Vol.6 Iss.1 2025 Communication & Education Review
https://doi.org/10.37420/j.cer.2025.009

Deep Learning-Based Movie Script Generation Models and Evaluation

Shiwei Mo*
School of Telecommunications Engineering, Communication University of China, China
***Corresponding author E-mail: 271499788@qq.com**

Abstract:
The main content of this article is the research on the movie script generation model based on deep learning technology and its potential value in the film industry and artificial intelligence field. First, the introduction introduces the research background, clarifies the goal of building a movie script generation model with deep learning as the core, and introduces the current research status of script generation models at home and abroad.The introduction to related technologies section provides a detailed analysis of the basic principles of deep learning and the format and characteristics of movie scripts, as well as how deep learning can provide new perspectives for film creation. In the design and implementation of the movie script generation scheme, the process of data collection and preprocessing, model selection and construction strategies, and model training and optimization methods are focused on.In order to compare the results of the models, both qualitative and quantitative analysis methods were used in the experiment. By displaying script samples generated by the model, the characteristics of the model were qualitatively analyzed. The results of the two models were quantitatively analyzed through the BERT classifier. Finally, this article also explores the potential of the model in future film creation and artificial intelligence, while also pointing out the limitations of current technology and possible development directions, and discussing social impacts and ethical considerations.

Keywords:
Deep learning, Movie script generation, GPT-4 model, BART model

1 Introduction

1.1 Research Background

Due to the film industry's pursuit of continuous content innovation and the rapid development of artificial intelligence technology in the field of text generation, movie script generation models have emerged. As film markets in various countries continue to expand globally,audience aesthetic demands are becoming increasingly diversified. Scripts, as a core element of film creation, also need to possess innovation and uniqueness to better balance marketability and artistry. Of course,the creative inspiration, life experience,and artistic accomplishment of individual screenwriters are indeed important in the scriptwriting process. However, in the face of the huge market demand and the trend of efficient and industrialized film production, the manual

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

creation method has shown limitations in terms of efficiency,output,and innovation[1].

At the same time, deep learning technology has made breakthrough progress in the field of natural language processing(NLP), providing new possibilities for breaking the traditional creation mode of movie scripts. Deep learning models, such as recurrent neural networks(RNN), long short-term memory networks(LSTM), gated recurrent units(GRU), and Transformer-based sequence-to-sequence(seq2seq) architectures, have been successfully applied to multiple NLP tasks such as text summarization, machine translation, and dialogue systems, demonstrating strong capabilities in text understanding and generation[2]. The development of these technologies has laid a solid theoretical foundation and technical support for the construction of movie script generation models.

Research on movie script generation based on deep learning has attracted widespread attention worldwide, and the integration of the two industries is of great significance. Deep learning movie script generation models have introduced convenient and innovative creation modes to the film industry. Scriptwriting is a multi-stage process, including character profiles,outlines,and main texts. However, traditional scriptwriting mainly relies on the labor of individual screenwriters, which is time-consuming and easily affected by personal factors.In contrast,models based on deep learning can quickly and massively generate character settings and initial scripts,thereby greatly shortening the scriptwriting cycle and improving production efficiency[3]. For example, the AI screenwriter developed by Alibaba Cloud can generate multiple script outlines in a short time for creators to choose and further refine, significantly improving the speed of script output. By learning a large amount of script data, the model can capture and integrate various script elements and narrative structures, providing creators with new story plots and character settings, stimulating creative inspiration, and enriching the diversity of film works.

The research and application of movie script generation models also have an important promoting effect on the development of artificial intelligence technology. First, the models need to solve advanced natural language processing tasks such as long text generation, plot coherence, and simulation of character emotions and motivations, providing practical application scenarios for frontier deep learning technologies such as reinforcement learning, variational autoencoders, and attention mechanisms, and also bringing challenges, thereby promoting in-depth research and optimization of the technology. Second, movie script generation involves cross-modal understanding and generation, such as converting visual elements like scenes and actions into text descriptions, which requires the models to have cross-modal learning capabilities, thereby promoting the theoretical and technical progress of artificial intelligence in the field of multi-modal interaction[4]. In addition, the successful application of related models helps to break the public's perception that AI is limited to data analysis and decision support, showing its great potential in the field of humanities and art.

The research and application of movie script generation models will also give rise to new business models. On the one hand, the models can be combined with the existing film production process to form an intelligent movie creation auxiliary system in script evaluation,casting suggestions, and shooting plan generation, improving the overall collaboration efficiency of the industry chain[5]. On the other hand, the application of the models may break the supply and demand pattern of the traditional screenwriting market, such as providing customized script generation services through AI platforms to meet the demand for fast and low-cost scripts for low-budget films and online short dramas[6].

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

1.2 The Phenomenon of Deep Learning Models Creating Movie Scripts

The construction of movie script models based on deep learning is an important direction in recent film intelligent production and a new combination of artificial intelligence and film creativity.

In terms of technology, research on movie script generation based on deep learning is mainly divided into two categories:one is the text generation method based on language models, which mainly uses pre-trained transformer models such as GPT-4 and BART to generate movie script texts. These models use a large amount of general text data for unsupervised pre-training, and then perform supervised fine-tuning on movie script data to learn the language rules and style characteristics of movie scripts[7][8]. These models can generate movie script texts with corresponding styles and contents according to given genres, themes, plots, and other conditions, and can also generate complete or partial movie script texts according to given beginnings or endings[8]. However, although the models can generate coherent and creative texts, it is difficult to ensure the quality of the generated texts, and there may be some grammatical and logical problems in the texts, and sometimes content repetition may occur[7]. In addition, the models cannot perceive text details as delicately as humans, such as character personalities, behavioral motivations,and emotional colors.

The other category is the structured generation method based on sequence-to-sequence, which uses the encoder-decoder framework to take the general plot, main characters, and emotional trends of the movie script as structured inputs to complete the generation of the movie script. The model can guide and constrain the generation of text dialogues through the provided information, making the dialogues more logical.The model can also generate matching scenes and dialogues according to the provided plot outline,and vice versa.In this way, it can ensure that the generated text meets the requirements and maintains a certain degree of innovation. However, the logic of the generated dialogues is not strong,and the emotional trends do not match the experience of daily life. In addition, these methods also require a large amount of structured data annotation and alignment, increasing the difficulty of data acquisition and processing. These two categories of methods have their own advantages and disadvantages,but they all have some common problems,such as the lack of logic,coherence,and originality in the generated scripts, and it is difficult to reach the level of human creators[7]. Foreign research mainly focuses on using deep learning models to generate movie scripts, such as the research team of Stanford University in the United States, which developed a movie script generation model called “ScriptGen”. Domestic research pays more attention to the application of deep learning in movie script analysis, such as the research team of Peking University,which uses deep learning models to perform emotional analysis on movie scripts.

In the commercial field, film production companies and related enterprises have also invested in the research and development of script generation tools. For example, IBM’s Watson AI has participated in the creation of part of the script for the movie “Morgan”, which generates adoptable plot paragraphs by analyzing a large amount of script data and combining the requirements of the director. Some start-up companies such as Cinelytic and ScriptBook have developed AI-based script evaluation and creation auxiliary platforms, using deep learning models to predict the market performance of scripts and even generate preliminary script outlines and specific scenes[9]. Although the application of deep learning in movie scripts is still short and shallow, it has brought considerable reform and promotion to the film industry, and the commercial prospects of the combination of the two are broad.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

However, despite many successful attempts,the application research of deep learning in the field of movie script generation still faces many challenges. The unique artistry and complexity of movie scripts require the generation models to not only have language expression capabilities but also understand the deep narrative logic, character shaping, and emotional rendering and other advanced semantic features, which puts forward high requirements for existing models. The scarcity and annotation difficulty of high-quality script data limit the effective training and performance improvement of the models[10]. How to ensure the originality of the generated scripts, avoid the risk of plagiarism, and how to balance the relationship between machine generation and human creativity are all ethical and legal issues that need to be solved urgently[11].

However, these studies, although they have studied the commerciality and feasibility of deep learning models creating movie scripts, have neglected the performance differences between models and the quality evaluation problems of the generated content.

2 Script Theory and Script Generation Technology

2.1 Format and Characteristics of Movie Scripts

Movie scripts, as the core blueprint for film production, have a format and characteristics that largely determine key elements such as the narrative logic, character shaping, and visual presentation of a film. This section introduces the basic format and characteristics of movie scripts to illustrate the standards that models need to follow when generating movie scripts.

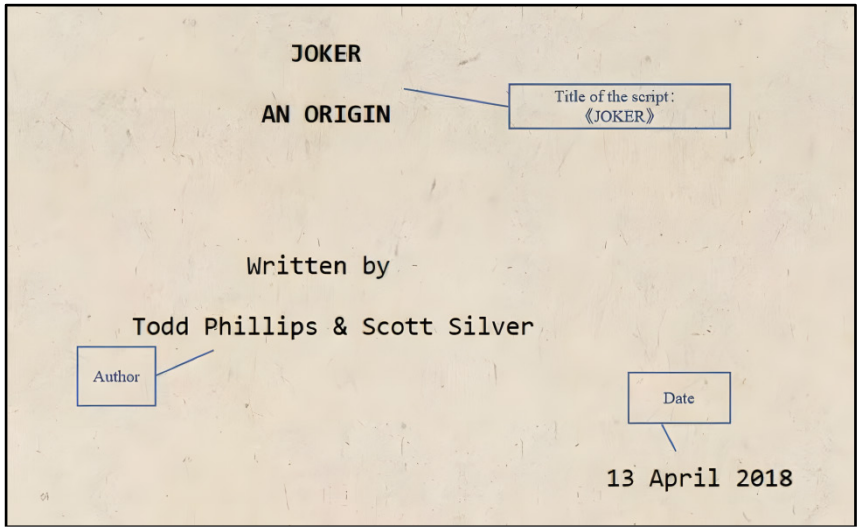


Figure.1 The first page of the script

As shown in Figure 1, taking the movie “Joker” as an example,the first page of a movie script usually includes the script title, the author's name,copyright information(such as the copyright holder, date, etc.), and contact information (such as phone number, email, etc.), which clearly and concisely identifies the script.

As shown in Figure 2, we can see the transitions and scene changes, scene titles,introductions of main characters first appearances, and action descriptions in the script. The script needs to clearly mark important

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

narrative nodes such as scene transitions and time jumps, often using standard symbols like “CUT TO:” and “FADE IN/OUT:”to guide,and can also use “TO BE CONTINUED…” or “THE END” to mark the pause points or endings of the script.Before each new scene begins, the script should provide a scene title, including the scene number and description. For example, “INT.DEPT. OF HEALTH, OFFICE-MORNING” in the figure indicates an indoor scene, located in the office of the Department of Health, in the morning. When the main character appears for the first time,their name should be written in capital letters,followed by a brief description in parentheses. The action description is written in the third person narrative style, detailing the character's actions, scene changes, and atmosphere rendering, and other non-verbal elements. The text should be vivid and intuitive to help directors, actors, and other production staff understand how the shots should be filmed and presented.For example, “Joker takes a deep breath, pauses to see if it’s over.”

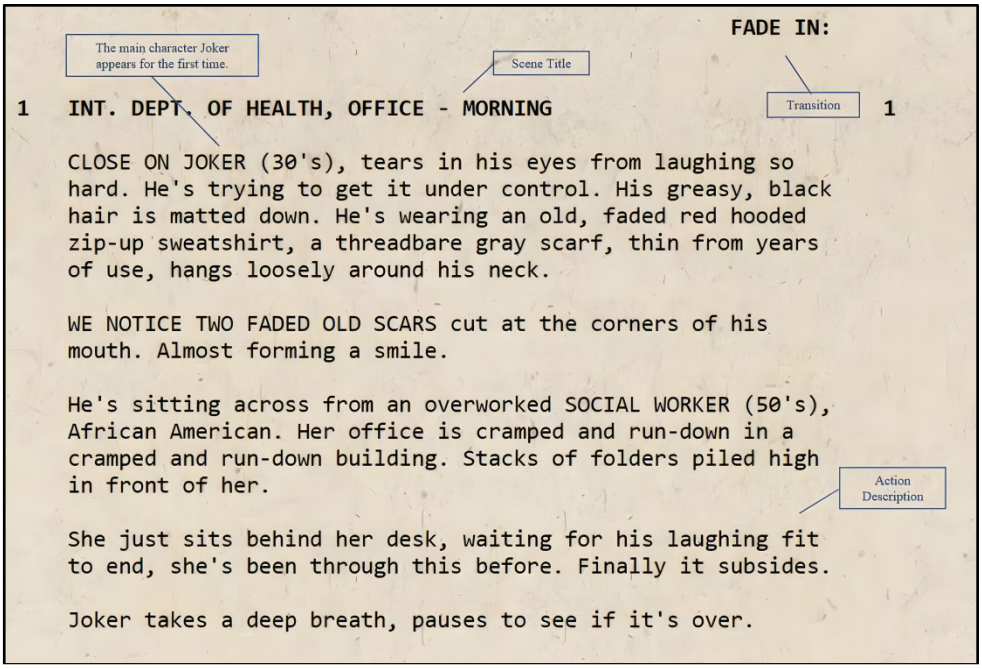


Figure 2 Script Fragment 1

As shown in Figure 3, we can see the remaining two important parts of a movie script, which are dialogue and supplementary information. Dialogue is the primary means of communication between characters in a script. The character’s name is written in capital letters in the center, with the lines starting on a new paragraph below the name.Moreover, dialogue should be concise and powerful, fitting the character’s personality and the development of the plot, avoiding excessive explanatory or redundant information. For supplementary information such as director’s notes, sound effect descriptions, dialogue descriptions, and off-screen narration, they are usually marked with parentheses and placed above the dialogue to distinguish them from the main content.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

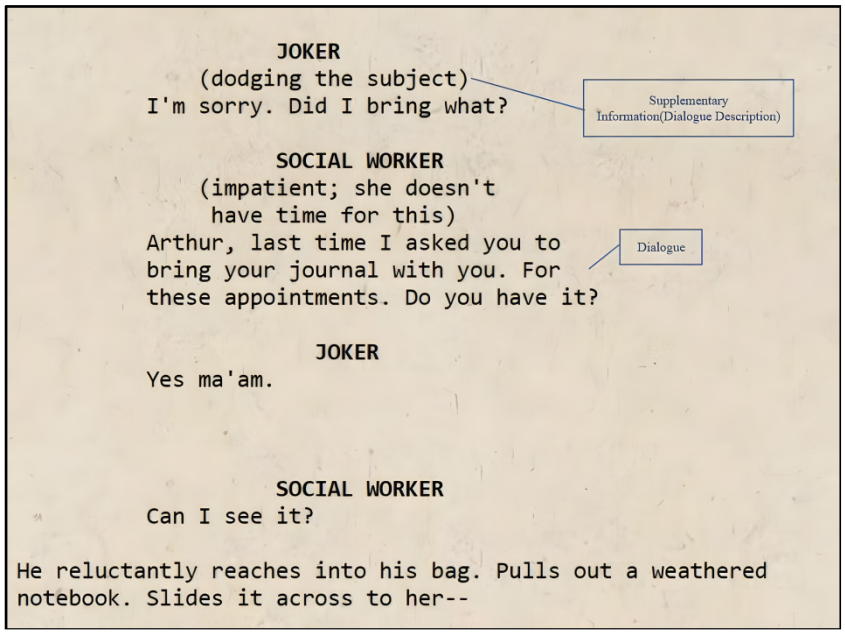


Figure 3 Script Fragment 2

Movie scripts need to tell a complete story within a limited space,requiring a compact plot, distinct conflicts,and a moderate pace, and characters are vividly portrayed through dialogue and actions. Therefore, deep learning models need to learn to identify and imitate effective narrative structures,such as the three-act structure and the golden section, and also need to understand and simulate the complex interactions and behavioral motivations between characters to ensure the dramatic tension of the generated script.

2.2 Basic Principles of Deep Learning

Deep learning, as an important branch of the field of machine learning,its basic principles and core concepts are the key to understanding and applying this technology. This section will explore in depth the core ideas, architectural features, and key components of deep learning, laying a theoretical foundation for the subsequent construction of movie script generation models based on this technology.

The essence of deep learning lies in simulating the hierarchical structure and information processing methods of the human brain's neural network to achieve efficient learning and representation of complex data. Its core ideas can be summarized as "End-to-End Learning" and "Feature Learning".

"End-to-End Learning" refers to the ability of deep learning models to start directly from raw input data(-such as text,images,etc.)[13], and through layer-by-layer nonlinear transformations, until the final prediction or decision-making results are output, without the need for manual extraction of intermediate features[12]. This approach greatly simplifies the cumbersome feature engineering process in traditional machine learning, enabling the model to automatically capture deep and abstract features of the data.

"Feature Learning" is a major advantage of deep learning over shallow learning. In deep neural networks, each layer is responsible for learning features at a specific level, gradually transitioning from primary features

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

such as edges and textures to advanced semantic concepts and other abstract features. This hierarchical feature learning mechanism enables deep learning models to effectively deal with high-dimensional, nonlinear data containing complex relationships, demonstrating strong representation and generalization capabilities.

The architecture of deep learning typically consists of multiple interconnected layers,including the input layer, hidden layers, and the output layer. The input layer receives raw data, the hidden layers perform feature learning and information transmission, and the output layer generates the model's prediction results[14]. The number of hidden layers and the number of neurons in each layer constitute the depth and width of the network, respectively, and together determine the complexity and expressive power of the model. Common types of deep learning architectures include Feedforward Neural Networks(such as multilayer perceptrons), Convolutional Neural Networks(CNNs) for processing image data, and Recurrent Neural Networks(RNNs) and their variants, Long Short-Term Memory(LSTM), which are suitable for processing sequential data such as text and speech.

Weights and biases are key parameters in deep learning models, defining the connection strengths and activation thresholds between neurons in each layer. During the training process, through the cooperation of the backpropagation algorithm and optimizers (such as gradient descent,Adam,etc.), the model adjusts these parameters from top to bottom according to the error signals calculated by the loss function, in order to minimize the difference between the model's predictions and the true labels. Regularization and Dropout techniques are widely used to prevent overfitting and ensure the model's robust performance on both the training and test sets.

In summary, the basic principles of deep learning lie in constructing multi-layer neural network architectures to achieve end-to-end learning and feature learning, using key technologies such as backpropagation, optimization algorithms, and activation functions to automatically learn complex and abstract feature representations from raw data, and to make accurate predictions or decisions based on these representations. These principles provide a solid foundation for building movie script generation models, enabling the models to learn key elements such as plot development, character shaping,and dialogue style from a large number of script samples, thereby generating high-quality movie scripts.

2.3 Applications of Deep Learning in Natural Language Processing

The application of deep learning technology in the field of Natural Language Processing(NLP) has not only significantly promoted the development of related theories and practices but also demonstrated strong processing capabilities and extensive application values in many practical application scenarios. Deep learning technology has provided a powerful driving force to the field of natural language processing, enabling it to handle text well and perform well in text-related tasks.

In terms of semantic understanding, deep learning can convert words into high-dimensional vectors through algorithms such as Word2Vec and GloVe, achieving quantitative representation of word meanings, effectively solving the limitations of traditional statistical methods in dealing with the diversity of word meanings and polysemy.These embedded vectors can capture semantic relationships between words, providing a solid foundation for subsequent tasks such as semantic analysis and emotion recognition.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

In text generation, deep learning architectures such as Recurrent Neural Networks(RNNs), Long Short-Term Memory(LSTM), Gated Recurrent Units(GRUs), and Transformers have effectively solved the nonlinearity and long-distance dependency issues of sequential data in natural language, performing well in generative tasks such as machine translation, text summarization, dialogue systems, and poetry generation, greatly improving the coherence and logic of the generated text.

In semantic parsing, deep learning models such as BERT, RoBERTa, and ALBERT, pre-trained language models, have obtained rich language knowledge through large-scale unsupervised training, significantly improving the accuracy of semantic parsing[15]. They have made breakthrough progress in fields such as reading comprehension, knowledge graph question answering, and conversational question answering systems, enabling machines to accurately understand user intentions and provide corresponding answers.

3 Design and Implementation of Movie Script Generation Scheme

3.1 Collection and Preprocessing of Dataset

The experiment requires a dataset that not only contains a large amount of text for training but also has matching style types.Based on the above requirements, the experiment selects the publicly available movie script website IMSDb as the data source. This website not only contains a large number of movie scripts but also has a rich variety of movie genres, and each movie has been genre-matched, as shown in Figures 4 and 5. The downloaded data consists of 1 147 movie scripts, each of which may have more than one genre label, and there is a significant amount of noise, so meticulous data cleaning work is required to remove irrelevant information, correct formatting errors, and ensure data quality.



Figure 4 Screenshot of the IMSDb Website

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

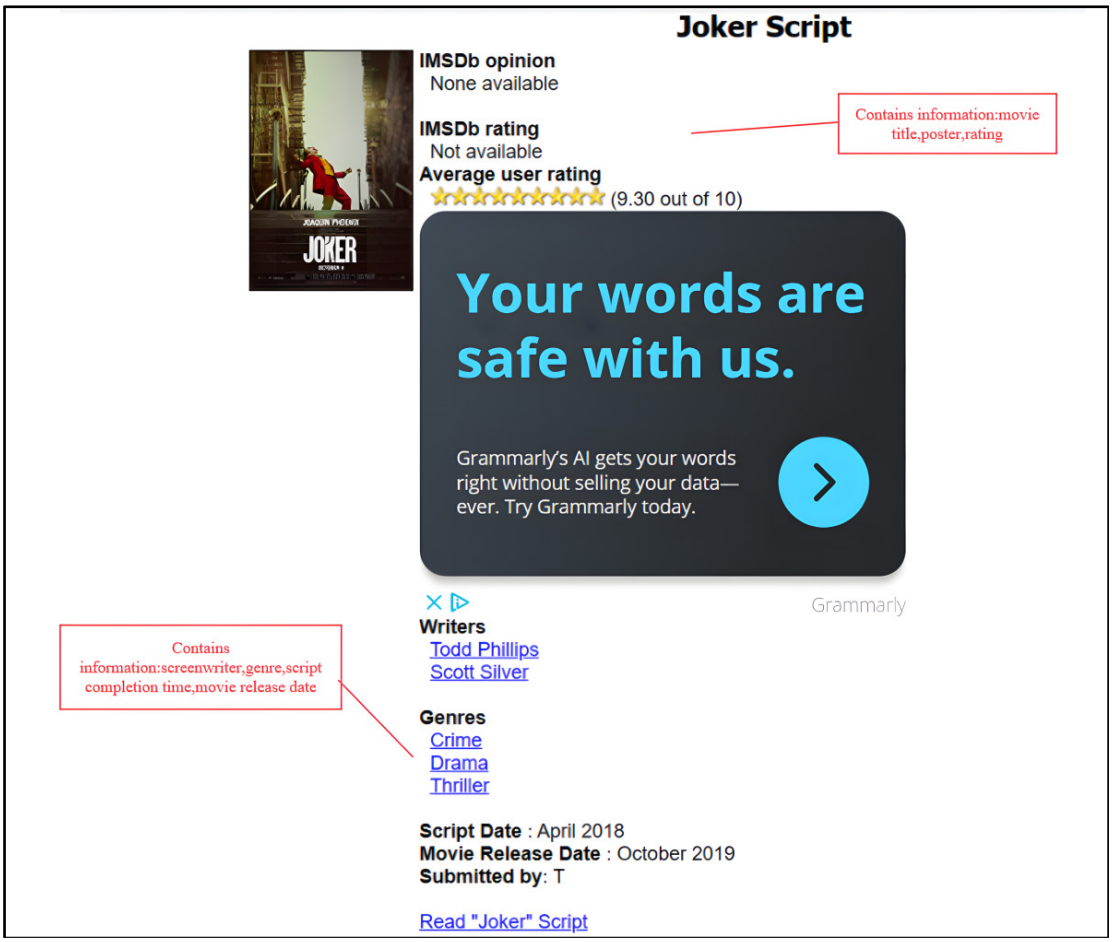
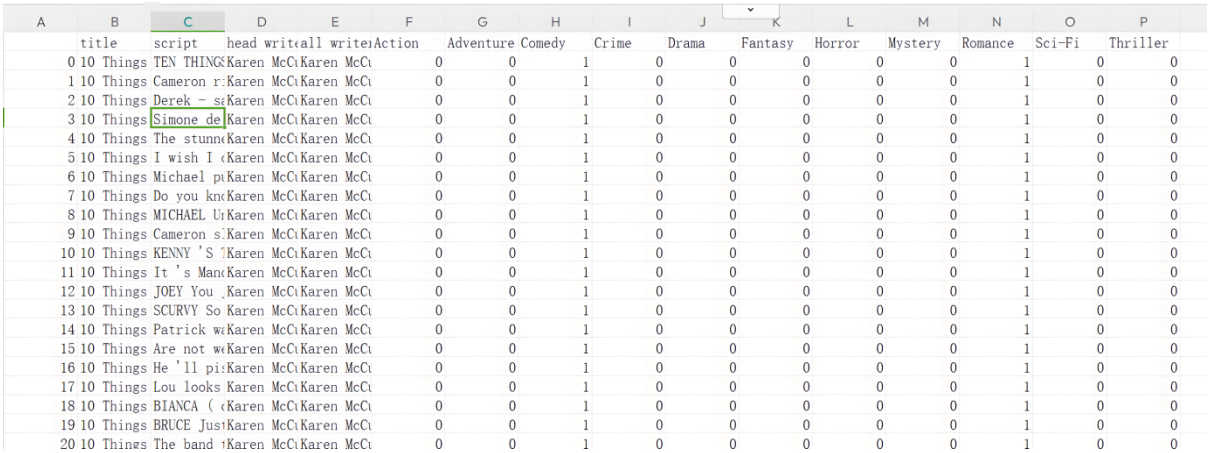


Figure 5 Screenshot of the Movie Details Page on the IMSDb Website

Firstly, HTML tags in the scripts need to be removed. The scripts may contain non-plot elements such as director's notes, scene descriptions,and shooting instructions. These non-essential elements can be identified and removed using text parsing techniques. Preprocessing utilizes parsing libraries such as regular expressions and Beautiful Soup to clear HTML tags and standardize the script format, enabling the model to accurately understand the structure of the script.

Next is text segmentation. The average size of a script is about 15 000 tokens, which is too large to be used as a single input. Therefore, the script needs to be divided into smaller sequence units as input. For example, in this experiment, it is converted into sequences with a batch length of 250. The modified data is stored in a CSV file, part of which is shown in Figure 6, including the script title, author, genre, and the text after segmentation processing.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
		title	script	head writ	all writ	Action	Adventure	Comedy	Crime	Drama	Fantasy	Horror	Mystery	Romance	Sci-Fi	Thriller
0	10	Things	TEN THING	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
1	10	Things	Cameron r	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
2	10	Things	Derek - sa	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
3	10	Things	Simone de	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
4	10	Things	The stunn	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
5	10	Things	I wish I (	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
6	10	Things	Michael p	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
7	10	Things	Do you kn	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
8	10	Things	MICHAEL U	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
9	10	Things	Cameron s	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
10	10	Things	KENNY 'S	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
11	10	Things	It 's Man	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
12	10	Things	JOEY You	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
13	10	Things	SCURVY So	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
14	10	Things	Patrick w	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
15	10	Things	Are not w	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
16	10	Things	He 'll pi	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
17	10	Things	Lou looks	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
18	10	Things	BIANCA (	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
19	10	Things	BRUCE Jus	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0
20	10	Things	The band	Karen McC	Karen McC	0	0	1	0	0	0	0	0	1	0	0

Figure 6 Screenshot of the Dataset

Then, remove genres with too few movies,selecting only those with at least 50 movies for training. After filtering, eleven genres remained,comprising a total of 1 049 movie scripts,as shown in Table 1.

Table 1 Composition of the Dataset

Genre	Number of Movies
Action	307
Adventure	186
Comedy	378
Crime	216
Drama	626
Fantasy	119
Horror	157
Mystery	110
Romance	200
Sci-Fi	169
Thriller	384

3.2 Model Selection and Optimization

The selection and optimization of models directly affect the quality of the generated scripts. This stage requires consideration of the current deep learning technologies and the particularities of movie scripts in long-text tasks,choosing models that can understand the content logic, style types of movie scripts, and generate coherent text.Based on the effective models currently available for long-text generation tasks, a viable option is the Transformer-based Sequence-to-Sequence(Seq2Seq) model, such as Google's Transformer or its improved versions like BART and the GPT series of models. These models,with their self-attention mechanisms, can effectively capture logical relationships in the text, making them particularly suitable for script-writing tasks that require strong narrative, logical, and coherence. This experiment selects the GPT-4 model for fine-tuning and chooses the BART model as the baseline model to compare the results generated by the GPT-4 model.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

3.2.1 GPT-4 Model

The GPT-4 model is a pre-trained language model developed by OpenAI. It is based on the Transformer architecture and is capable of generating coherent and diverse text. Through pre-training on a large-scale dataset, the GPT-4 model learns a rich vocabulary and context, thereby being able to understand and generate text with a certain degree of logic and coherence[16].

The GPT-4 model has shown excellent performance in the field of natural language processing and has demonstrated great potential in scriptwriting, leading to its application in the generation of movie scripts. In text generation tasks, the GPT-4 model can not only understand the meaning of words but also comprehend the logical relationships between sentences and paragraphs, thus generating coherent text or dialogue based on contextual relationships or given prompts. The GPT-4 model can also recognize text genres and generate script content of corresponding styles through different linguistic features. At the same time, the GPT-4 model serves as a vast reservoir of inspiration, continuously providing new ideas and expressions, making it a powerful and convenient technological force that can contribute to movie creation.

The input and output of the GPT-4 model can be demonstrated through the construction of simple sentences. When using the GPT-4 model for text generation, given words in the form of embeddings, the model will transform the input word embeddings into output word embeddings. This transformation generally does not change the dimensionality of the word embeddings, and the output word embeddings are referred to as hidden states.

3.2.2 BART Model

The BART model is an efficient pre-trained language model built upon the standard sequence-to-sequence (Seq2Seq) Transformer model[17]. It combines the features of a bidirectional encoder and an autoregressive decoder,undergoing pre-training by introducing noise during the encoding phase and reconstructing the original text during the decoding phase. This process enhances the BART model's ability to understand and generate text,making it highly suitable for text generation tasks, including movie script generation.

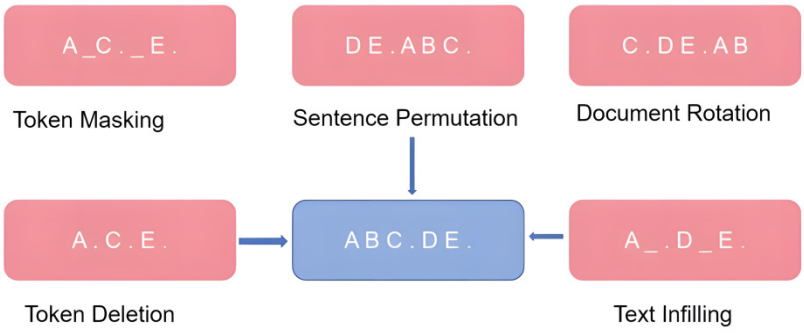


Figure7 Transformation of Input Noise by the BART Model

During pre-training, the BART model employs various noise strategies to corrupt the original text, such as Token Masking, Token Deletion, Text Infilling, Sentence Permutation, and Document Rotation, among others. The purpose is to enable the model to learn how to reconstruct text from incomplete or disordered inputs, thereby improving its understanding of text structure and content.The transformation of input noise by the

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

BART model is illustrated in Figure 7. In the model architecture, the encoder is responsible for extracting key features from input script segments and forming contextual representations. This can be achieved by stacking multiple layers of Transformer Encoders,utilizing self-attention mechanisms to capture long-range dependencies. The decoder then generates new script content step by step based on the contextual information provided by the encoder. To prevent information leakage during the decoding process, mechanisms such as Masked Multi-Head Attention or causal autoregression can be employed.

The BART model, equipped with additional language modeling and pre-trained on large text corpora, performs well in text generation tasks and has been used to achieve state-of-the-art results on text summarization datasets[17]. In the task of movie script generation, the BART model can also be fine-tuned and trained on the IMSDb dataset to learn specific styles of scriptwriting.

In this experiment, a controllable language generation task is added to the model's summarization problem, training the BART model on a fine-tuned dataset. The dataset adds a genre tag before the input sequence with a maximum length of 1024, for the model to generate text of a specified type.

The BART model accepts a piece of text and a genre tag as input,randomly replacing a portion of the words in the text with a mask symbol, such as "[MASK]". Then, the encoder of the BART model reads this masked text, learning information within the text, including the semantics of words,relationships between words,and grammatical structures. Subsequently, the decoder of the BART model attempts to predict the words that have been replaced by the mask. Since the decoder is autoregressive, it generates words one by one,considering all previously generated words and the genre tag each time a new word is produced. Finally, the BART model generates a new piece of text, which is a rewrite of the input text.While retaining most of the information from the input text, it modifies the original text and supplements relevant elements according to the given genre.

In the task of movie script generation,an initial prompt is typically given to the BART model, which then generates the remainder of the script based on this prompt. The BART model takes into account factors such as the style of the script, plot logic, and relationships between characters to produce a script that meets the requirements.

3.3.3 BERT Model

The BERT model is a deep learning model built upon the Transformer model, designed for natural language processing tasks. Its main feature is bidirectionality. Traditional language models, such as the GPT model, can only learn language from left to right or right to left, but the BERT model can simultaneously consider the context on both the left and right sides of a word. This bidirectional training approach enables the BERT model to understand the meaning of each word in context, thereby enhancing the model's ability to handle language tasks[18].

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

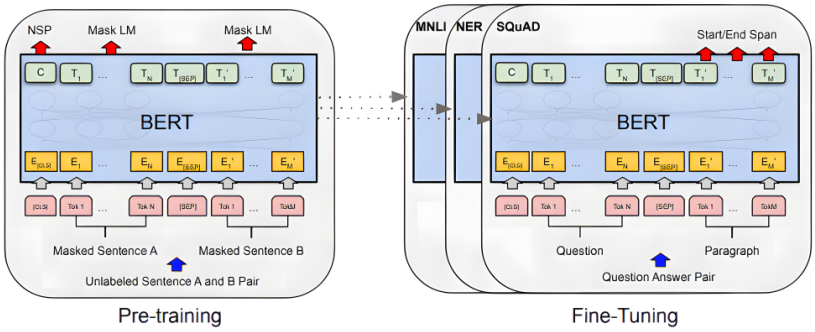


Figure 8 Pre-training and Fine-tuning of the BERT Model

As shown in Figure 8, the training process of the BERT model includes two stages: Pre-training and Fine-Tuning. Except for the output layer, the same structure is used in both pre-training and fine-tuning models, and the same pre-trained model parameters are used to initialize models for different downstream tasks. The [CLS] token is a special symbol added before each input instance,and the [SEP] token is a special separator token[19].

During the pre-training stage, the BERT model is trained on a large amount of unsupervised text using a self-supervised training method, encoding linguistic knowledge such as grammar, word meaning, and semantics into the Transformer-encoder layer in the form of parameters. Pre-training allows the model to learn general knowledge about text, rather than just a specific NLP task[20]. In pre-training, the BERT model randomly masks some words in the input text and then predicts these words based on the context. A sentence is input into the model multiple times, but the masked words are not the same each time. This is because if a word is completely masked and never appears, the model may not be able to handle these unseen words during the fine-tuning stage, which could affect performance.

In the fine-tuning stage, the BERT model first initializes the model with pre-trained parameters and then fine-tunes it using labeled data specific to the task. All parameters are updated during this period, and each different downstream task has its own fine-tuned model[21]. The self-attention mechanism in the Transformer allows BERT to model many downstream tasks by appropriately exchanging inputs and outputs, including QNLI(for determining whether text contains answers to questions), MRPC(also for judging whether two sentences are equivalent), SST-2(sentiment analysis of movie reviews), CoLA(semantic judgment of sentences,whether they are acceptable) and so on.

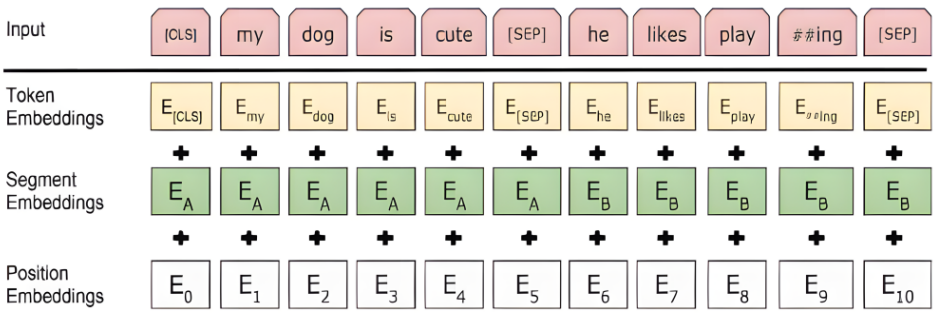


Figure 9 Input of the BERT Model

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

As shown in Figure 9, the BERT model includes two special tokens:[CLS]for classification tasks and [SEP] for sentence separation, in addition to three types of embeddings. The Token Embedding inserts[CLS] at the beginning of the input text after text segmentation and [SEP] at the end, then performs a token embedding lookup, which involves finding the embedding vectors of the vocabulary,with dimensions:

$$[\text{seq}_{\text{length}}, \text{embedding}_{\text{dims}}]$$

where the former represents the sequence length and the latter represents the dimension of the embedding vector. The Segment Embedding has only two values,1 and 0, used to distinguish the first sentence from the second sentence. For example, all tokens in the first sentence have a Segment Embedding value of 0, and all in the second sentence have a value of 1, with dimensions:

$$[\text{seq}_{\text{length}}, \text{embedding}_{\text{dims}}]$$

The Position Embedding inputs positional information into the model, where the positional information in the BERT model is calculated using sine and cosine functions,with dimensions:

$$[\text{seq}_{\text{length}}, \text{embedding}_{\text{dims}}]$$

and these parameters are not updated during training. Therefore, the input of the BERT model consists of token embeddings, segment embeddings, and position embeddings.

4 Evaluation of Movie Script Generation Scheme

4.1 Experimental Design and Setup

The experiment chose to use the GPT-4 pre-trained model from the Hugging Face library as the base model and selected the BART model as the baseline model. Both have demonstrated excellent performance in natural language processing tasks, meeting the requirements of the experiment.

Firstly, Byte Pair Encoding(BPE) was used to tokenize the scripts. BPE is a natural language processing technique that can break down text into smaller fragments, effectively handling rare and new words in the scripts.

Then, the scripts were segmented. The average size of a script is about 15 000 tokens, which is too large to be used as a single input. Therefore, the scripts needed to be divided into smaller sequence units as input. After tokenizing the scripts, they were converted into sequences with a batch length of 250. However, each sequence had 20 overlapping tokens. Keeping these overlapping sequences might lead to redundancy, but removing them would make it difficult for the model to maintain coherence over the full length of the script. Therefore,they were retained.

After data processing,model optimization was also required. The experiment used the AdamW optimizer with a linear decay schedule to fine-tune the model for 5 epochs. AdamW is an improved version of the Adam optimizer, showing excellent performance in training deep learning models. The experiment chose a learning rate of 0.003 to train the model, and the entire process took approximately 30 hours on a Google Colab V100GB GPU.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Finally, the model was used to generate script fragments on the Google Colab cloud platform. The experiment required a Python version 3.6 environment, but the default Python version on the cloud platform was 3.10, so it was necessary to install conda and then set up a Python 3.6 environment within it.

4.2 Presentation and Analysis of Experimental Results

4.2.1 Qualitative Analysis

Tables 2 and 3 show examples of scripts generated by the GPT-4 and BART models.

Table 2 Script Fragments Generated by the GPT-4 Model

Script	Horror	Comedy	Romance
The man went to the park.	, and I got off. Reilly walks past Holly's body. HOLLY The man? REILLY Yeah, he's your boyfriend. HOLLY He's not my boyfriend either. REILLY Oh, I see. (THEN) Kit'll never forgive you for that, buddy. Holly waves him off. HOLLY Don't apologize. REILLY Yeah, okay. He touches her breasts. HOLLY She's gorgeous. REILLY I don't think I've ever gotten confused about it. HOLLY Don't worry, you're going to love this. REILLY I miss it. HOLLY I never did.	PAUL Turn around. FATHER Stop it. (then into radio) Jesus, I'm sorry. I never imagined anybody being born. Your sister is coming. PAUL Not even you. FATHER Oh, you didn't know who she was and I wanted to see her. I thought she was just something wonderful. Father mimes the singing coming from the darkness. FATHER You see what that means? It means that our little girl is being agitated by her parents. PAUL What the hell? Can't we all just be dumb enough to think that? Father mimes the singing coming from the darkness. FATHER It means she's come. She's lonely. She's been over the last few days. No one has ever noticed her. MOTHER Your sister worries. FATHER Oh, Jesus.	Ray waves dismissively. RAY Then I should have gone. Lila SNAPS some papers off her uniform and trims them off. RAY That's just thin air. LILA I left my clothes. RAY Still doin' what? LILA Breathe. RAY Still doin' it. Lila walks toward the ramp. EXT. ACCLAIM IN FRONT OF SANTA MONICA BLVD. - DAY The ESCALADE EXPLODES. INT. ACCLAIM ON ROUTE Lila peers through the window and sees a car pull up. She gets out of the car and walks up to the front door. LILA You need a ride. RAY I'm not gonna leave my car. LILA Let's go. LILA You're not gonna leave me

Each script fragment was generated using the same text prompt with different genre requirements.The experiment used 20 text prompts and 7 genre tags to generate 140 script fragments for each model. Due to space limitations,only a portion is displayed in the table. After analyzing the script fragments generated by the mod-

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

els, some trends of different models can be observed.

The fine-tuned GPT-4 model used script blocks with a sequence length of 300 for script generation, resulting in more coherent scripts. Moreover,even though the genre indications in the text were not obvious,scripts generated with specified genre tags still showed clear differences. For example, Table 2 shows script fragments generated with the genre style tags of Horror, Comedy and Romance. The Horror genre script created an atmosphere of danger and tension, the Comedy genre script depicted the protagonist's amusing family life,and the Romance genre script began with a dangerous explosion plot, leading to the romantic theme of the protagonist's unwavering commitment in a crisis. However, there is a noticeable difference between the experimental results and actual scripts, making it difficult to qualitatively evaluate the results of GPT-4.

Table 3 Script Fragments Generated by the BART Model

Script	Horror	Comedy	Romance
The man went to the park.	The man went to the park. The man then went to a different park. He then went back to the first park. Then he went to another park. And so on and so on until he got to the third park. The man was found dead in a park.	The man went to the park. The man then went to a different park. He then went back to the first park. And so it went on for a few more days. He even went to another park in the middle of the day. And then he went back again.	The man went to the park. The man then went to a different park. He then went back to the one he had been at. He had been there for a few hours. He went to another park. It was the first time he had ever been there. He was so happy.
She puts her finger to her lips, and tells him to hide.	She puts her finger to her lips, and tells him to hide. She then looks at him in horror and says, “I'm going to kill you. He looks at her in horror, and she says, “'I'm gonna kill you.” She then puts her hand to her mouth, and says she wants to kill him.	She puts her finger to her lips, and tells him to hide. I'm sorry,” she says. “It's just that I don't want him to see me like that I just want to be with you,” she adds. “That's all I want”	She puts her finger to her lips, and tells him to hide. I'm not going to let you get close to me,” she says to him. You're going to have to hide from me,” he replies. “Yes,” she replies, “I am going to hide

As shown in Table 3, the scripts generated by the BART model indeed perform better in specific genres,namely romance and horror.When the input text prompt is “She puts her finger to her lips,and tells him to hide” and the option to generate a romantic script is selected, it outputs a plot fragment where the woman protects the man,warning him to hide. If the option to generate a horror script is chosen, it outputs a plot fragment where the woman kills the man. At the same time, some limitations of this method can also be observed. For example, in the sequence generation task of the BART model, after some operations, it copies from the input sequence[23]. Therefore, it is difficult for this model to generate novel text given a prompt. The initial weights of the model often lead to the model repeatedly restating the input sequence in the generated text. Moreover, the format of the generated text is quite different from that of a movie script.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

4.2.2 Quantitative Analysis

The experiment used a multi-label classifier implemented with a fine-tuned pre-trained BERT classifier to evaluate the quality of the generated movie scripts. The classifier was implemented using the Fast-BERT library from the Hugging Face library. Fast-BERT is an efficient implementation of BERT that can train high-quality BERT models in a shorter amount of time. The classifier used the LAMB optimizer to accelerate training. The LAMB optimizer is designed specifically for large-scale training and can effectively reduce training time while maintaining model performance[22].

The experiment utilized the BERT model to implement text classification, quantitatively comparing the performance of the two models by predicting the genre of the text generated by the GPT-4 and BART models and comparing it with the actual genre. The accuracy of the fine-tuned BERT model was 81.6%, indicating that 81.6% of all generated texts could be accurately classified. However, the F1 score was 0.54. The F1 score is the harmonic mean of precision and recall. Precision refers to the proportion of samples predicted as positive by the model that are truly positive, and recall refers to the proportion of all true positives that are predicted as positive by the model. The range of the F1 score is from 0 to 1, with higher scores indicating better performance.The classifier's overall accuracy was 81.6%, but the F1 score was only 0.54, suggesting that the model's performance in predicting target genres was not ideal, which may imply that the generated scripts do not have a clear correspondence with the specified genres in terms of style.

Table 4 Model Accuracy Tested with BERT Classifier

	Top-1	**Top-3**
GPT-4	0.13571	0.37857
BART	0.02857	0.22857

As shown in Table 4, when using the model to predict the genre of the generated scripts,the classifier's most likely predicted target genre matches the actual genre, with a Top-1 probability of about 14%, and the probability that the actual genre is among the three most likely predicted genres is about 38%. For the BART model, these probabilities are approximately 3% and 23% respectively. The results are better than chance.

The dataset for the experiment was split into an 80:20 training-test ratio,with various model configurations trained for 3 to 5 epochs, and hyperparameters such as regularization terms, learning rate decay, batch size, learning rate schedulers, and optimizers were fine-tuned. However, regardless of the parameter settings, the model always overfitted. The reason may be that stop words were not removed during script processing,thereby adding unnecessary noise,leading to overfitting of the training data. Additionally,as mentioned in the qualitative section,movie scripts are long and must be processed in blocks. The experiment on the Google Colab cloud platform tried using blocks of 300 and 512, as well as sequences of various lengths such as 8,16 and 32. The longer the input sequence used,the better the effect,but due to memory limitations,long input sequences could not be processed, and the large amount of noise they generated could not be handled.

Furthermore, the experiment also used the non-linear dimensionality reduction technique t-SNE to evaluate the style of the generated text.t-SNE technology can reduce high-dimensional data to two or three dimensions for visualization. When similar points in the high-dimensional space maintain proximity,mapping data points

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

to a low-dimensional space can preserve the similarity between points in the high-dimensional space. Thus, after dimensionality reduction with t-SNE, visualizable vocabularies can be generated in two or three-dimensional spaces[24]. In the visualization,words of different genres form different clusters, and using different colors to distinguish them helps intuitively perceive the differences in vocabulary style between genres.

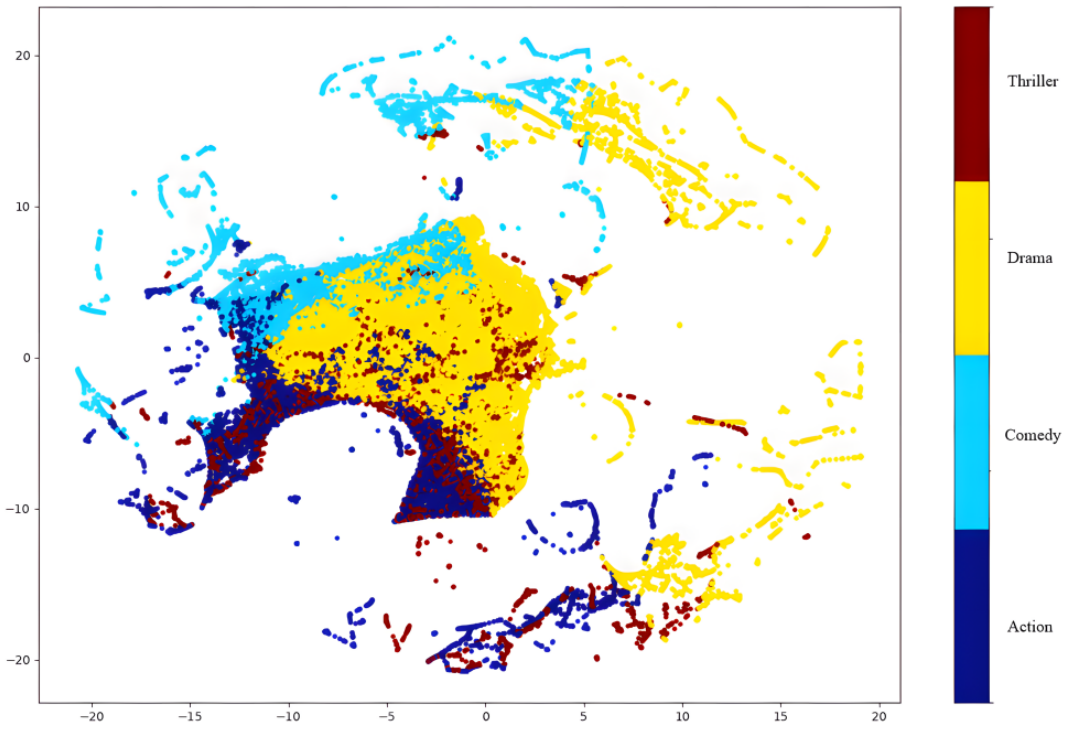


Figure 10 t-SNE Visualization of Action, Comedy,Drama,and Thriller Script Genres

As shown in Figure 10, it displays the distribution of vocabulary in script generation models for four genres:thriller, drama, comedy, and action. The style of vocabulary is assessed by comparing the position, density, and distribution of word clusters across different genres. If the word clusters of two genres are far apart in the visualization, this may indicate a significant difference in their vocabulary usage.

5 Applications and Prospects

5.1 Application of Movie Script Generation Models

This section will demonstrate script fragments generated using the high-performing GPT-4 model and modify them into usable short film script fragments.

The chosen script genre is horror, with the general plot being: A girl returns home and discovers a shadowy figure holding a knife. She believes it to be the murderer who has been eluding capture in the community,but it turns out to be her sister.In the end, she finds out that it was her sister's spirit that killed the murderer and protected her. The female protagonist is named “Monica” and the text prompt is set as “When Monica came home, she saw the shadow with a knife. She thought it was a murderer who fled from the community, but it

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

turned out to be her sister. Finally, she found that it was her sister's soul that killed the murderer and protected her". The generated script and the modified script are compared as follows.

Table 5 Script Fragments Generated by the GPT-4 Model and the Revised Fragments

Script Fragments Generated by the GPT-4 Model	Script Fragments Written by Humans
INT. MONICA'S LIVING ROOM - NIGHT The room is shrouded in darkness. MONICA, 30s, enters, her breath visible in the cold air. She glances around, uneasy. The clock ticks ominously. MONICA (whispering) Is anyone there? A SHADOW looms in the corner, holding a knife. Monica's heart races. She thinks of the recent community murders. MONICA (to herself) The murderer... here? She inches closer, fear gripping her. But as the shadow moves, it reveals itself to be JESSICA, Monica's younger sister, translucent and ethereal. JESSICA (softly) Monica... MONICA (stammering) J-Jessica? But you... you died! JESSICA I did. But my soul remains. I've been protecting you, watching over you. That man—the real murderer—he was about to strike. I intervened.	INT. MONICA'S LIVING ROOM - NIGHT A dark figure stands by the sink,holding a knife.The face is indistinguishable,and the knife glints in the light. Monica is scared against the wall,shivering,her breath clearly visible in the cold air.She looks around nervously as the clock ticks. The shadow steps closer,slightly illuminated by the hallway light,revealing it to be Jessica. Jessica: (concerned,holding a knife) It's me,Monica. Monica: (relieved,scolding as she turns on the light) Sis!Why didn't you turn on the light?You scared me to death! Jessica: (turning to wash the knife) I was cleaning up the house,didn't realize it got so late. Monica: Why didn't you tell me you were coming? Monica walks to the sofa,puts down her backpack and phone,then heads to the sink. Jessica approaches Monica's phone and switches it to airplane mode. Jessica: Monica,show me the night view of Beijing. Monica: Sure,let me use the bathroom first. Jessica: No!I...I'm so hungry,let's go out quickly. Jessica takes Monica's hand and pulls her out the door. EXT.STREET-NIGHT Jessica: The night view of Beijing is so beautiful. Monica: Yeah,my sister should come to keep me company more often. Monica takes out her phone from her trouser pocket.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

	Monica: (talking to herself) How did it accidentally switch to airplane mode? Monica turns off the airplane mode,and the phone rings.Monica walks slowly forward while listening,and Jessica stays in place. Relative: Monica!You finally answered the phone. Monica: I accidentally turned off my phone,sorry. Relative: Oh,poor child,you have to hold on,your sister... Monica: What happened?Did my sister come to Beijing today... Relative: (interrupting) Jessica passed away today from sudden cardiac arrest. Monica: (suddenly turns her head) How could that be,my sister she was just... At the spot where Jessica was,there is only a coat left on the ground. EXT.Apartment Building Entrance-NIGHT Monica returns to the apartment building entrance in a daze,and a neighbor approaches. Neighbor: Monica,thank goodness you're okay.A burglar just climbed out of your room,the one who's been breaking in with a knife recently! Monica: (confused) What?Burglar? (epiphany) My sister...she saved me. (crying in sorrow) Sister... INT.Corridor Outside Monica's Home-NIGHT The burglar,holding a knife,collapses on the ground,his eyes filled with terror and his body rigid.

The GPT-4 model tells stories in a simple and understandable manner.Since the GPT-4 model generates long movie fragments, it handles the plot more directly,starting with terrifying scenes and then revealing the protection of loved ones through the characters' actions, allowing the audience to better receive the information. However, the revised script is positioned as a short film script with a duration of five minutes, and through modifications, each scene is imbued with a suspenseful and horror atmosphere, with the element of

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

familial affection only revealed at the end, leading the audience to an epiphany alongside the protagonist, Monica.

Moreover, it is possible that because the datasets used to train the GPT-4 model are English movie scripts,the vocabulary and plot handling in the generated scripts tend to be more Western-oriented. Eastern movies,on the other hand, are usually more subtle, using details to create atmosphere and portraying familial affection in a more delicate manner. The horror and familial elements in the generated script are direct and obvious, but upon closer examination, they seem somewhat absurd.For example, why does the soul suddenly appear? How does the soul kill the murderer? Therefore, through modifications, I have the character of the sister appear as a soul to her younger sister on the night of her death, providing a motive for the character's actions and also strengthening the power of familial bonds. Secondly, I have the sister's spirit injure the murderer hiding in the bathroom, then show his state of death from extreme fright, adding an element of the supernatural and sparing the protagonist from suspicion.

In summary, movie script generation models based on deep learning have a very broad development prospect in the fields of film creation and artificial intelligence. As demonstrated above,incorporating script fragments generated by the GPT-4 model in the creative process can save time in constructing a script outline from scratch, thereby allowing for better refinement of the script within a limited time frame.

5.2 Reflections and Prospects

Movie script generation models, as an emerging artificial intelligence technology applied to film creation, are closely linked to the digital transformation of the film industry and have a broad application prospect in the future.

From the perspective of film creation,in the traditional scriptwriting process, screenwriters need to spend a lot of time and energy on plotting, character shaping, and dialogue polishing. However, with the help of deep learning models,by simply inputting key information such as specific themes, styles, and character settings, the model can quickly generate a preliminary script. This not only alleviates the workload of screenwriters but also, through the model's learning of a vast amount of script data, it can excavate and integrate unique elements from different scripts, achieving cross-boundary collisions and fusions of creativity, and giving birth to novel and unique script concepts.

From a business perspective, movie script generation models are expected to promote personalized customization and precise marketing in the film industry. Nowadays, with the diversification of audience aesthetics, the market demands more novel and personalized scripts. By combining deep learning technology in the early stage, it is possible to accurately grasp and precisely predict audience preferences and market trends, thereby better launching scripts that meet the aesthetic needs of the audience[25]. In the later stage, combined with AI prediction algorithms, the market acceptance and box office potential of movie scripts can be assessed, thus more precisely customizing marketing strategies and reducing unnecessary investments.

However, although script generation models have shown certain innovative potential and practical application value, they are currently facing new challenges.

The first is technological limitations. On the one hand,the performance of deep learning models is highly

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

dependent on the quality and scale of training data. For movie script generation, a large number of high-quality scripts are needed as training sets. However, due to restrictions such as copyright protection, data acquisition difficulties,and the non-standardized format of scripts, the available script data is relatively scarce and unevenly distributed, which may lead to the model learning limited script patterns and making it difficult to generate rich and diverse plot content. On the other hand, training complex movie script generation tasks with deep learning models requires a large amount of computing resources, and the computational cost is too high.

In addition, although the emergence of movie script generation models has brought revolutionary tools and more possibilities to the film industry,the social impact and ethical issues generated by related technologies cannot be ignored. A study by IBM shows that with the help of AI assistance, the completion time of the first draft of the script can be shortened by about 30%, reducing the workload of screenwriters and also helping to accelerate the progress of film projects, better adapting to the fast-paced market environment. However, this improvement in efficiency may also lead to a reduced demand for human screenwriters in the industry, triggering adjustments in employment structure and potentially impacting the professional group of screenwriters[26].

Moreover, as an art form that integrates visual and auditory language, the core of film lies in telling touching stories and providing engaging viewing experiences. Although deep learning technology can imitate human narrative logic to a certain extent, there is still a long way to go to truly capture the subtle changes in human emotions, appreciate the rich connotations of culture, and understand the complexity of social phenomena. Over-reliance on AI may cause film works to drift away from real life, thereby losing humanistic depth and generating a series of “mechanized” scripts that only have completion but lack artistic quality, which are neither sincere nor likely to be commercially successful.

We must actively seize the opportunities brought by technological innovation to promote the innovative development of the film industry, while also deeply exploring the usage norms of artificial intelligence, safeguarding the unique humanistic power of film, and ensuring that the future development of the film industry is both vibrant and warm-hearted.

References

[1].J. Zhu, H. Yang, H. He, W. Wang, Z. Tuo. (2023). MovieFactory: Automatic Movie Creation from Text using Large Generative Models for Language and Images . arXiv:2306.07257.

[2].Liu,C.,Shmilovici,A.,&Last,M. (2020). Towards story-based classification of movie scenes. PLOS ONE, 15(2).

[3].C. Lei, Y. J. Hong. (2023). Innovative Applications of AIGC Technology in Film Digital Creation and Production Platforms . Journal of Beijing Film Academy.(in Chinese)

[4].H. Dongmei, L. Tong. (2023). AIGC: The Driving Force of Change in Film and Television Content Production . Science and Technology Communication. (in Chinese)

[5].C. Jun, Z. Jianjun, L. Menghe. (2023). Research and Prospects on AI and Intelligent Film Production. Modern Film Technology, Issue 10.(in Chinese)

[6].Afshar Aref, Gholamreza Nouri, Shahin Ghazvineh, Seyed Hossein Hosseini Lavassani. (2024). Machine-Learning Applications in Structural Response Prediction: A Review. Practice Periodical on Structural

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Design and Construction, 29(3).
[7].J. Zhu, H. Yang, H. He, W. Wang, Z. Tuo, W.-H. Cheng, L. Gao, J. Song, and J. Fu. (2023). MovieFactory: Automatic Movie Creation from Text using Large Generative Models for Language and Images . arXiv,2306.07257 [v1].
[8].P. W. Mirowski, K. W. Mathewson, J. Pittman, and R. Evans. (2023). Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals . Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), Hamburg, Germany. ACM, New York, NY, USA, 34.
[9].M. Ghosh, S. S. Roy, H. Mukherjee, S. M. Obaidullah, et al. (2021). Movie title extraction and script separation using shallow convolution neural network .
[10].P. Gandhi, P. Bhattacharyya. (2020). Survey: Automatic Movie Plot and Script Generation. arXiv,2008(00012) .
[11].F. Zhiwei. (2024). Term Problem in the Development from ChatGPT to Sora . China Terminology, 26(2), 88-92. (in Chinese)
[12].W. Jingjing. (2020). Prediction of Drug Target Interactions Based on Multimodal Autoencoders . PhD. .Dissertation. Taiyuan University of Technology.(in Chinese)
[13].W. Junjie, L. Guannan, W. Jingyuan, Z. Yuan, B. Hui, L. Hao. (2020). Data Intelligence: Trends and Challenges . Systems Engineering - Theory & Practice, 40(12), 3121-3134. (in Chinese)
[14].Z. Hong, G. Bibu. (2023). Exploring Large AI Models: Current Status, Challenges, and Future . Telecommunications Express, Issue 7, 6-11.(in Chinese)
[15].Y.Shuo,W.Siyu,W.Chao,C.Zhenyu,Z.Yue.(2023).Application Prospects of Large Language Models Similar to ChatGPT in Power Systems.Electrical Age,Issue 10,50-53.(in Chinese)
[16].C.Jingyi.(2023). Application of ChatGPT in Digital Sports Teaching.Snow and Ice Sports Innovation Research,Issue 9,133-136.(in Chinese)
[17].M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension . Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7871–7880.
[18].J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding . Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186.
[19].J. Liu, Y. Gao. (2022). Survey of Word Embedding Representation Learning. Control Theory and Technology, 39(7), 1171-1193. (in Chinese)
[20].C.Fangzhou.(2020). Research on Machine Reading Comprehension Based on Deep Learning. PhD. Dissertation.Beihang University. (in Chinese)
[21].Z. Zhengkun. (2020). Research on Generative Text Summarization Model Based on Transformer. PhD. Dissertation.Huazhong University of Science and Technology. (in Chinese)
[22].Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, and C.-J. Hsieh. (2020). Large Batch Optimization for Deep Learning: Training BERT in 76 Minutes . ICLR 2020.
[23].M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer.

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

(2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension . Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7871–7880.

[24].H. Xinyue. (2017). Semantic and Sentiment Analysis in Twitter Event Detection. PhD.Dissertation. University of Electronic Science and Technology of China. (in Chinese)

[25].J. Zhao. (2023). Research on Event Reasoning Methods for Application Reviews Based on Data Augmentation. PhD. Dissertation. Journal of Dalian Maritime University.(in Chinese)

[26].Q. Zhao. (2023). The Impact of Intelligent Text-to-Image Generation Technology on the Creative Process of Digital Imaging Art. Journal of Jilin College of the Arts, Issue 5, 406-414.(in Chinese)

[27].L. Lanting. (2023). Research on Intent Recognition Methods for Human-Computer Interaction. Master. Dissertation.Journal of Northeast Normal University. (in Chinese)

[28].Y. Zhu, R. Song, Z. Dou, J.-Y. Nie, and J. Zhou. (2020). ScriptWriter: Narrative-Guided Script Generation. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, July 2020, 8647–8657.

[29].J. Wang, X. Liu, X. Ouyang, G. Zhang, and Y. Zhang. (2024). A method of applying deep learning based optical flow algorithm to river flow discharge measurement. Measurement Science and Technology, 35(6).

[30].S. Ouyang, B. He, H. Luo, and F. Jia. (2024). SwinD-Net: a lightweight segmentation network for laparoscopic liver segmentation. Computer Assisted Surgery, 29(1).

© 2025 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)