combining DeepSeek's computation of user's emotions from the perspective of large language model with the bionic intelligence technology of “Bionic Intelligence”. The model is upgraded with Seq2Seq as the main model, and the attention mechanism is introduced to form an emotional dialog generation model based on recurrent convolutional network. Design a robot system to simulate the biological brain for emotional expression, and propose and design the humanoid robot for emotional expression in conjunction with the requirements of multicultural background. Analyze the performance of the emotional dialog generation model based on recurrent convolutional networks on different datasets. Access DeepSeek tool to multiple apps for emotion recognition experiments and humanoid robot multilingual emotion expression experiments.

2 Research base

2.1 Bionic Intelligence Technology for “Bionic Intelligence”

Bionic intelligence has become an interdisciplinary subject spanning many disciplines such as physics, mathematics, biology, chemistry, management, information science, system science and sociology, and plays an increasingly irreplaceable role in many scenarios such as social production, human life and military defense [23-25].

From the perspective of imitation object, the current research and application of bionic intelligence mainly focuses on the two fields of “imitation of intelligence” at the micro level and “imitation of shape” at the macro level, trying to develop high-end intelligence originated from the imitation of living creatures respectively. Among them, "imitation intelligence" focuses on simulating the structure and operation mechanism of the biological brain (which can be regarded as simulating the "spiritual world"), and "imitation" focuses on imitating the appearance and movement mechanism of organisms (which can be regarded as simulating the "physical world").

As an important part of bionic intelligence, cognitive computing that mimics the biological brain aims to simulate the in-depth interaction and continuous learning process between the human brain (or other biological brain) and the environment, so as to have certain cognitive abilities of high-IQ organisms at the functional level, so as to complete specific cognitive tasks such as discovery, learning, understanding, reasoning, and decision-making of "data-information-knowledge". To help decision makers gain insight into the value of different types

of massive data, and ultimately explore and develop general artificial intelligence along the line of “neuroscience-based and computational science-supported”.

Human brain (as well as other biological brain) activities are complex and continuous dynamical processes, the complexity of which is far beyond the upper limit of the current computing resources to simulate. The human brain has three major characteristics relative to traditional computers: first, low energy consumption, the power of the human brain is about twenty watts, while the current supercomputers that try to simulate the human brain need several million watts. The second is fault tolerance; the human brain loses neurons all the time without affecting the brain's information-processing mechanisms, whereas a microprocessor can't function if it loses a transistor. Third, no programming is required; the human brain learns and changes spontaneously as it interacts with the outside world, without having to follow paths and branches limited by preset algorithms, as programs implementing artificial intelligence do, and is one of the most promising paths to general-purpose AI. Cognitive computing research centered on imitating the biological brain has begun to receive more and more attention from academia and industry, and is considered to be one of the most important development directions in the “Post-Moore Era”, and may become a breakthrough for future intelligent computing.

2.2 DeepSeek Core Technology Advantages

2.2.1 Multi-pronged Latent Attention

DeepSeek's Multiple Latent Attention (MLA) technology provides significant resource savings while maintaining model capabilities through intelligent compression and dynamic computational optimization [26, 27]. Its core advantages can be summarized as follows.

(1) Efficient memory storage.MLA adopts optimized attention computation to enhance long text processing capability while reducing computational resources. Through efficient storage and computation optimization, DeepSeek's processing capability in long text tasks far exceeds that of the traditional Transformer structure.

(2) Intelligent information screening. Dynamically adjust the information retention ratio according to the task difficulty, for example, retaining more details in mathematical reasoning, reducing redundant calculations in simple conversations, and reducing the overall computation.

(3) Hardware depth adaptation. Optimize the computation process for different graphics cards, and increase the speed by several times compared with traditional methods with almost no loss of accuracy. This technology enables large models to run smoothly on ordinary graphics cards, which is especially suitable for dealing with long document analysis, code generation and other scenarios, providing a technical foundation for the popularization of AI applications.

2.2.2 DeepSeekMoE Architecture

DeepSeek's Mixed Expert (MoE) architecture significantly improves model efficiency through intelligent task allocation and dynamic resource scheduling, and its core design is as follows.

(1) Refined expert division of labor: DeepSeekMoE adopts a dynamic expert activation strategy, where each token activates only part of the experts, which reduces computational redundancy and improves reasoning efficiency.

(2) Hybrid parallelization strategy. Combining the expert parallelism and data parallelism techniques with the dynamic routing algorithm to reduce cross-node communication, the training speed is increased several times.

(3) Resource optimization mechanism. Real-time monitoring of hardware load, dynamic adjustment of redundant expert distribution, significantly improve resource utilization and reduce memory occupation. This architecture has outstanding performance in code generation,

long text comprehension and other scenarios, with inference speed up to two times that of traditional solutions, and has been widely used in high-precision fields such as finance and industry.

2.2.3 Multi-token prediction

DeepSeek's multi-token prediction (MTP) technology breaks through the single-step prediction limitation of traditional autoregressive models through layered decoding architecture and dynamic routing optimization, and its core mechanisms are as follows.

(1) Layered prediction: MTP adopts a layered prediction structure, where the main model is responsible for the basic prediction, multiple MTP modules predict future tokens, and the synergistic optimization of multi-token prediction is achieved by sharing the embedding layer and output header.

(2) Dynamic depth adjustment: MTP module adjusts token allocation through sequence-level load balancing.

(3) Speculative decoding acceleration: MTP accelerates the reasoning by precomputing the future token probability distribution, providing a new efficiency-accuracy balance paradigm for long logic chain tasks.

2.3 DeepSeek empowers the dissemination of multicultural development

2.3.1 Generative Artificial Intelligence Enabling Multicultural Communication

Based on the inheritance and convergence of many advanced technologies such as Natural Language Processing (NLP), Computer Vision (CV), Reinforcement Learning (RL), and Multi-Modal Convergence, DeepSeek significantly improves the operational efficiency of the model through the technological innovations of the Mixed Expert (Mo E) architectural model, Multi Leader Potential Attention (MLA) mechanism, and the DualPipe Algorithm, which becomes a disruptive integration of the organic links and integration of the modern digital technologies. The rise of DeepSeek-like generative AI is not only a concentrated manifestation of multidisciplinary cross-fertilization, but also expands the scope of its ability to analyze and solve complex problems, providing a brand new power source for the continuation of the vitality of multicultural civilization.

(1) Multi-technology integration of DeepSeek-type generative artificial intelligence helps explore and trace the origin of multiculturalism

Firstly, based on natural language processing technology, it converts people, events, objects and other factors containing multicultural resources into data language, then abstracts regularities or correlations through deep learning neural network models, and extracts abstract contents such as emotions, imagery expressions and national cultures based on the collaborative processing of multimodal large models. For example, the semantic parsing of oracle bones and documents by natural language processing technology, and the multi-spectral analysis of Dunhuang murals by computer vision technology.

(2) DeepSeek generative AI helps to expand the content material of multiculturalism.

DeepSeek generative AI focuses on the richness and diversity of the data corpus, especially information, data, facts and other types of content, showing a unique ability to create and generate, and humans are good at concepts, symbols, opinions, and other content levels to form a complementary advantage.

DeepSeek-type generative AI can provide basic materials for innovating content topics, product forms, and refining iconic new concepts, such as launching large-volume, short and fast news reports and cultural products. In the user evaluation, DeepSeek has demonstrated its excellent ability to write ancient poems according to the rhyme of the words, to synthesize

arguments to write current affairs commentaries, or to imitate the sequel of a novel according to the style of the writer.

2.3.2 Multiple Corpora Deepen Multicultural Emotional Identity

Emotions come from the long-term evolution of everyday language, and it is a mistake to believe that “intelligence and emotion are separate from each other”. DeepSeek generative artificial intelligence can be sensitive to the different linguistic habits and multicultural backgrounds implied in the user's utterances, and is able to mine and generate potential common issues, creating a “zero distance” immersive field between audiences and multiculturalism, thus bringing the psychological distance closer and triggering emotional perception. It can explore and generate potential common issues, creating a “zero-distance” immersive field for audiences and multiculturalism, which in turn can narrow the psychological distance and trigger emotional perception.

Emotional recognition of multiculturalism lies not only in the surface of speech, but also in its deep logical persuasion; DeepSeek-type generative AI, on the basis of establishing an emotional connection with the user, shows logical reasoning ability close to or even surpassing that of the user, and is able to produce content containing complex cognitive structures such as emotional counseling, diagnosis of medical conditions, and in-depth sharp evaluation. This human-needs-centered “all-around assistant” role not only meets the diversified demands of individuals on the functional level, but also realizes electronic care and response to users on the emotional dimension. This provides a technologically mediated path to promote users' emotional recognition of cross-cultural communication.

3 DeepSeek-based humanoid robot system for emotional expression

3.1 User Sentiment Computation for DeepSeek in the Perspective of Large Language Modeling

DeepSeek is an intelligent tool developed by DeepSeek, aiming to provide efficient and accurate information retrieval and data analysis services through advanced artificial intelligence technology. As a large language model, DeepSeek is not only extremely adaptable, open and scalable, but also has excellent computational performance and multimodal support.

As a class of intelligent assistants based on large-scale pre-trained language models, DeepSeek adopts a variety of deep learning and reinforcement learning techniques to handle complex content generation tasks, and realizes text comprehension, generation, translation and summarization functions by pre-training on massive text data. It can be seen that DeepSeek features multi-modal learning, distributed computing, and natural language processing.

User Emotion Computing is the underlying technology of DeepSeek, which aims to recognize and understand human emotional states by analyzing text, speech or other forms of input, and is mainly composed of emotion recognition, emotion analysis, emotion generation, and context-awareness technologies. The application of user emotion computing technology enables DeepSeek to not only understand and execute commands, but also to perceive and adapt to human emotions, thereby enhancing the overall experience of user-system interaction.

First, with the support of DeepSeek's user emotion computing technology, it can intelligently identify the user's emotional needs at the levels of subjective expectations and objective behaviors, and capture the user's emotional changes during the service process, so as to improve the accuracy and robustness of the user's demand analysis. Secondly, DeepSeek's

user emotion computing technology is integrated into discipline navigation systems, academic search engines, online Q&A platforms and other tools, which can comprehensively analyze users' emotions in different contexts, and serve cultural communication and emotion transfer in different cultural contexts.

3.2 Deep Learning Based Emotional Dialogue Generation Modeling

In this chapter, deep learning techniques are used to extract emotional representations of text, and a recurrent convolutional network-based emotional dialog generation model incorporating static emotion vectors and emotion transfer models is proposed using the Seq2Seq model as the basic framework.

3.2.1 The Seq2Seq model

Seq2Seq model is an “Encoder-Decoder” structure network model for mapping one sequence to another, also known as NMT model [28]. Seq2Seq model has already achieved good results in many tasks, and is now widely used in the fields of machine translation, text generation, language modeling, and speech recognition.

The Seq2Seq model is learned in a way that involves two main processes: firstly, a variable-length input signal sequence is converted into a fixed-length vector representation in the Encoder. Secondly, a variable-length signal sequence of the target is generated from this fixed-length vector expression in the Decoder. “The structure of the Encoder-Decoder model is shown in Figure 1.

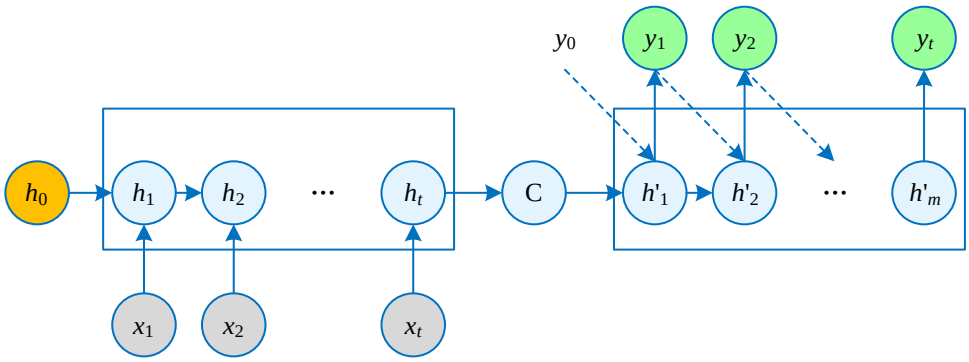

Figure 1: Structure diagram of "Encoder-Decoder" model

The Encoder is usually an RNNs that will read each character in the input sequence $X\left(x_1, x_2, \ldots, x_t, \ldots\right)$ in sequential order. As each symbol is read, the hidden state of the RNNs changes according to Equation (1). Eventually, after reading to the end of the sequence, the hidden state of the RNNs represents the feature information of the whole input sequence C. Namely:

$$h_t = f\left(h_{t-1}, x_t\right) \tag{1}$$

In this equation, h_t represents the hidden state of the RNNs at the current time step t. The f represents a nonlinear activation function.

The Decoder in the Seq2Seq model is usually also an RNNs, which generates the next character based on the hidden state h_t given by the Encoder, which in turn yields the output

sequence. Unlike the Encoder's hidden state update mechanism, the update of h'_m in the Decoder requires not only h'_{m-1}, but also relates to y_{t-1} and C, which is given by the following update formula:

$$h'_m = f\left(h'_{m-1}, y_{t-1}, C\right) \tag{2}$$

This yields the hidden state h'_m of the decoder. Also from Fig. 1, it can be seen that the conditional distribution of the next character can be computed from h_m', y_{t-1} and C:

$$P\left(y_m \mid y_{m-1}, y_{m-2}, \ldots, y_0, C\right) = g\left(h'_m, y_{m-1}, C\right) \tag{3}$$

Seq2Seq model is outstanding in text tasks, but it has some problems. Among them, the Seq2Seq model generates the target text as follows:

$$\begin{cases} y_1 = f(C) \\ y_2 = f\left(C, y_1\right) \\ y_3 = f\left(C, y_2, y_1\right) \end{cases} \tag{4}$$

It can be seen that the semantic encoding C of the input sentence used by the model is the same regardless of any target word generated, and thus there are two drawbacks. First, for the Encoder's last hidden state information, which has a large correlation with the end-of-sentence vocabulary, its information about the beginning of the sentence is easily lost. Therefore, especially for some too long sentences, a lot of information will be lost in the last state vector, which leads to a significant decrease in model performance. Secondly, the practice of giving the same weight to each word in a sentence leads to insufficient differentiation of each word, which is not reasonable.

Two improved versions of the attention mechanism, Global Attention and Local Attention, are proposed for the fixed state vector problem. When computing the context vector c_t's, Global Attention pays attention to each hidden state generated by the encoder. Assuming that h_t denotes the state of the decoder at moment t, the length of the input sample sequence is n, and that the full hidden state of the encoder is $C = \left\{c_1, c_2, \ldots, c_n\right\}$, then the context vector c_t can be computed in the following weighted way to obtain it:

$$c_t = \sum_{i=1}^{n} a_t c_i \tag{5}$$

where a_t^i is a calibrated vector of length n, which represents the importance of each hidden state sequence at time step t, computed using the Softmax method:

$$a_t^t = \frac{\exp\left(score\left(h_i, c_i\right)\right)}{\sum_{j=1}^{n} \exp\left(score\left(h_i, c_j\right)\right)} \tag{6}$$

where c_i denotes the hidden state of the encoder at time step i. c_j denotes the hidden state

of the encoder at time step j. n is the length of the input sequence.The Score function denotes an arbitrary function of the comparison vectors, which is generally commonly used in the way of dot product computation and the use of parameter matrices. Three commonly used methods are proposed as follows:

$$score\left(h_i, c_i\right) = \begin{cases} h_i^T c_i \\ h_i^T W_a c_i \\ v_a^T \tanh\left(W_a \left[h_i; c_i\right]\right) \end{cases} \tag{7}$$

where W_a and v_a are learnable parameters. The $\left[h_i; c_i\right]$ denotes the splicing of the two vectors.

It can be seen that by introducing the attention mechanism, for each word to be predicted, there is a different hidden state c_t of the encoder, which makes the model pay attention to each element in the prediction, thus improving the prediction performance of the model.

3.2.2 Problem definition

In some cases, the chatbot can be instructed to generate a reply that matches by specifying the type of emotion. And more naturally, the chatbot should be able to sense the user's emotional state so as to give a reply that matches the current emotional context.

For the first case above, for a given input $X = \left(x_1, x_2, x_3, \ldots, x_n\right)$ and emotion category, to generate a response $Y = \left(y_1, y_2, y_3, \ldots, y_m\right)$, where the categories of emotions can be categorized into six $\left\{Anger, Disgust, Happiness, Like, Sadness, Other\right\}$, the tasks to be accomplished in this paper are shown in Table 1.

Table 1: Task schedule

Given source sentence	Given emotional classification	Reply
It's great to eat delicious food today.	Like	Delicious food makes people happy
I didn't hit the car, delayed the trip, and didn't see my favorite star.	Sadness	That's a pity

3.2.3 Modeling of Emotional Dialogue Generation

(1) Sentiment Analysis Model

In this paper, we wish to design a text sentiment classification model capable of combining the ability of recurrent neural networks to capture sequential dependencies at a distance and the ability of convolutional neural networks to capture localized features. In order to realize such an effect as much as possible, the following recurrent convolutional structure network is designed.

Step1: Word Embedding Layer. Pre-trained word vectors from word2Vec are used for initialization in this layer.

Step2: Bidirectional recurrent neural network layer. In this layer LSTM is used as the basic unit of the recurrent neural network to obtain the state of the hidden layer in both directions at each moment.

Step3: Convolutional layer. The bidirectional recurrent neural network has two hidden layer

states in two directions corresponding to each moment, and the outputs in these two directions are convolved, i.e., undergo a nonlinear change.

Step4: Pooling layer/attention layer. In this layer, the attention mechanism can be used to compute the effect of the convolution result at each moment on the final classification result.

Step5: Fully connected layer. The final sentiment categories are computed using the fully connected layer Softmax function on the results of the previous step.

In the above model, it is considered that at the moment t, $\overrightarrow{h_t}$ captures the input information and the semantic information to its left under that moment. The localized representation of this position under the t moment is obtained by a single convolution operation. The convolution result for the t moment is computed by equation (8) as follows:

$$h_t = \mathrm{F}\left(W\overrightarrow{h_t} + U\overleftarrow{h_t} + b\right) \tag{8}$$

where F is a nonlinear transformation function, $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are the implied layer states of the recurrent neural network in both directions at the moment of t, W and U are the parameter matrices of dimension $\left|h_t\right| \cdot \left|\overrightarrow{h_t}\right|$, b is the parameter matrix with dimension $\left|h_t\right|$ and h_t is the convolution result at the moment t.

In the model, a method is proposed to obtain the global semantic representation of the text, which is calculated as shown in equation (9):

$$c_i = \max_{k=0}^{T} h_{ki} \tag{9}$$

Another way is to use the attention mechanism, which was introduced to effectively account for the effect of words on the classification effect.

The depth of the network has a positive effect on the experimental results of the model, so a hierarchical recurrent convolutional network model was designed by increasing the depth of the network based on the above model. In this hierarchical recurrent convolutional network model, the two-way recurrent neural network is still used as the basis of the whole network model, and the LSTM unit is used as the base unit of the recurrent neural network.

(2) Emotional dialog generation model

Based on the encoder-decoder model, in the input layer of the model, in addition to the input word vectors at the current moment and the implicit layer state information at the previous moment, the sentiment embedding e_{post} of the input sequence is also added, and the corresponding sentiment input of the reply is e_{reply}. In the experiment, the sentiment classifier based on hierarchical recurrent convolutional neural network designed in the previous subsection is used to obtain the sentiment embedding vectors of the text sequence e, and the sentiment embedding vectors and word vectors as well as the hidden layer states are used as inputs to the recurrent neural network, whose computational expressions are shown in Eqs. (10) and (11):

$$\overrightarrow{h_t} = LSTM\left(\overrightarrow{h_{t-1}}, \left[w_t; e_{post}\right]\right) \tag{10}$$

$$\overleftarrow{h_t} = LSTM\left(\overleftarrow{h_{t-1}}, \left[w_t; e_{post}\right]\right) \tag{11}$$

where $\overrightarrow{h_{t-1}}$ denotes the forward hidden layer state at the moment $t-1$, w_t is the input word vector at the moment t, and e_{post} is the sentiment embedding vector of the input sequence, and all three are used as inputs to the LSTM in order to obtain the forward hidden layer state at the moment t $\overrightarrow{h_t}$.

In the decoder, the emotion embedding vector e_{reply} of the text is added along with the introduction of the attention mechanism, so that in the decoder, the computational expression is shown in Equation (12):

$$h_t = LSTM\left(h_{t-1}, \left[w_t; c_t; e_{reply}\right]\right) \tag{12}$$

where h_{t-1} denotes the implicit layer state at the moment $t-1$, w_t is the input word vector at the moment t, c_t is the context vector at the moment t, which is computed as in Eq. e_{reply} is the sentiment embedding vector of the replies, and the four are used as inputs for obtaining the decoder's implicit layer state h_t at the moment t.

Based on this, the corresponding responses can be generated by means of a sentiment transfer network. For example, in this experiment there are M different emotions, which can be represented as $\mathrm{E} = \left[e_1, e_2, \ldots, e_m\right]$. This sentiment matrix is randomly initialized at the beginning of training so that it is trained together during model training. It is also possible to use the pre-training parameters of the sentiment classification model above as starting values. There is a sentiment transfer matrix T in the model which is shaped like:

$$\begin{bmatrix} T_{11} & \cdots & T_{m1} \\ \vdots & \ddots & \vdots \\ T_{1m} & \cdots & T_{mm} \end{bmatrix} \tag{13}$$

where T_{ij} denotes the probability that emotion i changes to emotion j, it can be seen that for any emotion k there is $\sum_{j=1}^{m} T_{kj} = 1$. By calculating $T \odot E$, the sentiment vector after the transfer can be obtained, such as $e_1' = \sum_{k=1}^{m} T_{k1} * e_k$, and at this point, it is considered that the sentiment embedding in the input sequence has a significant influence on the generation of the sentiment of the reply has a relationship, and the computational method similar to the attention mechanism can be used to obtain the computational formulas (14), (15), and (16):

$$w_k = f\left(e_k', e_{post}\right) \tag{14}$$

$$a_k = \frac{\exp\left(w_k\right)}{\sum_m \exp\left(w_k\right)} \tag{15}$$

$$e = \sum_{k=1}^{m} a_k * e_k' \tag{16}$$

3.3 Design of robotic systems that mimic biological brain cognition

3.3.1 Emotional expression in humanoid robots

In this paper, the M6 speech recognition module is used to recognize and synthesize speech offline, so in order to realize human-computer dialogue, a speech library containing question and answer statements needs to be edited and stored into the M6 module first. The speech library is divided into 3 parts: the greeting part, the compliment/criticism part and the task part, and the question statements are labeled. The answer utterances are similarly labeled in 3 parts: the type of external event, the kind of expression, and the intensity of the emotion expressed. Among them, the emotional intensity is coded by fuzzy encoding, and the emotional intensity is classified into 5 levels: very weak, weak, average, strong and very strong, corresponding to the Arabic numerals 1-5, respectively.In this paper, a total of 45 question statements and 30 answer statements are established in the speech library, which is used for experimental validation of the artificial emotion model. On this basis, the speech library can be extended to 2000 question utterances and 3500 word answer utterances, which can include most of the daily language and meet the pragmatic demand.

The same question statement annotation may correspond to multiple different question statements, and the same answer statement annotation may correspond to multiple different answer statements. And the same answer statement may also correspond to multiple annotations. When selecting the answer statements, one statement is randomly selected as the answer statement among all the answer statements that satisfy the conditions.

From the emotion types generated by the artificial emotion model, the emotion type corresponding to the maximum emotion intensity is selected as the emotion type of the answer statement vector, and the fuzzified emotion intensity level is used as the emotion intensity of the answer statement vector. These two, together with the external event types, constitute the answer statement vector. The robot answer statement vector corresponds to the answer statement labeling, and the corresponding answer statement is selected by calculating the answer statement vector. That is:

$$A_{words} = \left[event_{class} \quad emotion_{class} \quad \hat{E}_F \right] \tag{17}$$

where $emotion_{class}$ represents the type of emotion with the highest emotional intensity among the emotions produced. $event_{class}$ represents the type of external event. $\hat{E}_F$, $\hat{E}_{\max}$ correspond to the fuzzification result, $\hat{E}_F = 1,2,3,4,5$.

Fuzzy logic is used to fuzzify the intensity of emotions generated by the artificial emotion model. Each emotional intensity is defined as a set of 5 fuzzy sets: very weak, weak, average, strong, and very strong, corresponding to Arabic numerals 1-5, whose corresponding emotional intensities are 0.1, 0.3, 0.5, 0.7, and 0.9, respectively (the range of emotional intensities is $[0,1]$). Each emotional intensity is described by these 5 fuzzy sets. The normal distribution affiliation function is determined as the affiliation function on each fuzzy set, see equations (18)-(22). In this paper, the “maximum affiliation principle” is chosen as a method to define the intensity of emotional fuzzification. That is:

$$F_1(x) = \begin{cases} 1 & x \le 0.1 \\ e^{-\left(\frac{x-0.1}{\sigma}\right)} & x > 0.1 \end{cases} \tag{18}$$

$$F_2(x) = e^{-\left(\frac{x-0.3}{\sigma}\right)} 0 \le x \le 1 \tag{19}$$

$$F_3(x) = e^{-\left(\frac{x-0.5}{\sigma}\right)} 0 \le x \le 1 \tag{20}$$

$$F_4(x) = e^{-\left(\frac{x-0.7}{\sigma}\right)} 0 \le x \le 1 \tag{21}$$

$$F_5(x) = \begin{cases} e^{-\left(\frac{x-0.9}{\sigma}\right)} & x \le 0.9 \\ 1 & 0.9 < x \le 1 \end{cases} \tag{22}$$

In Eqs. (18)-(22), $\sigma = 0.08$.

3.3.2 HCI experiment hardware system

Face-to-face verbal communication is the main and most effective way of communication between human beings, so the realization of speech communication between robots and human beings is a necessary problem to be solved in the field of human-robot emotional interaction. Due to the continuous maturity of speech recognition and synthesis technology in recent years, the speech recognition and synthesis module can provide robots with speech recognition and synthesis technology that reaches the application level.

According to the requirements of large storage capacity, scalability, high recognition accuracy, long recognition distance and other requirements put forward by the construction of humanoid avatar robot platform for speech recognition module, we finally chose the WEGASUN-M6 speech recognition module produced by Shenzhen Times Electronics Company. This module can realize non-specific human voice recognition, its recognition distance can reach 5 meters, the maximum storage capacity of 2000 phrases and its single recognition of up to 75 words of the utterance.

The module has 4 modes of speech detection: button detection mode, dialog mode, executive recognition mode, and customized password mode. In addition, this module also has the function of recognizing Mandarin and English offline, and supports 3 kinds of recognition result output modes (single-byte output, 6-byte output, 32-byte output). The maximum length of the module's speech synthesis can be up to 3500 words of text.

4 Experiments on the expression of emotions by humanoid robots

4.1 Model Analysis of Emotional Dialogue Generation

4.1.1 Data sets

The model was experimented on two publicly available conversation datasets, namely the Douban conversation dataset and the emotional first aid dataset. Douban is one of the most popular social software, and the Douban conversation dataset contains 1.1 million binary conversations (i.e., sessions between two people) with more than two rounds. The Emotional First Aid dataset is the first publicly available question-and-answer corpus in the field of counseling and includes 20,000 counseling data, which are publicly available counseling conversations. The dataset is richly annotated with detailed annotations such as discourse status

(negative or positive) and interlocutor identity information (counselor or consultant).

Six emotion categories were set up in the experiment, happy, sad, angry, disgusted, fearful, and other. “Other” indicates that the sentence does not have any emotional information. The statistical information of the dataset used in the experiment is shown in Table 2. It can be found that the rate of emotional dialog data in the Emotional First Aid dataset is higher than that of the emotional dialog data in the Douban Conversation dataset.

Table 2: Statistical information on the data set used in the experiment

Data set	Emotional category						Verification set	Test set	Total amount
	happy	sad	angry	disgust	fear	other			
Douban Conversation	96836	75661	60522	64544	13254	178075	1500	1500	488892
Emotional First Aid	1132	6536	1428	1028	4565	1421	1500	1500	16110

4.1.2 Contrasting models

(1) In order to demonstrate the superiority of the model proposed in this paper, three state-of-the-art approaches are chosen for comparison:

The HRED model introduces an additional context encoder to model the interaction structure of a multi-round dialog.

The Emo-HRED model takes into account context-level affective information to dynamically simulate human affective interactions, and generates responses with more positive affective tendencies by training on an anticipatory library containing positive emotions.

EACM, a dialog system capable of sensing emotions, is able to generate emotional responses based on a single round of dialog. The semantic coherence and syntactic accuracy of the responses generated by this model are significantly reduced due to the lack of contextual content.

ECCM model: an emotional dialog generation model based on the self-attention mechanism.

(2) Since the appropriateness of emotions cannot be evaluated automatically, the validity of the model is verified through objective and manual evaluation.

Three metrics, Perplexity, Distinct-1 and Distinct-2, are used to objectively evaluate the performance of the proposed method and the baseline model.Perplexity measures the ability of the model to describe the syntactic structure of the dialog and the syntactic structure of each sentence.The smaller the value of Perplexity, the better the performance of the language model. Distinct-1 and Distinct-2 are commonly used to measure the diversity of the generated responses; the higher the value of Distinct, the richer the vocabulary in the response.

Manual evaluation: 100 triads were randomly selected from the entire test set and two evaluators were invited to evaluate the generated responses based on two criteria: semantic coherence and emotional appropriateness. These two criteria corresponded to whether the response was syntactically fluent and logically contextualized, and whether the category and intensity of emotions in the response were contextualized, respectively. To ensure the validity of the results, these two evaluators were unaware of the model to which the response corresponded during the evaluation process. In accordance with the prevailing indicator system, the evaluation indicators were categorized into five levels, ranging from 5 to 1, representing strongly agree, agree, not sure, disagree, and strongly disagree.

4.1.3 Experimental results and analysis

Three metrics, Perplexity, Distinct-1 and Distinct-2, are utilized to judge the accuracy and

diversity of the responses generated by the recurrent convolutional network-based emotional dialog generation model and the comparison model.

The experimental results for the Douban dialog dataset and the emotional first aid dataset are shown in Fig. 2. Overall, the recurrent convolutional network-based emotional dialog generation model achieves the best performance on all metrics in both datasets, which indicates that it can generate effective emotional responses in multiple rounds of dialog.

In particular, the Recurrent Convolutional Network-based Emotional Conversation Generation model in the Douban conversation dataset improves 9.5% on the Perplexity metric over the best performing comparison model (ECCM model). This indicates that the responses generated by the recurrent convolutional network-based sentiment dialog generation model are semantically smoother and contextually clearer. The Recurrent Convolutional Network-based Emotional Conversation Generation model in both datasets improves 47.44% versus 34.87% over the ECCM model on the Distinct-1 metric, respectively.

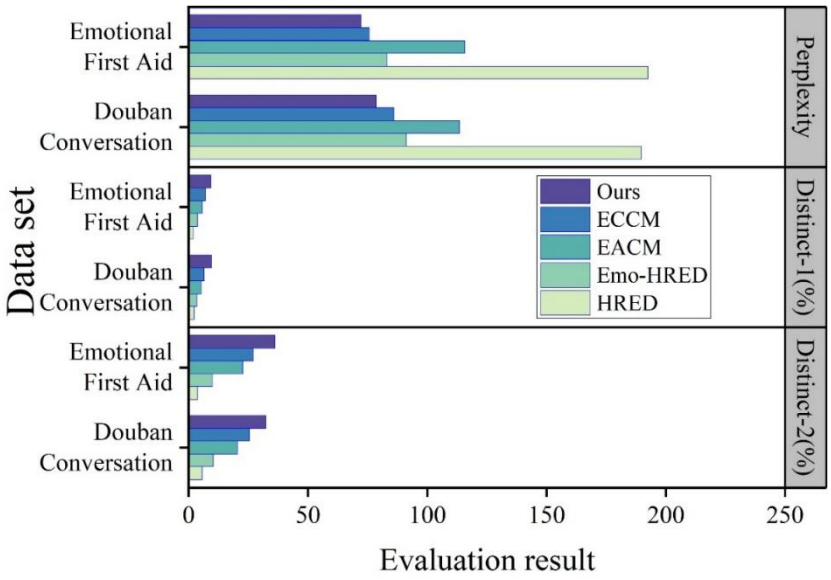

Figure 2: Experimental results on two data sets

A manual evaluation was used to focus on comparing the semantic coherence (SC) and emotional appropriateness (EA) of the responses of the recurrent convolutional network-based emotion dialog generation model and the comparison model.

The results of the manual evaluation are shown in Fig. 3, where it can be observed that EACM performs the worst in terms of semantic coherence due to the difficulty in learning contextual information.The SC score of the EACM model is 2.99.In contrast, HRED performs the worst in terms of affective appropriateness, and the EA score of the HRED model is 2.31, since this model does not take into account the affective information itself.

Among the comparison models, ECCM has the best semantic coherence and sentiment appropriateness, with scores of 3.65 and 3.87, respectively.Obviously, the Recurrent Convolutional Network-based Emotional Dialogue Generation model scores the highest in both metrics, which suggests that the Recurrent Convolutional Network-based Emotional Dialogue Generation model produces a more anthropomorphic response than that of the comparison models. This may be due to the fact that the emotion dialog generation model based on recurrent convolutional networks is able to encode the emotional and semantic information of multiple rounds of dialog separately and fuse them into the decoding process. Thus, the advantage of the cyclic convolutional network-based affective dialog generation model is the injection of emotional signals into the response generation process.

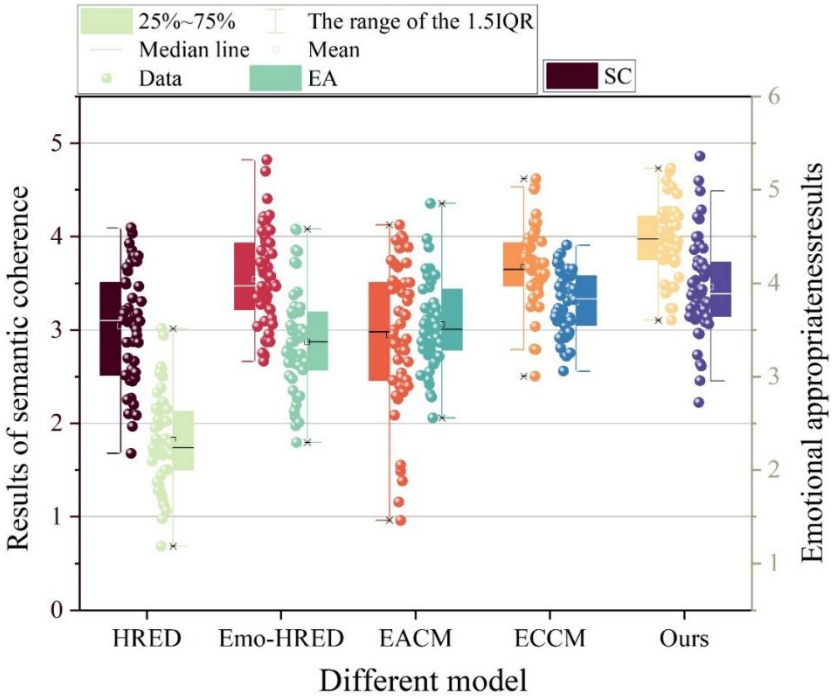

Figure 3: Results of artificial evaluation

4.2 DeepSeek Access to Emotion Recognition for Multi-Channel APPs

4.2.1 Data preparation

This chapter evaluates the emotion dialog generation model in DeepSeek based on native user reviews from the App Store. In terms of App selection, this chapter follows the following three principles: (1) The App is a popular application in the App Store marketplace that is regularly updated by the developer. (2) The selected App shall cover different categories. (3) The App should contain a large number of user reviews that can verify the effectiveness of user intent recognition. Based on the above principles the generalizability of the emotion dialogue generation model can be ensured to a certain extent. Finally, four Apps are selected in this chapter, and the native comment dataset is shown in Table 3.

Table 3: Native comment data set

App name	Categories	Comment number
Wechat	Socializing	13590
Tencent conference	Office	6251
Gode map	Navigation	4230
Baidu	Tools	8053

4.2.2 Experimental results and analysis

In this chapter, after preprocessing, intent classification, and keyword extraction of the native review data from the App Store, Gephi is used to generate a semantic network to recognize user intent at a finer granularity.

The sentiment recognition and intent classification of the sentiment dialog generation model in DeepSeek is shown in Figure 4. In the figure, Bug A, Fun A, Exp A, and Oth A represent bug reports, feature-related, user experience, and other positive comments, respectively.Bug B, Fun B, Exp B, and Oth B represent bug reports, feature-related, user experience, and other neutral

comments, respectively. Bug C, Fun C, Exp C, and Oth C represent bug reports, feature-related, user experience, and other negative comments, respectively.

The results show that user concerns are different for different apps. WeChat and AutoNavi users focus on function-related, Baidu.com focuses on user experience. While Tencent Conference concerns are in error reporting. In addition, there is a huge difference in the percentage of positive and negative comments in different categories across apps.

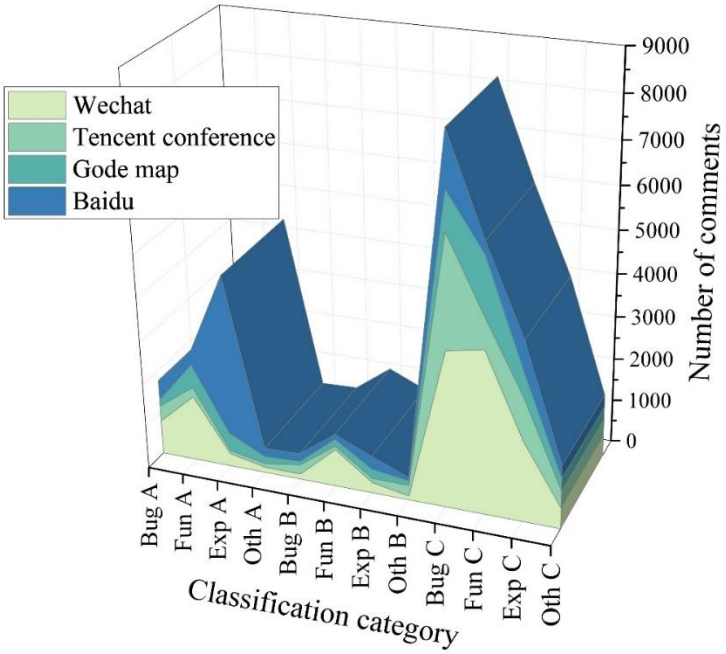

Figure 4: The emotional identification and intentional classification of the model

4.3 Humanoid Robot Emotion Expression Experiment

This chapter carries out the humanoid robot emotion expression experiment, which adopts the within-group experimental design of Human Factors, is a multilingual emotion expression experiment, and a total of 30 subjects were tested. The experiment explores the effects of different languages and different emotional expressions of the humanoid robot on the user's emotional experience during the interaction between human and humanoid robot through subjective emotional experience. We also synthesized the indicators of the three-dimensional emotion model of perceived pleasantness, activation, and dominance (PAD emotion model), and selected the humanoid robot's emotion expressions that can be accurately recognized by the user, are reliable, effective, and have strong emotional expressions, so that subsequent experiments can determine the humanoid robot's emotion expressions for cross-cultural communication.

4.3.1 Experimental design

This experiment is a humanoid robot emotion expression experiment, the subject is required to wear a wearable eye-tracker to record the whole experimental process according to the humanoid robot's multilingual emotion expression. After each round of humanoid robot multilingual emotion expression, the subjective assessment questionnaire was filled in to recognize and judge the humanoid robot's emotion.

In this experiment, personal information data (age, gender, education, interaction experience with the robot, etc.) as well as subjective perception questionnaire data (perceived

pleasantness, activation, dominance) and objective physiological measurements of the raw data of the subjects were collected. And the data were described and statistically analyzed after overall cleaning of all data (missing, outliers checking for exclusion, etc.) and preprocessing.

4.3.2 Experimental results and analysis

The descriptive statistics of the humanoid robot's Chinese emotion expressions are shown in Table 4. It can be seen that the humanoid robot can be universally recognized for the expressions of six emotions after accessing DeepSeek's emotion dialog generation model. Among them, the humanoid robot's Chinese expression about fear has a 98.12% recognition rate.

Table 4: Descriptive statistical results of Chinese emotional expression

Multilingual emotional expression		Recognition rate/%			PAD valueS			Validity
		Positive	Negativity	Neutrality	P	A	D	
Chinese	happy	95.61	1.52	2.87	2.34	1.34	1.21	5.269
	sad	0.95	95.78	3.27	-0.51	0.72	0.08	6.892
	angry	1.1	96.2	2.7	-1.89	1.56	0.45	7.015
	disgust	0.76	97.34	1.9	-0.91	0.19	1.34	6.421
	fear	0.5	98.12	1.38	-0.78	1.16	-0.15	4.139
	other	1.18	3.42	95.4	0.25	-1.73	0.65	6.886

The results of the descriptive statistics of English emotion expression are shown in Table 5. The DeepSeek model accessed by the humanoid robot is capable of multiple emotion expressions. The six emotion categories expressed by the humanoid robot have more than 90% recognition rate.

Table 5: Descriptive statistics of emotional expression in English

Multilingual emotional expression		Recognition rate/%			PAD valueS			Validity
		Positive	Negativity	Neutrality	P	A	D	
English	happy	93.56	2.36	4.08	2.56	1.36	1.31	5.631
	sad	1.52	96.12	2.36	-0.56	0.75	0.10	6.787
	angry	1.33	95.48	3.19	-1.93	1.63	0.41	7.052
	disgust	0.96	96.57	2.47	-0.81	0.21	1.35	6.331
	fear	1.59	94.21	4.2	-0.74	1.19	-0.21	4.208
	other	3.54	5.68	90.78	0.32	-1.83	0.63	6.854

The results of descriptive statistics of French emotional expressions are shown in Table 6. The recognition rate of the humanoid robot's French emotional expressions reaches more than 90% in terms of happiness, sadness, anger, disgust, and fear, but more French samples are needed for model training in terms of expressions without any emotional information.

Table 6: French emotional expression descriptive statistics

Multilingual emotional expression		Recognition rate/%			PAD valueS			Validity
		Positive	Negativity	Neutrality	P	A	D	
French	happy	90.52	2.53	6.95	2.12	1.42	1.53	5.529
	sad	3.62	92.11	4.27	-0.75	0.45	0.16	6.601
	angry	2.56	93.25	4.19	-2.01	1.00	0.38	6.987
	disgust	1.86	94.05	4.09	-0.89	0.35	1.47	6.123
	fear	1.34	93.58	5.08	-0.92	1.37	-0.35	4.562
	other	10.56	8.9	80.54	0.43	-1.55	0.76	6.604

5 Conclusion

This paper unites the core technological advantages of DeepSeek tool to design an emotion expression robot system that simulates biological brain cognition. A deep learning-based emotion dialogue generation model is proposed to analyze the possibility of DeepSeek's access to multi-App emotion recognition method in humanoid robot multicultural communication.

The emotional dialog generation model designed with Seq2Seq as the base model achieves good performance on two different datasets, which indicates that the emotional dialog generation model based on recurrent convolutional networks can produce effective emotional responses in multiple rounds of dialogs, presenting good semantic coherence and emotional appropriateness.

The emotional dialog generation model in DeepSeek tool can effectively categorize the focus of each app user, and classify the user's emotion recognition and intention when accessing multiple apps. The humanoid robot implanted with DeepSeek's emotional dialog generation model can effectively express emotions in multiple languages. Linguistic communication and emotional expression, as the most basic development of cultural communication, can promote multi-cultural emotional transmission in a multicultural context.

Funding

This work was supported by the 2024 Xiamen Institute of Technology Research Project on Education and Teaching, “Innovative Research on Ideological and Political Teaching Models of the ‘Intercultural Communication’ Course under the Belt and Road Initiative” (Project No. XJJY24016).

References

[1] Lifintsev, D., Zelihic, M., Grebliauskiene, B., Wellbrock, W., Patel, S. V., & Sharma, R. K. (2025). Young professionals' perspectives on cross-cultural communication: Assessing competence and employer support across regions. International Journal of Cross Cultural Management, 14705958251319695.

[2] Liao, Z., Pang, Q., & Xiao, H. (2025). Glocalization: Cross-cultural communication of tourism research. Tourism Management, 108, 105129.

[3] Zilola, S. (2025). PRAGMATICS IN CROSS-CULTURAL COMMUNICATION. EduVision: Journal of Innovations in Pedagogy and Educational Advancements, 1(3), 549-555.

[4] Hwang, H. C., & Matsumoto, D. (2016). Emotional expression. The Expression of emotion: Philosophical, psychological and legal perspectives, 137.

[5] Hareli, S., Kafetsios, K., & Hess, U. (2015). A cross-cultural study on emotion expression and the learning of social norms. Frontiers in psychology, 6, 1501.

[6] Matsumoto, D., & Hwang, H. S. C. (2019). Culture, emotion, and expression. Cross-cultural psychology: Contemporary themes and perspectives, 501-515.

[7] Yi, W., & Bexci, M. S. (2025). A Study on the Emotional Expression of Chinese Image

Shaped by Film and Television Culture from the Perspective of Cross-cultural Communication. Journal of Theory and Practice in Humanities and Social Sciences, 2(1), 40-45.

[8] Ruining, H. (2024). A Cross-Cultural Exploration of Cultural Influences on Emotional Expression and Regulation. Wah Academia Journal of Social Sciences, 3(01), 142-158.

[9] Khasawneh, M. A. S. (2023). The potential of AI in facilitating cross-cultural communication through translation. Journal of Namibian Studies: History Politics Culture, 37, 107-130.

[10] Peng, Y., Chen, Q., & Shih, G. (2025). DeepSeek is open-access and the next AI disrupter for radiology. Radiology Advances, 2(1), umaf009.

[11] WU, Y. (2024). Thoughts on AI innovation and open source development: Lessons from DeepSeek. Bulletin of Chinese Academy of Sciences (Chinese Version), 40(3), 446-452.

[12] Kaswan, K. S., Dhatterwal, J. S., Batra, R., & Yadav, D. K. (2023, November). ChatGPT: a comprehensive review of a large language model. In 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI) (pp. 738-743). IEEE.

[13] Liao, H. (2025). DeepSeek large-scale model: technical analysis and development prospect. Journal of Computer Science and Electrical Engineering, 7(1), 33-37.

[14] Bevara, R. V. K., Mannuru, N. R., Lund, B. D., Karedla, S. P., & Mannuru, A. (2025). Beyond ChatGPT: How DeepSeek R1 may transform academia and libraries?. Library Hi Tech News.

[15] Faray de Paiva, L., Luijten, G., Puladi, B., & Egger, J. (2025). How does DeepSeek-R1 perform on USMLE?. medRxiv, 2025-02.

[16] Kotsis, K. T. (2025). ChatGPT and DeepSeek Evaluate One Another for Science Education. EIKI Journal of Effective Teaching Methods, 3(1).

[17] B Parghi, N., Chauhan, C. R., & Karavadiya, D. A. (2025). A Comparative Study of DeepSeek and Other Ai Tools. International Journal of Innovative Science and Research Technology, 10(3), 1125-1132.

[18] Okaiyeto, S. A., Bai, J., Wang, J., Mujumdar, A. S., & Xiao, H. (2025). Success of DeepSeek and potential benefits of free access to AI for global-scale use. International Journal of Agricultural and Biological Engineering, 18(1), 304-306.

[19] Poo, M. M. (Ed.). (2025). Reflections on DeepSeek's breakthrough. National Science Review, 12(3), nwaf044.

[20] McGee, R. W. (2025). Using DeepSeek to Make Publishing Decisions. Working Paper. February 6.

[21] IS, S., & DR, A. (2025). Advancements in AI-Powered NLP Models: A Critical Analysis of ChatGPT and DeepSeek. Available at SSRN 5125445.

[22] Martens, B. (2025). How DeepSeek has changed artificial intelligence and what it means for Europe (No. node_10755). Bruegel.

[23] DivyaBabu & Terli SankaraRao. (2024). A secure routing protocol using trust-based clustering and bionic intelligence algorithm for UAV-assisted vehicular ad hoc networks. Transactions on Emerging Telecommunications Technologies,35(5).

[24] Fuyu Wang,Jiawen Fan,Mengkai Chen,Haoxuan Xie,Juma Nzige & Weining Li. (2023). Research on a bionic swarm intelligence algorithm and model construction of the integrated dispatching system for the rescue of disaster victims. International Journal of Bio-Inspired Computation,22(2),117-127.

[25] Eugenio Martelli,Laura Capoccia,Marco Di Francesco,Eduardo Cavallo,Maria Giulia Pezzulla,Giorgio Giudice... & Marco Panagrosso. (2024). Current Applications and Future Perspectives of Artificial and Biomimetic Intelligence in Vascular Surgery and Peripheral Artery Disease. Biomimetics,9(8),465-465.

[26] Jinlin Wu. (2025). The rise of DeepSeek: technology calls for the "catfish effect".. Journal of thoracic disease,17(2),1106-1108.

[27] Elizabeth MontalbanoDark Reading. (2025). DeepSeek's ByteDance data sharing raises fresh security concerns. Urgent Communications.

[28] Lei Wang,Jiajun Wang,Dawei Tong & Xiaoling Wang. (2024). A Novel Long Short-Term Memory Seq2Seq Model with Chaos-Based Optimization and Attention Mechanism for Enhanced Dam Deformation Prediction. Buildings,14(11),3675-3675.