

The limitations of large language models in short drama subtitle translation from the perspective of Skopos theory

Zhengxuan Lu, Jingjing Feng*

Jinling Institute of Technology, Nanjing, China

*Corresponding Author. Email: fengjjng@jit.edu.cn

Abstract. Driven by China's cultural export strategy, vertical short dramas have surged in popularity; however, subtitle translation speed and quality have emerged as critical bottlenecks hindering their global reach. Although large language models are widely used for preliminary subtitle translation, they remain inadequate in handling culture-loaded expressions, conveying emotional nuances, and restructuring syntactic patterns. Drawing on the three rules of Skopos theory, this study conducts a comparative analysis of Chinese-to-English subtitle translations generated by DeepSeek-V4-Pro, Doubao 2.1 Turbo, and DeepSeek-R1, using two widely viewed short dramas as case studies. The findings reveal that the core deficiency is not grammatical inaccuracy but a lack of Skopos-oriented awareness: while the models can transfer semantic content, they fail to calibrate their outputs according to dramatic context and character traits, resulting in a persistent gap between semantic fidelity and pragmatic effectiveness. Each model exhibits distinct shortcomings, and none adequately reconciles fidelity with colloquial naturalness.

Keywords: Skopos theory, short dramas, large language models, subtitle translation

1. Introduction

Vertical short dramas are currently the category with the fastest capacity expansion in China's digital cultural export sector. The explosive momentum of this content format—characterized by fast pacing, intense conflicts, and short episode lengths—is evident from a set of data: according to estimates by DataEye Research Institute [1], the overseas micro-short drama market reached approximately US\$4 billion in 2025 when revenues from In-App Advertising (IAA), subscriptions, and web-based platforms are included. Short dramas have become an important vehicle for China's digital cultural exports, and subtitle translation, as the core textual channel for overseas audiences to understand the plot, directly affects the international communication effect of Chinese stories. The domestic performance of two dramas on the Hongguo platform is equally striking: *Mrs. Su's High-Profile Divorce* surpassed 2 billion views within 35 days of its launch, and *Summer Findela* has accumulated over 3 billion views.

However, the rapid growth of production capacity has been accompanied by a conflict between speed and quality from the very beginning. The average cycle from project initiation to launch for a single drama in the industry is about one week, while traditional manual subtitle translation for a micro-short drama takes at least three days [2]. To break through this bottleneck, end-to-end AI dubbing and translation systems covering

subtitle extraction, translation, dubbing, and synthesis have been adopted in the industry. The short drama *The One in Front of Me Is the One in My Heart* generated multilingual versions using such systems, and its overseas views quickly exceeded 10 million.

The focus of this paper is precisely the unresolved quality issues behind this efficiency improvement. The quality of subtitle translation directly affects overseas audiences' understanding of the plot and their experience of the so-called "sense of gratification", thereby restricting the communication effect of short dramas overseas. To systematically examine the above issues, this study adopts the perspective of Skopos theory, selects two short dramas with contrasting styles from the Hongguo platform, uses three mainstream domestic large language models to generate English subtitles respectively, and conducts a horizontal comparison from three dimensions: the conversion of culture-loaded expressions, the conveyance of emotional nuances, and subtitle segmentation and syntactic restructuring, so as to investigate the common problems that may arise when different models process such texts.

2. Literature review and theoretical framework

2.1. Overview of subtitle translation research

Gottlieb [3] defines subtitle translation as a form of "diagonal translation", a trans-semiotic conversion process that transforms spoken source language into written target language and presents it synchronously on screen. This format itself determines that subtitle translation is inevitably subject to strict constraints of screen space and display time: a single subtitle usually consists of no more than two lines, stays on screen for only 2 to 7 seconds, and must match the characters' speech rate and lip movements [4]. These temporal and spatial constraints require the translation to be accessible and fluent within an extremely short reading window, making audiences barely aware of the subtitles. This ideal state is summarized as "subtitle invisibility", which remains an important benchmark for measuring the quality of subtitle translation to this day [5].

Regarding the integration of Skopos theory and subtitle translation, Yao [4] analyzed the subtitle translation strategies of *Kung Fu Panda* from the perspective of Skopos theory, while Liu [6] proposed translation methods such as linguistic condensation, domestication, and foreignization from the same theoretical perspective. All the above studies take human translation as their research object, focusing on translators' strategic choices. In terms of human-machine collaborative subtitle translation, Sun [7] examined the application of generative artificial intelligence in subtitle translation using the TV drama *Bian Shui Wang Shi* as a case study, and found that the human-machine collaboration model has significant advantages in improving translation efficiency and quality. However, most of the above studies use TV dramas or films as cases; vertical short dramas, due to their unique stylistic features and production rhythm, have not yet been included in the scope of discussion. Furthermore, as large language models have begun to participate in the subtitle translation process, there is still insufficient systematic discussion on how machine translation behavior embodies or deviates from the principles of Skopos theory, and what unique quality defects may arise.

2.2. Skopos theory and its three core rules

Founded by Hans Vermeer in the 1970s, Skopos theory lies at the core of functionalist translation theory. As Zhang [8] points out, Skopos theory, which "takes the purpose of the text as the primary criterion of the translation process", differs from the traditional equivalence view. It does not pursue formal equivalence between the target text and the source text, but emphasizes that the translation should select the optimal processing method according to its intended function. Skopos theory comprises three core rules in a

hierarchical progressive relationship: the Skopos rule—translation strategies are determined by the communicative purpose, serving as the supreme criterion; the coherence rule—the target text must be smooth and natural, conforming to the expressive conventions of the target language; the fidelity rule—the target text must maintain intertextual coherence with the source text, but its form and degree are determined by the purpose and subordinate to the Skopos rule [9].

2.3. Applicability of Skopos theory to subtitle translation

The practical characteristics of subtitle translation are highly consistent with the three rules of Skopos theory. Yin [10] points out that the unique communicative nature of Skopos theory makes it highly appropriate for film subtitle translation. Translators can make precise adjustments to the translation based on the background and needs of the target audience. In short drama subtitle translation, the Skopos rule requires the translation to help target audiences understand the plot and receive emotional impact in an extremely short time—whether a translation is good or bad should be judged by whether it accomplishes this task, rather than by literal equivalence. Temporal and spatial constraints force translators to compress information and simplify sentence structures, providing a practical impetus for the coherence rule: that is, the translation must be smooth and natural, and be understood by audiences as if it were an original text in the target language. Since images and sound already supplement a great deal of information, translators may appropriately omit or adjust content, granting greater flexibility to the fidelity rule. Core plots, character relationships, and emotions must be accurately reproduced, while specific wording can be flexibly adjusted. Guo [11] also notes that the main purpose of subtitle translation is to most effectively convey relevant information to audiences in a specific cultural context under the constraints of time and space.

However, when applying Skopos theory to AI translation research, there is an additional particularity to consider: unlike traditional human translation, even when provided with prompts (text instructions input by users to constrain the direction of translation), AI can hardly develop a genuine "sense of purpose". It cannot actively weigh the cross-cultural communicative functions of the translation as humans do. Therefore, whether and to what extent its output can embody the Skopos rule is itself a question to be examined.

2.4. Definition of analytical dimensions

Based on the above discussion, subtitle translation faces three core difficulties within the framework of Skopos theory. First, the Skopos rule requires the translation to enable target audiences to instantly understand the plot and experience the "sense of gratification". However, the large number of culture-loaded expressions in short dramas lack direct equivalents in English, posing a major difficulty [12]—this is the core issue of the first dimension. Second, the conveyance of emotional nuances is also dominated by the Skopos rule. Emotions such as provocation, sarcasm, and retort in Chinese lines must deliver their due impact to target audiences, while the coherence rule plays a supporting role here, ensuring that emotions are conveyed on the basis of fluent translation—this is the core issue of the second dimension. Third, syntactic restructuring mainly involves the coherence rule, while also taking into account the Skopos rule. Chinese is primarily paratactic, with logical relations implicit in word order; English is primarily hypotactic, requiring logical relations to be explicitized through conjunctions. Within the extremely short display time of vertical screen subtitles, how to restructure Chinese run-on sentences into short sentences conforming to English grammar and rhythm while ensuring smooth comprehension by audiences is the core issue of the third dimension. Although the fidelity rule is not the main object of investigation in the aforementioned three difficulties, it assumes a "bottom-line" function in the analytical framework of this study—that is, it examines whether there is obvious information loss or distortion in the translation, serving as a basic guarantee for quality assessment.

Accordingly, this study conducts translation analysis from the following three dimensions:

Dimension 1: Conversion of culture-loaded expressions. This mainly corresponds to the Skopos rule—whether the translation enables target audiences to instantly grasp the intention of the lines and experience the "sense of gratification" through appropriate domestication or free translation.

Dimension 2: Conveyance of emotional nuances. This mainly corresponds to the Skopos rule—whether the translation accurately reproduces the emotional intensity of the original lines, such as provocation, sarcasm, or retort, and whether its colloquial expression delivers similar emotional impact to target language audiences.

Dimension 3: Subtitle segmentation and syntactic restructuring. This mainly corresponds to the coherence rule—under the constraint of extremely short subtitle display time, how the translation maintains semantic integrity and fluency through syntactic adjustment and logical conjunctions while avoiding the loss of core information.

3. Research design

As can be seen from the above literature review, there is still a lack of systematic research on the application of AI in vertical short drama subtitle translation. The following elaborates the design of this study from three aspects: case selection, AI model configuration, and research methods and analytical procedures.

3.1. Case selection and corpus sources

This study selects two representative hit short dramas from the Hongguo platform as cases:

Case 1 is *Mrs. Su's High-Profile Divorce* (launched in June 2026), an urban female revenge-themed drama with outwardly emotional lines and dense gratification-driven narrative tropes. Case 2 is *Summer Findela* (launched in September 2025), a sweet romance drama with more restrained emotional expression, rich subtext, and complex emotional layers. The two dramas present a sharp contrast in line style, which is conducive to examining the performance of AI models in dealing with different types of translation difficulties.

All original lines are sourced from the subtitles of episodes publicly broadcast on the platform.

3.2. AI translation models

This paper selects three mainstream domestic large language models available in 2026 to generate English subtitles: DeepSeek-V4-Pro (DeepSeek official website, deep thinking mode); Doubao 2.1 Turbo (ByteDance Doubao App, "Expert Mode"); DeepSeek-R1 (accessed via Tencent Yuanbao App, deep thinking mode). All three models are product forms directly accessible to ordinary users, representing the current practical application level of domestic AI translation tools.

Both sets of prompts include background information such as drama genre, plot synopsis, target audience, emotional tone, and character settings. However, due to the vastly different styles of the two short dramas, the prompts are set differently in terms of emotional tone and translation priorities, as outlined below:

The core requirements of the prompts for *Mrs. Su's High-Profile Divorce* are: (1) emotional conveyance takes priority over literal correspondence, and the "thrill of revenge" must be retained; (2) colloquialism; (3) preset domestication strategies for culture-loaded expressions (e.g., "净身出户" translated as "walk away with nothing"); (4) the female protagonist's Voice-Over (VO) monologue maintains a calm, sober, and angry tone; (5) accuracy is required for commercial/legal terminology (e.g., "股权转让" translated as "equity transfer").

The prompts for *Summer Findela* emphasize: (1) emotion priority over literal correspondence, with priority given to conveying "sweetness", "gratification", and "angst"; (2) colloquialism—using natural

expressions such as gonna/wanna; (3) free translation/domestication of culture-loaded expressions (e.g., "老登" translated as "fuddy duddy"); (4) the female protagonist's VO inner monologue maintains literary grace and elegance.

The essential difference between the two sets of prompts is that they stipulate in advance the communicative purpose that the translation should serve, based on the emotional tone and target audience of each drama. In essence, they simulate the Skopos rule at the prompt level.

3.3. Research methods and analytical procedures

This study adopts the methods of case analysis and translation comparison, carried out in three specific steps. Taking the three rules of Skopos theory as the unified evaluation criterion, it conducts item-by-item comparison of the translation outputs of the three models. The specific operational procedure is as follows: First, the three AI models are made to generate complete English subtitles for the two dramas respectively under uniform prompt conditions. Each model performs two independent generations, and the version with better translation quality is selected for subsequent analysis. Second, typical difficult sentences with analytical value are selected from Chinese lines along three dimensions: density of culture-loaded expressions, intensity of emotional expression, and syntactic complexity. Four sentences are selected from each of the three dimensions for in-depth case analysis. The selection criteria are: for culture-loaded expression category, sentences with the highest density of culture-loaded expressions that serve as the semantic core of the whole sentence are selected; for emotional nuance category, sentences whose emotional conveyance highly depends on Chinese modal particles and sentence rhythm are selected; for syntactic restructuring category, sentences whose logical relations rely entirely on word order and semantic progression rather than conjunction cohesion are selected. Finally, the translation quality of the 12 example sentences is evaluated sentence by sentence in combination with the three rules of Skopos theory, focusing on the handling of culture-loaded expressions, the effect of emotional conveyance, and the quality of segmentation and logical restructuring, and summarizes the common defects and differentiated performance of the three models.

4. Analysis of limitations of large language models in English subtitle translation of short dramas

This chapter analyzes the typical problems of the three models in English subtitle translation of short dramas from three dimensions: culture-loaded expressions, emotional nuance conveyance, and syntactic restructuring. Four most representative translation examples are selected for each dimension, and specific analysis is carried out around the corresponding principles of each dimension. Overall, DeepSeek-V4-Pro tends toward complete literal translation, Doubao 2.1 Turbo tends toward simplified compression, and DeepSeek-R1 tends toward extreme simplification but lacks stability.

4.1. Limitations in the dimension of culture-loaded expressions

Should "任劳任怨" be translated as "uncomplaining" or "long-suffering"? Is "出这口气", "satisfaction" or "revenge"? The translation of culture-loaded expressions is never a matter of finding equivalent words, but of reconstructing equivalent cultural functions in the target language. Existing studies analyzing the translation of culture-loaded expressions in film subtitles under the framework of Skopos theory have found that the coherence rule requires the translation to be naturally readable in the target language, while the fidelity rule requires the translation to accurately convey the intent of the original text—yet both must take the Skopos rule as their premise [13]. In terms of overall trends, DeepSeek-V4-Pro mainly adopts explanatory literal

translation, retaining relatively complete information but often diluting cultural weight with neutral vocabulary; Doubao 2.1 Turbo and DeepSeek-R1 tend toward simplified compression, with the most obvious loss of cultural information. Four typical translation examples are selected below for specific analysis:

(1) 我做了十年任劳任怨的贤妻,他却在外养了三年的情人.

(Episode 2 of *Mrs. Su's High-Profile Divorce*)

DeepSeek-V4-Pro: I spent ten years as a devoted, uncomplaining wife... while he spent three years keeping a mistress.

Doubao 2.1 Turbo: Ten years I've been the devoted wife, and he's kept a mistress for three.

DeepSeek-R1: Ten years devoted wife, he kept a mistress three years.

In the plot, the female protagonist gives up her career and endures ten years of mistreatment from her husband's family, only to be rewarded with her husband's infidelity. She uses the phrase "a dutiful wife who bears hardship without complaint" to take stock of this marriage.

DeepSeek-V4-Pro uses "uncomplaining" to render "任劳任怨", making it the only model that addresses this culture-loaded expression directly, with the best performance in the fidelity rule. However, "uncomplaining" merely describes a good temper, whereas the protagonist's forbearance in the plot is a passive sacrifice shaped by long-term marital discipline. "Long-suffering" would better convey the complex image of the female protagonist who endures unfair treatment silently over a long period of time. Therefore, there is still a gap in emotional conveyance in terms of the Skopos rule. Doubao 2.1 Turbo compiles the phrase into "devoted wife", filtering out information about specific efforts, resulting in the greatest loss in fidelity, but its grammar is fluent with the best coherence. DeepSeek-R1 similarly compresses the culture-loaded expression, and the phrase "Ten years devoted wife" is grammatically fragmented, failing to meet the coherence rule.

The three models have distinct shortcomings: DeepSeek-V4-Pro has complete information but insufficient emotional weight; Doubao 2.1 Turbo is fluent but loses cultural information; DeepSeek-R1 has unstable structure. None of the three achieves a balance among fidelity, Skopos, and coherence.

(2) 就算我拿不到一分钱,我也要出这口气.

(Episode 46 of *Mrs. Su's High-Profile Divorce*)

DeepSeek-V4-Pro: Even if I don't get a single cent... I want to have this satisfaction.

Doubao 2.1 Turbo: Even if I get nothing, I'm getting my revenge.

DeepSeek-R1: Even if I don't get a single cent out of this, I'm still getting my revenge.

In the plot, the female protagonist says this in response to threats during negotiations with her ex-husband. It juxtaposes money with dignity and serves as a declaration of revenge.

DeepSeek-V4-Pro's "have this satisfaction" has the weakest intensity. "Satisfaction" sounds too polite in a face-to-face declaration scenario, unable to bear the intensity of confrontation, thus losing points in the Skopos rule. Doubao 2.1 Turbo and DeepSeek-R1 use "getting my revenge" accurately: in response to bribery, they reply with revenge in a direct tone, satisfying the Skopos rule with complete concessive structures. This case contrasts with that of "a dutiful wife who bears hardship without complaint": in the latter, cultural information needs to be retained, and the compression by Doubao 2.1 Turbo and DeepSeek-R1 leads to point loss; here, "vent this anger" means "revenge" in the declaration scenario, so literal translation is instead accurate. DeepSeek-V4-Pro tends toward safe paraphrasing in this case, rather than adjusting intensity according to the confrontational strength of the dialogue.

(3) 盲婚哑嫁,你也能接受?

(Episode 1 of *Summer Findela*)

DeepSeek-V4-Pro: A blind marriage, and you're fine with that?

Doubao 2.1 Turbo: You're okay with a blind arranged marriage?

DeepSeek-R1: A blind arranged marriage, and you're cool with it?

After learning about the arranged marriage arranged by her family, the female protagonist questions the male protagonist, whom she has never met. "盲婚哑嫁" refers to a marriage arranged by parents in which the parties involved have no right to choose.

DeepSeek-V4-Pro translates it literally as "blind marriage", retaining the imagery of "blind" with acceptable fidelity, but it is easily misinterpreted as a marriage between blind people, carrying the risk of ambiguity. Doubao 2.1 Turbo's "blind arranged marriage" supplements "arranged" to highlight the meaning of arrangement, balancing fidelity and Skopos, making it the most accurate. DeepSeek-R1 adopts the same core treatment as Doubao 2.1 Turbo, with "cool" in the ending making it more colloquial. A translation more consistent with the Skopos rule would be "You're okay with an arranged marriage to a complete stranger?" The phrase "to a complete stranger" supplements the core dilemma implied by "blind" and "mute"—marrying a stranger—balancing fidelity and Skopos, and making up for the connotation of "forced silence in marriage" omitted by all three models. The character "哑" implies that the female protagonist cannot even voice her opposition in this marriage arrangement; the absence of this connotation would lead audiences to mistakenly think she is merely arranged to marry, rather than being silenced by force.

Overall, Doubao 2.1 Turbo and DeepSeek-R1 solve the problem of cultural information transmission by supplementing "arranged", while the literal translation of DeepSeek-V4-Pro instead creates comprehension barriers. The common limitation is that none of the three translations separately conveys the connotation of passive silence and lack of voice implied by "哑嫁".

(4) 我听说,那些老男人一旦开了窍,老房子着火,简直让人受不了.

(Episode 3 of *Summer Findela*)

DeepSeek-V4-Pro: I've heard once those old guys get a clue, it's like an old house catching fire—utterly irresistible.

Doubao 2.1 Turbo: I hear once those old guys catch feelings, it's intense—like a house on fire. Unstoppable.

DeepSeek-R1: I heard once those old dogs catch fire, it's unreal.

This line comes from a teasing phone call between the female protagonist and her best friend. "老房子着火" is a metaphor for middle-aged people becoming passionately and uncontrollably emotional once they fall in love.

DeepSeek-V4-Pro adopts literal translation to preserve the form, but "utterly irresistible" means "completely impossible to resist", carrying a strong connotation of attraction. It loses the playfulness of teasing and the vivid imagery of loss of control, thus losing points in the Skopos rule. Doubao 2.1 Turbo breaks it down into "catch feelings" plus "like a house on fire", with accurate meaning and satisfactory Skopos rule performance, but omits the word "old", diluting the sense of contrast. DeepSeek-R1 uses slang substitution with ingenuity, but "it's unreal" is vague in meaning, leaving gaps in both fidelity and Skopos. A translation more consistent with the Skopos rule would be: "I hear that when these older guys finally get a clue, they go up like an old house on fire—it's a whole lot to handle". The phrase "go up like an old house on fire" fully retains the metaphorical imagery; "older guys" placed at the beginning highlights the meaning of "old"; and the colloquial exclamation at the end fits the pragmatic scenario of teasing between best friends, more fully balancing imagery retention and pragmatic effect.

The three models take different approaches: DeepSeek-V4-Pro preserves form through literal translation but deviates in emotion; Doubao 2.1 Turbo has accurate meaning but impairs the imagery; DeepSeek-R1 is creative but vague in expression. The common limitation is that they can only perform semantic conversion of culture-loaded expressions, struggling to balance both imagery integrity and pragmatic effect.

4.2. Limitations in the dimension of emotional nuance conveyance

Semantic fidelity does not equal accurate emotional delivery. The emotions in Chinese lines are often embedded in sentence rhythm, word order pauses, and modal particles, yet model translation tends to package complete semantics into grammatically standard sentence structures, neutralizing emotions into plain statements. In terms of overall trends, DeepSeek-V4-Pro has the most prominent problems in this dimension—high fidelity but flat emotional delivery; Doubao 2.1 Turbo occasionally achieves better rhythm by simplifying sentence structures; DeepSeek-R1's excessive compression often leads to emotional fragmentation. Four typical translation examples are selected below for specific analysis:

(1) 原来这就是暗格藏着的秘密.

(Episode 2 of *Mrs. Su's High-Profile Divorce*)

DeepSeek-V4-Pro: So this is the secret hidden in the compartment.

Doubao 2.1 Turbo: So this is the secret in that hidden compartment.

DeepSeek-R1: So this is the compartment's secret.

The female protagonist utters this exclamation after seeing with her own eyes the evidence of her ex-husband's infidelity hidden in the car. The tone marks a shift from being deceived to sudden realization.

DeepSeek-V4-Pro has complete semantics and the best fidelity, but its sentence structure is somewhat formal, and the sense of discovery is reduced to a statement, resulting in flat conveyance of the sudden realization implied by "原来". Doubao 2.1 Turbo has a slightly better colloquial feel, but similarly wraps emotion in prepositional phrases. DeepSeek-R1 is the most concise, but the sense of dynamism is staticized by the possessive case, leaving gaps in Skopos performance for all. A translation more consistent with both the Skopos and coherence rules would be "So this is what he's been hiding". The present perfect continuous tense indicates the continuity of the action "he has been hiding", which matches the three-year time span of deception in the plot. It focuses on the other party's act of concealment, consistent with the protagonist's character of immediately settling scores after discovering evidence, and is also closer to colloquial habits.

The common limitation of the three models is that they neutralize the sudden realization of "discovering a secret" into "stating a fact". All three can translate "what there is", but fail to convey the emotional twist of sudden realization in "原来".

(2) 我那十年的青春,我那十年的真情,我都喂了狗了.

(Episode 51 of *Mrs. Su's High-Profile Divorce*)

DeepSeek-V4-Pro: Those ten years of youth... those ten years of true love... I fed them all to the dogs.

Doubao 2.1 Turbo: Ten years of my life, ten years of my heart... fed to the dogs.

DeepSeek-R1: My ten years of youth, my ten years of true love—fed to the dogs.

This line is thrown out by the female protagonist in response to her ex-husband's tentative question about whether she has any regrets after receiving the divorce certificate. It is not a breakdown accusation, but a celebratory closure.

DeepSeek-V4-Pro has complete semantics and the best fidelity, but its overly complete sentence structure turns light irony into a heavy accusation, misaligned with the celebratory scenario. Doubao 2.1 Turbo has a crisp rhythm and the best Skopos rule performance, but "life" and "heart" blur the distinction between "youth" and "true affection". DeepSeek-R1 retains the parallel structure with an em dash, balancing fidelity and rhythm, making it the best performer in emotional conveyance among the three models. A translation more consistent with the Skopos rule would be "Ten years of my youth, ten years of my heart—all went down the drain". "Down the drain" is an English colloquial expression meaning "wasted, all for nothing". Compared with "fed to the dogs", it has less intensity of self-deprecation and more cold negation of "meaninglessness", consistent with the protagonist's character and with accurate emotional conveyance.

The common limitation of the three models is that they uniformly render "喂了狗" as "fed to the dogs", preserving the metaphor but leaning toward a breakdown scenario, failing to adjust emotional intensity according to the specific context of "receiving the divorce certificate".

(3) 如果知道我后来会那么地爱你,第一次见面的时候,我一定不会迟到.

(Episode 1 of *Summer Findela*)

DeepSeek-V4-Pro: If I'd known I'd love you this much later on, I never would've been late that first day we met.

Doubao 2.1 Turbo: If I'd known I'd love you this much later on, I never would've been late to our first meeting.

DeepSeek-R1: If I'd known I'd love you this much, I never would've been late that first day.

This line is the female protagonist's reflection on their first meeting when her marriage is happy, a sigh of relief with the feeling of "fortunately I didn't miss out". The stretched pronunciation of "那么地" carries deep affection, and the resolute negation of "一定" conveys the weight of promise.

DeepSeek-V4-Pro has the most complete semantics and the best fidelity, but "this much later on" quantifies the deep affection carried by the stretched pronunciation, resulting in flat emotion. Doubao 2.1 Turbo has the same structure as DeepSeek-V4-Pro, but the ending "our first meeting" is somewhat formal, with slightly less colloquial intimacy. DeepSeek-R1 omits "later on", with a tight rhythm but lost sense of temporal progression. A translation more consistent with the Skopos rule would be "If I'd known back then how much I'd come to love you, I never would have been late the first time we met". "Back then" pulls back to the past perspective; "I'd come to love you" matches the weight of time with a dynamic process; "never would have been late" conveys the weight of promise through resolute negation.

The common limitation of the three models is that they render "那么地" as an adverb of degree plus a time marker, translating the degree but failing to convey the deep affection and sense of fate within the stretched pronunciation.

(4) 我好像,真的有点爱上周晟安了.

(Episode 25 of *Summer Findela*)

DeepSeek-V4-Pro: I think... I might really be falling for Zhou Sheng'an.

Doubao 2.1 Turbo: I think... I might actually be falling for Zhou Sheng'an.

DeepSeek-R1: I think... I really might love Zhou Sheng'an.

This line is the protagonist's sigh when confessing her feelings to her best friend after her relationship with the male protagonist heats up. The three progressive layers of "好像", "真的", and "有点" present the complex psychology of "unexpectedly falling in love despite a forced arranged marriage".

DeepSeek-V4-Pro retains the three progressive layers with the best fidelity, but "really" is rather plain, lacking a sense of surprise. Doubao 2.1 Turbo's "I think... I might actually be falling for" conveys the surprise of "unexpectedly truly" with "actually"; "falling for" is milder and more colloquial than "love", fitting the stage of "just starting to have feelings", with the best Skopos rule performance among the three models. DeepSeek-R1 replaces "falling for" with "love", skipping the gradual process and compressing three psychological layers into two.

When dealing with multi-layered psychological progression, the three models tend to compress layers for the sake of grammatical fluency. DeepSeek-V4-Pro is the most complete but lacks surprise; Doubao 2.1 Turbo balances both layers and tone; DeepSeek-R1 has the greatest semantic leap.

4.3. Limitations in the dimension of subtitle segmentation and syntactic restructuring

Long paratactic Chinese sentences rely on word order itself as narrative logic, and when translated into English, implicit relations must be explicitized into grammatical cohesion. All models can complete this basic conversion, but at the micro level such as preposition selection, information block division, and the weight of modifying components, they often fail to match the narrative rhythm. Syntactic restructuring is not a purely grammatical operation—the order of information presentation directly determines the narrative rhythm of subtitles and the audience's comprehension efficiency. In terms of overall trends, DeepSeek-R1 occasionally has advantages in this dimension due to its tendency toward conciseness; DeepSeek-V4-Pro has the most complete grammar but slower rhythm; Doubao 2.1 Turbo has clear information but often adds extra translations. Four typical translation examples are selected below for specific analysis:

(1) 给我丈夫开了十年车的司机老周,在退休当天,告诉了我一个秘密.

(Episode 1 of *Mrs. Su's High-Profile Divorce*)

DeepSeek-V4-Pro: On the day my husband's driver of ten years officially retired, Old Zhou told me a secret.

Doubao 2.1 Turbo: Lao Zhou, my husband's personal chauffeur of ten years, spilled a long-concealed secret the day he retired for good.

DeepSeek-R1: Old Zhou, my husband's driver of ten years, told me a secret the day he retired.

This sentence is the female protagonist's opening narration for the entire drama. The four pieces of information are naturally connected by word order in Chinese, and as the first sentence of a short drama, it requires a crisp rhythm.

DeepSeek-V4-Pro opens with "On the day...", with complete grammar and the best fidelity, but the focus falls on the time frame rather than the core event, resulting in a slow rhythm. Doubao 2.1 Turbo has the clearest information hierarchy and the best coherence, but "personal chauffeur" and "spilled a long-concealed secret" involve excessive added translation; an opening line requires directness rather than embellishment. DeepSeek-R1 is the most concise, with four information points naturally output in Chinese word order, its rhythm fitting the opening of a short drama, with good performance in both Skopos and coherence.

The three models have different tendencies: DeepSeek-V4-Pro values grammatical completeness but has a slow rhythm; Doubao 2.1 Turbo values clear information but over-embellishes; DeepSeek-R1 values conciseness and directness, with the best rhythm for this sentence, though there is still room for refinement in preposition selection. The common dilemma of all three in this sentence is that they take "complete information" or "grammatical correctness" as their primary goal, rather than "letting the audience grasp the core event in the shortest time"—this is a typical manifestation of models being unable to adjust the order of information presentation according to narrative rhythm.

(2) 也就是说,十年婚姻,他背地里偷偷转移的资产,至少有两个亿.

(Episode 5 of *Mrs. Su's High-Profile Divorce*)

DeepSeek-V4-Pro: So what you're saying is... over ten years of marriage, the assets he secretly siphoned off behind my back total at least two hundred million yuan.

Doubao 2.1 Turbo: So after ten years of marriage, he's secretly moved at least 200 million in assets.

DeepSeek-R1: So in ten years of marriage, the assets he secretly siphoned off behind my back total at least 200 million.

This line is the protagonist's summary and confirmation after hearing the information from her best friend. "也就是说" connects the preceding and following content, carrying the summarizing meaning of "highlighting the key point".

DeepSeek-V4-Pro has the most complete semantics and the best fidelity, but the whole sentence is rather long, slowing down the colloquial rhythm. Doubao 2.1 Turbo is the most concise with a crisp rhythm and the best Skopos rule performance, but the vivid action of "siphoned off" is compressed. DeepSeek-R1 has a moderate structure, but its attributive clause is tight, with slightly weaker colloquial feel. A translation more consistent with both the coherence and Skopos rules would be "Ten years of marriage, and he secretly moved at least two hundred million yuan in assets—behind my back". "Ten years of marriage" stands alone as the opening, directly juxtaposed with "two hundred million yuan", creating an immediate contrast; the em dash places "behind my back" at the end as the emotional focal point.

The common tendency of the three models is to incorporate "ten years of marriage" into a temporal adverbial using prepositions, which is grammatically standard but weakens the topical sense. DeepSeek-V4-Pro values completeness but has a slow rhythm; Doubao 2.1 Turbo values rhythm but has semantic compression; DeepSeek-R1 is moderate but has tight sentence structure. None of the three makes temporal information an independent block to form a natural colloquial connection with event information.

(3) 不管我们两个合不合适,这个婚我们都得凑合着结.

(Episode 1 of *Summer Findela*)

DeepSeek-V4-Pro: Suited or not, we're stuck with this marriage.

Doubao 2.1 Turbo: Whether we're compatible or not, we have to make this marriage work.

DeepSeek-R1: Whether we fit or not, we're stuck getting hitched.

This line is the protagonist's statement when she first meets the male protagonist under the family arranged marriage. "凑合着结" is a mixture of passive compromise and helpless resignation.

DeepSeek-V4-Pro uses "stuck with this marriage", with the most accurate emotion, balancing both fidelity and Skopos. Doubao 2.1 Turbo's "make this marriage work" leans toward "working hard to manage the marriage", which is completely opposite to the passive compromise of "凑合", resulting in the most obvious point loss in the Skopos rule. DeepSeek-R1's "stuck getting hitched" has semantic deviation, as "getting hitched" refers to the act of getting married rather than the state of marriage. A translation more consistent with the Skopos rule would be "Whether we're a good match or not, we're stuck with this marriage anyway". "Stuck with this marriage" accurately conveys the passive state of maintaining a marriage perfunctorily, and "anyway" supplements the sense of helplessness of "having to get married no matter what", with emotional nuance conveyance meeting the Skopos rule.

The three models each have deviations in handling the emotional keyword "凑合": DeepSeek-V4-Pro is the most accurate; Doubao 2.1 Turbo understands it in the opposite direction; DeepSeek-R1 has semantic offset. The common limitation is the insufficient precision in identifying the emotion of passive compromise.

(4) 约法三章第一章,婚后,不得以任何理由,干涉对方的事业.

(Episode 2 of *Summer Findela*)

DeepSeek-V4-Pro: Chapter One of the ground rules. After marriage, neither party shall interfere with the other's career, for any reason.

Doubao 2.1 Turbo: Rule number one: After marriage, neither of us can interfere in the other's career for any reason.

DeepSeek-R1: Rule one—after we're married, neither of us interferes with the other's career.

This line is the first marriage rule set by the female protagonist for the male protagonist when they first meet. Spoken in a calm tone, it establishes boundaries for the forced arranged marriage.

DeepSeek-V4-Pro opens with "Chapter One of the ground rules", with a strong legal tone and the best fidelity, but as a colloquial scenario of first meeting, the tone is too stiff. Doubao 2.1 Turbo's "Rule number one" is more colloquial with the best Skopos rule performance, but replacing "neither party shall" with "can't"

weakens the binding force of the rule. DeepSeek-R1 is the most concise with a crisp rhythm, but loses the scope limitation of "for any reason". A translation more consistent with the coherence rule would be "First rule: after we get married, neither of us interferes with the other's career—for any reason". "First rule" balances solemnity and colloquialism; "after we get married" has a more conversational feel than "after marriage"; the em dash places "for any reason" at the end to emphasize the exclusivity at the core of the rule.

The three models have different tendencies when handling rule-based lines: DeepSeek-V4-Pro values formality but lacks adaptation to colloquial scenarios; Doubao 2.1 Turbo values colloquialism but weakens the binding force of rules; DeepSeek-R1 values conciseness but has semantic gaps. The common limitation is the insufficient precision in balancing the dual contexts of "rule declaration" and "colloquial dialogue".

5. Conclusion

Through sentence-by-sentence comparative analysis of 12 typical cases from two dramas, it is found that there is a core paradox in large language models' short drama subtitle translation: they can translate words, but not the functions of words in the plot; they can translate sentences, but not the emotions carried by sentence rhythm. The root cause of this defect is not insufficient grammatical ability, but that models generate translations based on corpus probability distribution, lacking active judgment of "translation purpose"—the "optimal output" at the semantic level is not equal to the "best choice" at the pragmatic level. Specifically: DeepSeek-V4-Pro excels in completeness and fidelity, suitable for preliminary translation of information-intensive lines, but its emotional conveyance is flat and its intensity is imbalanced across different scenarios; Doubao 2.1 Turbo excels in fluent colloquialism, suitable for rapid generation in dialogue scenarios, but has obvious loss of cultural information and occasional reverse comprehension—for example, translating "凑合着结" as "make this marriage work", misinterpreting passive compromise as active management; DeepSeek-R1 excels in minimalist rhythm with occasional highlights, but lacks grammatical stability, as syntactic fractures such as "Ten years devoted wife" appear repeatedly. The common problem of the three is that it is always difficult to balance between fidelity and colloquial rhythm, and between cultural information fidelity and emotional intensity conveyance.

The prompts used in this study are customized according to drama styles, aiming to make model outputs closer to the translation purpose. However, prompt optimization cannot fundamentally replace the cross-cultural judgment of human translators.

Judging from the analysis of 12 typical cases, none of the three models currently has the ability to independently complete high-quality Chinese-to-English subtitle translation for short dramas, but as auxiliary tools, they can be selected differentially according to scenarios: DeepSeek-V4-Pro is suitable for preliminary translation of information-intensive lines; Doubao 2.1 Turbo is suitable for rapid generation in dialogue scenarios; DeepSeek-R1 can provide reference for concise expression. On the fast track of short drama overseas expansion, large language models and human translators are not in a simple substitution relationship. In the future, a collaborative model of "AI preliminary translation + human polishing + cultural proofreading" can be explored—AI undertakes efficiency, and translators undertake precision, jointly serving the international communication of Chinese digital culture.

However, this study also has certain limitations: one of the core characteristics of large language models is that the same input may produce different outputs in different rounds. This study performed two generations for each model and selected the better version, but did not conduct systematic stability tests with multiple generations. Therefore, the analytical conclusions may not completely exclude the influence of randomness in

model output. Future research can further verify the reliability of the above findings by averaging multiple generations or introducing stability evaluation indicators.

References

- [1] DataEye Research Institute. (2026). 2025 overseas micro-short drama data report. *DataEye*.
- [2] Zhao, Z. G. (2025, April 18). China's first intelligent translation and dubbing center for micro-short dramas inaugurated in Malanshan. *Hunan Daily*.
- [3] Gottlieb, H. (1994). Subtitling: Diagonal translation. *Perspectives: Studies in Translatology*, 2(1), 101–121.
- [4] Yao, S. (2016). Subtitle translation under the guidance of Skopos theory: A case study of Kung Fu Panda. *Henan Agriculture*, (2), 50–51.
- [5] Li, H. Q., & Bo, Z. J. (2005). Norms and film subtitle translation. *Chinese Science & Technology Translators Journal*, 18(2), 44–46.
- [6] Liu, L. A. (2008). An analysis of film subtitle translation from the perspective of Skopos theory. *Anhui Literature (Second Half of the Month)*, (12), 100–101.
- [7] Sun, J. Y. (2025). Research on human-machine collaborative subtitle translation from the perspective of functional equivalence: A case study of Bian Shui Wang Shi. *Foreign Language Education and Translation Development Innovation Research*, (17), 49–52.
- [8] Zhang, J. L. (2004). Skopos theory and translation methods. *Chinese Science & Technology Translators Journal*, 17(1), 35–37.
- [9] Vermeer, H. (2001). *A framework for a general theory of translation*. Shanghai Foreign Language Education Press.
- [10] Yin, X. F. (2024). Strategies of film subtitle translation from the perspective of Skopos theory. *Movie Literature*, (18), 130–134.
- [11] Guo, S. Q. (2013). *Subtitle translation of the Painted Skin film series from the perspective of Skopos theory* [Master's thesis, Beijing International Studies University].
- [12] Li, X. P., & Zhou, X. W. (2020). Translation of culture-loaded expressions in film subtitles from the perspective of Skopos theory: A case study of Ne Zha: Birth of the Demon Child. *Popular Literature & Art*, (18), 144–145.
- [13] Zhang, J. Y., & Zhang, X. M. (2024). An analysis of culture-loaded expression translation in Chang'an subtitles from the perspective of Skopos theory. *English Square*, (7), 7–10.