| RESEARCH ARTICLE

SubGPT: An Automated Tool for Subtitle-Template Translation

Hussein Abu-Rayyash[1] ✉ and Ali F. Al-Ramadan[2]

[1]Kent State University, Department of Modern & Classical Language Studies, USA
[2]Kent State University, Department of Computer Science, USA

Corresponding Author: Hussein Abu-Rayyash, **E-mail**: haburayy@kent.edu

| ABSTRACT

SubGPT is an automated subtitle translation tool that integrates OpenAI's GPT to generate translations from Source Language Subtitle Template (SRT) files. The tool employs customized prompts to produce translations adapted for different dialects and vernaculars, while maintaining the spatial and temporal aspects of the original subtitles. A mixed-methods approach is used to evaluate SubGPT's performance, combining cosine similarity metrics and qualitative assessments from professional subtitlers. The results demonstrate the tool's effectiveness in handling diverse linguistic challenges and its potential for improving efficiency and accessibility in the audio-visual translation (AVT) industry. The tool contributes to the advancement of the AVT field through the integration of large language models for culturally and linguistically automated translation, with potential applications in different domains such as entertainment and education.

1. Introduction

Subtitling enables accessibility and cross-cultural exchange of audio-visual (AV) content by providing textual representations of the spoken source language dialogue. However, creating high-quality subtitles that convey both linguistic meaning and cultural nuances is challenging for AI-driven translation. Traditional subtitling heavily relies on human subtitlers, resulting in a process that is expensive, time-consuming, and inconsistent across languages and contexts. Recent developments in artificial intelligence (AI) and natural language processing (NLP) present opportunities for automating high-quality subtitling. However, subtitling is entering a new era as traditional software evolves into flexible, online platforms that incorporate translation memories, machine translation, and AI (Athanasiadi, 2017).

This project presents SubGPT, an automated subtitling tool that integrates OpenAI's GPT to generate translations of the text in Source Language Subtitle Template (SRT) files. SubGPT uses customized prompts for GPT to produce culturally relevant translations adapted to dialects and vernaculars. As stated by Athanasiadi (2015), high-quality subtitling transcends mere language translation, demanding a profound understanding and incorporation of cultural context. SubGPT rises to this challenge, leveraging the capabilities of GPT to address the critical need for cultural specificity, setting a new standard in automated subtitling by delivering translations that truly resonate with the target audience on a deeper, more meaningful level.

The effectiveness of SubGPT in producing high-quality, culturally adapted subtitles across diverse languages is empirically demonstrated. The generated subtitles are semantically and typographically similar and/or different from the source text such as between English and Arabic. In evaluating SubGPT, both cosine similarity metrics and assessments from a professional subtitler were used, providing a balanced evaluation that includes numerical measures of accuracy and qualitative insights from experts.

Thus, SubGPT represents a substantial advancement in the AVT field, setting new standards for culturally tuned and temporally synchronized automated multi-language subtitling through the integration of GPT.

2. Literature Review
2.1 The Intersection of AVT and AI
The landscape of audio-visual translation (AVT) has evolved significantly since its beginnings, which were characterized by manual subtitling and revoicing techniques (Díaz-Cintas and Remael, 2007). This transition has been further catalyzed by the proliferation of streaming platforms, necessitating a reevaluation of conventional AVT methods due to their cultural and technical constraints (Pedersen, 2007). Early advancements in AVT technology primarily manifested in specialized software applications, such as Subtitle Edit and Aegisub, which facilitated the automation of subtitling tasks, including transcription and timing, contributing to workflow efficiencies (Gambier, 2018). However, this automation tended to prioritize temporal synchronization over cultural appropriateness. As noted by Linde and Kay (1999), this focus on technical accuracy often compromised the complex localization of content, resulting in decontextualized subtitles. Furthermore, the advent of cloud-based subtitling solutions has perpetuated this issue by emphasizing swift linguistic translation without considering cultural ramifications (Bolaños-García-Escribano and Díaz-Cintas, 2020). With global streaming subscriptions projected to reach approximately two billion viewers by 2025, the need for culturally sensitive AVT strategies has intensified (Hepburn, 2021).

In this evolving space, AI is a disruptive force capable of revolutionizing existing paradigms in translation studies, including AVT. Large Language Models (LLMs), such as OpenAI's GPT, have advanced semantic processing capabilities (Kheiri and Karimi, 2023). Similarly, Yamada (2023) states that the auto-regressive architecture of GPT has shown promise for handling context-dependent translation tasks requiring cultural cognizance. Specifically, GPT can employ strategies like dynamic equivalence, where culturally specific expressions are replaced with relevant terms or concepts in the target language and culture. Parallel to these developments, reusable subtitle templates have proven instrumental in globalizing AVT workflows (Georgakopoulou, 2019). These templates allow translators to decouple timing information from linguistic translation, facilitating an expedited process (Artegiani and Kapsaskis, 2014).

The growing influence of global media emphasizes the need for AVT to expand its scope beyond mere linguistic translations to encompass culturally informed translations (Bielsa and Bassnett, 2008). In both entertainment and educational contexts, there is a high demand for culturally localized subtitles, as opposed to merely linguistically accurate ones (Abu-Rayyash et al., 2023). The emergence of translation models that understand context opens new opportunities to combine linguistic accuracy with cultural understanding in audio-visual translation. As a result, the stage is set for the development of AI-powered tools capable of auto-generating subtitles that are both linguistically precise and culturally driven.

2.2 Subtitle-template Files
Subtitle template files have become a core component of the globalized AVT industry over the past two decades (Georgakopoulou, 2019). These files serve as a standardized base for creating multi-language subtitle files, retaining timing information while allowing translators to focus on linguistic transfer (Artegiani and Kapsaskis, 2014). Although initially created in English, template files are frequently generated in languages other than English to serve as a pivot for additional target languages (Nikolić, 2015). The main benefits of using subtitle template files are efficiency in managing large multi-language projects and facilitating quality control. However, some concerns have been raised regarding their impact on translation quality and subtitler compensation (Georgakopoulou, 2006; O'hagan, 2005).

Subtitle template files represent an effort to find "a compromise between different national styles of subtitling" (Georgakopoulou, 2019, p. 140). Parameters such as reading speed, line length, and shot change handling are standardized based on viewer preferences, although concessions are still made to account for regional subtitling traditions. Over time, templated workflows have likely conditioned viewers to accept more "homogenized subtitling styles" (Nikolić, 2015, p. 199). Recent reception studies have started to examine viewing behaviors like reading speeds and shot change re-reading to inform template style guidelines (Krejtz et al., 2013; Szarkowska and Gerber-Morón, 2018).

Currently, the subtitling industry predominantly uses template-based methods. However, there is room for improvement in automation by using AI tools that can generate translations that fit both the language and culture from a template. According to Georgakopoulou (2019), template files provide an excellent base content resource for translation. Advanced machine translation (MT) and natural language processing techniques can then be applied to provide an initial draft translation from the template (García-Escribano, 2023). This allows human post-editors to focus their efforts on refining and customizing the machine-generated translations efficiently through a "full post-editing of the templates" (García-Escribano, 2023, p. 12). Such AI-based authoring solutions promise to balance the benefits of templated workflows with customized, high-quality subtitle experiences for diverse

global audiences. The use of AI for automated template creation and translation could potentially show promising results, although human editing remains crucial for quality assurance.

2.3 GPT Prompt Engineering

The recent advancements in Machine Translation (MT) capabilities of conversational agents like GPT have been extensively studied (Yamada, 2023). Comprehensively, Jiao et al. (2023) offer an analysis of GPT's MT competence, revealing that it performs comparably to commercial systems in high-resource European languages. However, it was found to struggle with languages that are either low-resource or linguistically distant from those it has been trained on. Their study also highlights the enhancement of translation quality through pivot prompting and indicates that GPT performs better in spoken language translation. Later, the GPT-4 model further enhances these capabilities. In the same vein, Gao et al. (2023) pioneer a new methodology for translation prompts that involves incorporating task information, context domain information, and Part-of-Speech tags. This approach was shown to surpass the performance of commercial systems in various translation directions, particularly when using few-shot prompts.

Moslem et al. (2023) focus on the need for both linguistic consistency and domain-specific terminology in translations. Their research suggests that large-scale language models can be harnessed to improve real-time adaptive MT, especially for high-resource languages. For less-supported languages, they propose a collaborative approach involving strong encoder-decoder models and fuzzy matches. Furthermore, Gu (2023) addresses a specific problem concerning the translation of Japanese attributive clauses into Chinese by using a pre-edit method and a two-step prompt with GPT, leading to a noticeable improvement in translation accuracy. Despite these advancements, the academic discourse in the translation studies field has primarily centered on language-specific issues, low-resource language pairs, and challenges that conventional MT systems are not equipped to resolve.

2.4 Theoretical Framework

Our study integrates core AVT theory centered on the evolution of quality assessment models for judging subtitle output. Early scholarship focused on parameters like reading speed, shot transition re-reading, and degree of viewer satisfaction (Gambier and Di Giovanni, 2018). However, as the intricacies of cultural awareness and semantic fidelity garnered more attention, multifaceted frameworks emerged, combining quantitative metrics and qualitative judgments (Bolaños-García-Escribano & Díaz-Cintas, 2020). Contemporary paradigms assess elements ranging from terminology accuracy, grammar, spelling, mistranslations, and omissions relative to source content using normalized error scales to dynamic replacement of idioms and regional vernacular (Pedersen, 2007). Integrating such crucial protocols through semantic analysis and professional subtitler evaluation allows for a layered insight into machine-generated subtitles.

Furthermore, the proliferation of AI and neural networks has profoundly impacted AVT platforms and workflows. As Natural Language Processing (NLP) and Machine Learning (ML) models have matured in capabilities, the translation field has rapidly shifted from purely manual techniques to blending automation efficiencies with human expertise (Athanasiadi, 2017). Modern conversational frameworks like GPT showcase remarkable human-like generation trained on immense datasets (Kheiri and Karimi, 2023). This potential has focused scholarship on customizing model prompts to maximize accuracy for target languages and cultures (Gao et al., 2023). Furthermore, using reusable subtitle templates as structured inputs balances automation with allowing for refinement (Georgakopoulou, 2006). Consequently, the SubGPT tool development applies such guidance on prompt engineering and templates to drive quality output.

Therefore, SubGPT bridges core concepts, including multi-parameter translation quality models evaluating technical precision, linguistic accuracy, semantic similarity, and cultural contextualization. It also draws upon the sophisticated natural language generation capabilities of modern AI models like GPT that integrate immense dataset training. Additionally, the methodology applies lessons from scholarship on effectively leveraging reusable subtitle templates to inform automated workflows (Georgakopoulou, 2006). This grounds the SubGPT solution, which evaluates output quality through cosine similarity analysis between source and target vectors and qualitative judgments by professional subtitlers. It also enables ongoing refinement of the automated subtitling tool based on these results. As Pedersen (2007) highlighted, harnessing technology to facilitate cultural exchange remains an integral challenge for the field to drive solutions that blend efficiency with understanding.

2.5 Research Gap and Significance

While existing research has explored harnessing large language models like GPT for generalized machine translation use cases (Yamada, 2023) and formulating prompt engineering techniques to heighten translation performance (Gao et al., 2023), there remains a tangible gap in studies explicitly examining the suitability of GPT for automated AVT. In particular, prior work has concentrated on optimizations for translation in narrow domains (Moslem et al., 2023) rather than constructing an end-to-end solution tailored for subtitling, making use of reusable subtitle templates (Georgakopoulou, 2019) paired with prompt tuning to preserve cultural awareness. Therefore, an opportunity exists to address this gap by developing and critically evaluating an

automated subtitling tool designed for contextual dependent AVT challenges. Assessing output through multi-parameter translation quality models (Bolaños-García-Escribano and Díaz-Cintas, 2020) can reveal multilayered performance insights. Therefore, the proposed SubGPT tool represents an effort to address this by:

1. developing a complete automated subtitling tool powered by GPT and specially tuned for subtitling challenges;
2. incorporating reusable subtitle templates to structure translation workflows; and
3. evaluating output quality through cosine similarity metrics, professional subtitlers' judgments, and iterative refinements.

3. Methodology
3.1 SubGPT Architecture
The SubGPT tool is developed using the Qt framework, which consists of a comprehensive set of intuitive and modular libraries that allow for the development of cross-platform applications for desktop, embedded, and mobile platforms. The programming language of choice is Python. Qt has Pythonic bindings provided by the PySide library, which allows for a seamless integration of the Qt framework with Python. The underlying framework and architecture of SubGPT is centered around OpenAI's GPT API. The tool utilizes SRT files as input and uses customized prompts for the API to produce culturally relevant translations adapted to different dialects and vernaculars. Figure 1 shows the interface of the tool.

Figure 1. SubGPT Interface
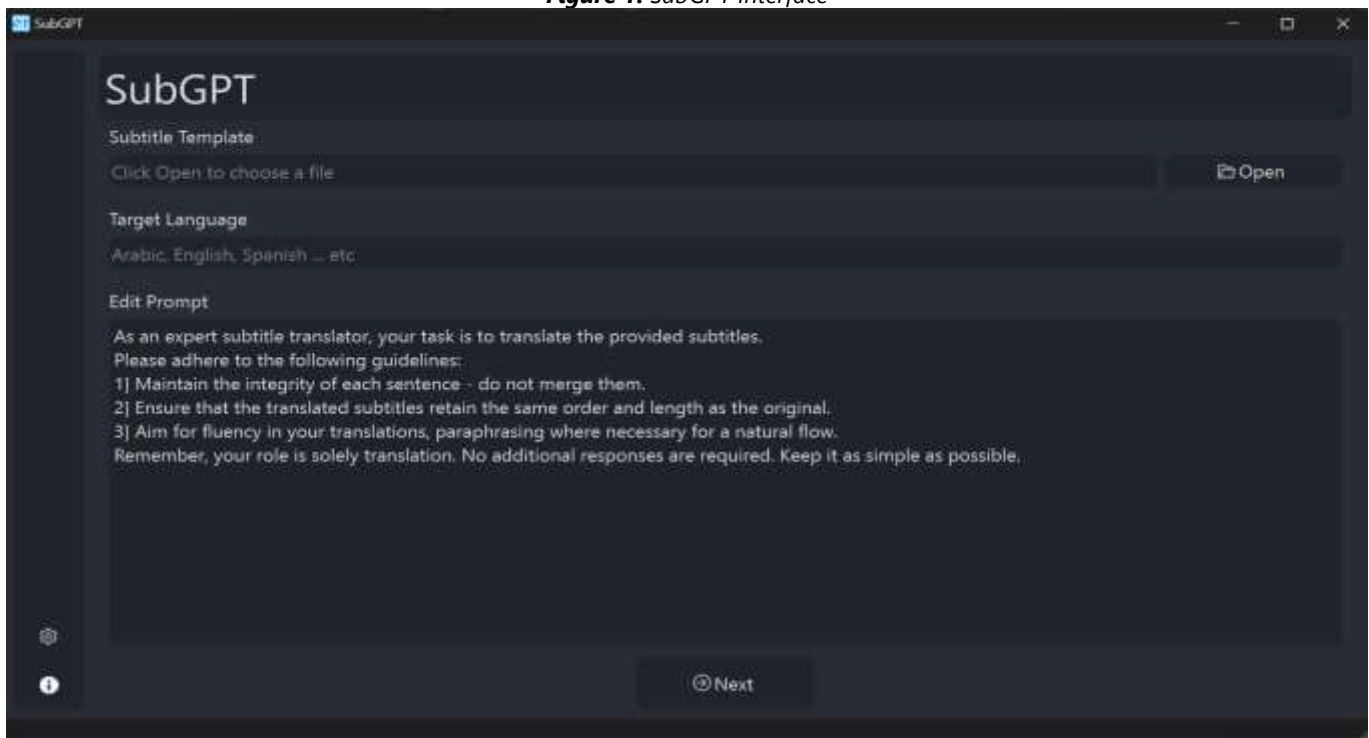
3.2 Tool's Evaluation
For testing and validation of SubGPT, both quantitative and qualitative approaches are used. The quantitative approach involves using cosine similarity to compare the embeddings of the generated translated output with the embeddings of the subtitle-template files or input. Embeddings, in the context of NLP and machine learning, refer to the representation of words, phrases, or documents as dense vectors of real numbers. These vectors capture the semantic and syntactic relationships between words or documents in a high-dimensional space. Cosine similarity serves as a metric for comparing vectors in a high-dimensional space, thereby providing quantitative insights into semantic parallels between different languages. These embeddings can be obtained using the OpenAI embeddings API, which provides pre-trained models for generating embeddings for text data. Despite computational efficiency, this method is not without limitations. For instance, it may overlook cultural nuances or idiomatic expressions that are crucial in translation studies (Yamada, 2023). To mitigate these potential threats to validity, a qualitative evaluation by professional human subtitlers is incorporated, supplementing the quantitative cosine similarity metrics.

The source materials selected for experimentation consist of a carefully curated set of Arabic movie excerpts, chosen for their linguistic richness, diverse vernaculars, and idiomatic expressions. Spanning various genres, including drama, comedy, action, and historical films, these excerpts ensure a comprehensive representation of the Arabic language in different contexts. The selection

process prioritizes movies that showcase a wide range of vocabulary, syntax, semantics, cultural references, and regional dialects, capturing the nuances and complexities of the Arabic language. To ensure the suitability of these excerpts for subtitling, the following criteria were applied: 1) the presence of clear and concise linguistic complexity, allowing for effective translation and comprehension; 2) the inclusion of cultural-bound references, highlighting the importance of cultural adaptation; and 3) the use of idiomatic expressions, testing the ability to accurately convey the intended meaning. This selected set of data serves as a valuable resource for testing and validating the translation capabilities of automated subtitling tools across various linguistic and cultural aspects of the Arabic language, while also providing a rich resource for applications such as language teaching, translator training, and AVT research.

Translation quality assessment has seen meaningful evolution in paradigm metrics, with models like BLEU (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005), and cosine similarity (Mikolov et al., 2013), each with its unique capabilities and constraints (Han et al., 2021; Lin and Xie, 2020; Saadany and Orasan, 2021). BLEU remains widely adopted given its simplicity but is limited in application by sensitivity to minor syntactic variations, even when core semantic meaning is preserved (Saadany and Orasan, 2021). METEOR strives to address this through more flexible matching, allowing for paraphrases but retains inherent weaknesses in accounting for conceptual equivalents (Lin and Xie, 2020). Alternatively, cosine similarity offers a compelling approach comparing complete vector representations of source and target text enabled by modern transformer embeddings to quantify semantic similarity regardless of phrasing structure. Rather than strictly matching surface forms, it can accommodate synonymous terms or differing idiomatic expressions that relay the same underlying meaning (Han et al., 2021). Given AVT intricacies requiring cultural awareness and localization beyond literal form-to-form mapping, cosine similarity's robustness to these dynamics makes it an optimal match. Our study supplements cosine similarity with qualitative human assessments to provide insights on translation quality. Therefore, cosine similarity enables efficient, vector-based semantic similarity analysis supported by contextual human judgment.

The selection of the English-Arabic language pair is purposeful to test the tool across distinct linguistic and cultural contexts. Arabic is chosen due to its semantic and structural distance from English. Each translation is subjected to a cosine similarity threshold, which is adjustable based on parameters such as text genre and target audience when translating between English and Arabic. In the human evaluation phase, professional subtitlers with expertise in the target languages are recruited to assess the translated content. Their role is twofold: first, to compare the tool-generated translations against other subtitlers translation of the source texts, and second, to assess how accurate the translations captured both the semantic and the cultural complexities of the source text. This evaluation is conducted using a predefined rubric.

In accordance with Pym (1992), the evaluation phase also involves generating multiple translations to allow for a more complex selection process. By providing the human evaluator with a variety of translation options, we aim to emulate real-world translation scenarios. In professional practice, translators do not just create a single translation but evaluate multiple alternatives to choose the most contextually and semantically appropriate version (Yamada, 2023). Therefore, this methodology serves a dual purpose: it quantitatively measures the translation's accuracy through cosine similarity metrics and qualitatively assesses its suitability through expert evaluation.

4. Results and Discussion
4.2.1 Quantitative Evaluation
To evaluate the output quality, cosine similarity was used to compare the semantic vectors of the input source text and the translated output. This method allowed for measuring the preservation of meaning in subtitle translation despite variations in phrasing and word choice that arise from localization.

The evaluation results showed an average cosine similarity score of 0.82 across different selected texts from the dataset, indicating that the automated subtitle translation framework performed well in maintaining the meaning of the source text when dealing with a variety of cinematic linguistic challenges between English and Arabic. However, performance varied depending on the complexity of the genre. The framework excelled in translating mainstream twentieth-century comedic animations, achieving a score of 0.88, as it successfully adapted youthful vernacular. On the other hand, it encountered some difficulties with early nineteenth-century docudrama exchanges containing less familiar cultural elements and complex vocabulary, resulting in a lower similarity score of 0.56 for this subset. These findings suggest that while SubGPT demonstrates overall robustness in handling diverse subtitle translation tasks, there is room for improvement in dealing with certain genres and time periods that present more complex linguistic and cultural challenges.

4.2.2 Qualitative Evaluation
A qualitative experiment was conducted to evaluate SubGPT's ability to adaptively localize Arabic transcriptions of sitcom dialogue into English equivalents by comparing them against professional translations (as shown in Table 1). Eight key criteria were used to

assess the preservation of cultural nuance: Colloquial Adaptation, Cultural Resonance, Fluency, Naturalness, Humor Preservation, Age Resonance, Sitcom Style, and Overall Quality. Experts rated the outputs using a five-point scale from Poor to Excellent based on how well they matched professional quality standards in comedy subtitling.

***Table 1.** Original Utterances and Professional Translations*

Original Utterance	Professional Translation	Utterances' Pragmatic Function
لا أراكم الله مكروهًا.	My condolences.	Condolences
مات أبي ورأينا مكروهًا بالفعل.	You gave enough of that, Hajj.	Sarcasm
ما خطبك يا هذا؟ لماذا ترمقني بهذه النظرات	Why do you keep looking at me like that	Confusion/Question
اغرب عن وجهي، لا أريد أن أراك مجددًا!!	Scram. I don't want to see you again.	Anger
ألقوا بهذا الرجل من فوق السطح.	Throw this man over the roof.	Threat
هلّا نجلس ونتناقش يا "بوحة."	Let's sit down and talk, Booha.	Invitation to talk
انفضّ العزاء. وقد مات أبوك ودُفن.	The funeral is over, and your father was buried.	End mourning period
وشؤون الأحياء أولى بالاهتمام.	Time to take care of the living.	Moving on
كفى يا "بوحة" "يا ابن أخي.	Come on, Booha, my nephew.	Frustration
لم يقصد عمك "همّام" مضايقتك.	Your Uncle Hammam didn't mean it.	Explanation
دعنا نتحدث عن الأمور المهمة.	Now let's talk about the important stuff.	Invitation to focus
اقترب. اجلس يا "بوحة!"	Come. Sit down.	Command
اسمع يا "بوحة" يا ابن أخي.	Listen, Booha, my nephew.	Listen up
كما يقولون، اطرق الحديد...	It's best to be serious...	Idiom
لكن سخّنه قبلًا!	And to be serious is best.	Joke
ولا تؤجل عمل اليوم... لأن المدير سيطردك.	And those who are too shy... Can never have kids.	Joke
ركّز واقبل نصيحتنا.	Let's focus on the important stuff and listen to us.	Focus/Advice
ورثت عن أبيك ثروة طائلة يا "بوحة."	Your father left you a big fortune, Booha.	Inheritance
نصيبك في الميراث نصف مليون جنيه.	He left you half a million pounds.	Inheritance
كنت محقًا، الماضي لا يمكن تغييره.	As you said, Hajj, what's gone is gone.	Agreement
رحمك الله يا أبي!	God rest your soul, Dad.	Condolences
"اركض" في سلام!	May you rest in peace!	Farewell/Joke

The expert raters used sample source text excerpts from an Arabic sitcom script and their corresponding professional human translations as a benchmark. This allowed them to evaluate whether the machine-generated subtitles effectively preserved linguistic nuances and sociocultural resonances compared to the gold standard. The source excerpts contained various pragmatic elements such as condolences, sarcasm, confusion, anger, threats, invitations, jokes, focus, and inheritance, while the professional subtitles demonstrated skilled cultural and tonal adaptation to the target language. The study aimed to identify SubGPT's limitations in conveying humor and emotional expressions by translating Arabic transcriptions into entertaining, culturally authentic English subtitles. Results revealed deficiencies in humor preservation using default settings, prompting researchers to customize the prompts to enhance performance.

Figure 2. Custom Prompt Composition
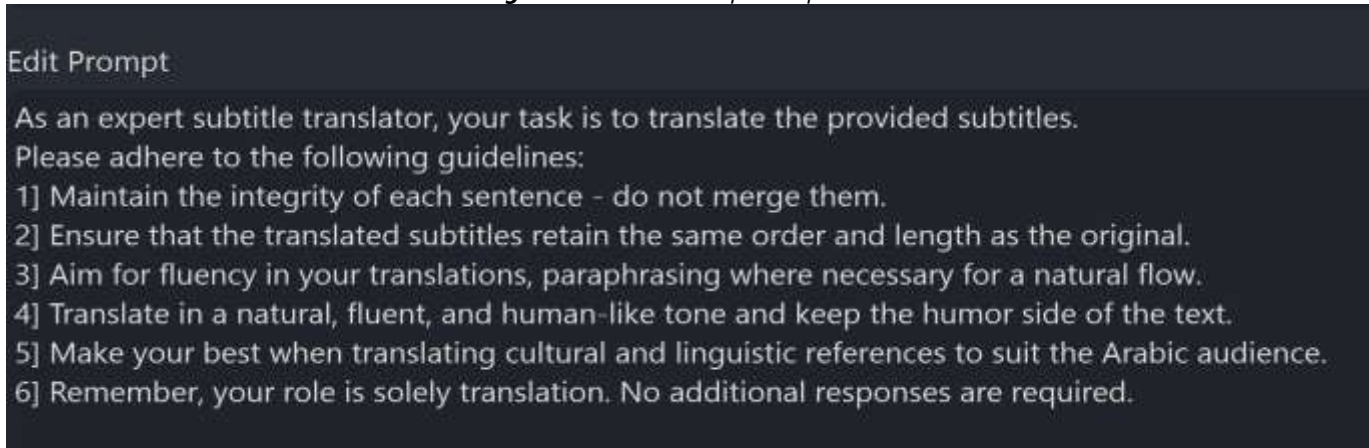
The custom prompt (Figure 2) provides tailored directives to improve output quality by focusing on translation accuracy. The guidelines emphasize maintaining the original integrity of the text, such as order, length, and sentencing, while targeting natural fluency. Crucially, the instructions stress the importance of retaining a humorous tone adjusted for audience resonance. This encoding acts as auxiliary steering for handling complexities around punchlines, banter, and conflict more accurately. By addressing criteria shortcomings through customized prompts, the performance was measurably boosted compared to generic applications, demonstrating the viability of guided augmentation for achieving better entertainment localization.

Table 2. SubGPT Subtitle Scoring by Translation Approach

Criteria	GPT-4 Translation	GPT-4 + Cosine	GPT-4 + Prompt	GPT-4 + Cosine + Prompt
Colloquial Style	Fair	Good	Excellent	Excellent
Cultural Adaptation	Good	Very Good	Good	Excellent
Fluency	Good	Very Good	Good	Excellent
Naturalness	Fair	Good	Very Good	Excellent
Humor Preservation	Poor	Fair	Good	Very Good
Age Resonance	Good	Very Good	Excellent	Excellent
Sitcom Style	Fair	Good	Excellent	Excellent
Overall Quality	Fair	Very Good	Excellent	Excellent

Table 2 compares SubGPT outputs, including the baseline and key variants, across eight criteria dimensions to evaluate entertainment localization capabilities. GPT-4 alone showed "Fair" performance, conveying simple cultural concepts like funeral condolences ("رحمك الله يا أبي!") or inheritance transfers ("fortune"). However, it struggled with localizing complex timing, emotions, sarcasm, and humor in comedy dialogue, as seen in the blunted impact of playful phrases ("بالفعل مات أبي ورأينا مكروهًا"). Still, some vocabulary grasping ("half a million") indicated potential if properly stimulated.

Augmenting the translation process with cosine similarity improved the tool's ability to generate age-appropriate content, which is crucial for sitcoms. This was evident in better cultural adaptation and age resonance scores, as the translations more effectively contextualized playful workplace banter, such as "ولا تؤجل عمل اليوم" ("Don't postpone today's work"). Furthermore, custom-tailored prompting helped align the situational comedy timing with the source material by incorporating age-appropriate elements based on previously identified deficiencies. This approach injected an informal tone that resonated with everyday experiences, such as a caring but sly authority figure common in family settings, as seen in the translation: "Listen, Booha...Don't postpone today's work!"

While directive prompting can add stylistic details, it struggled to capture the double meanings often found in clever punchlines that keep comedy flowing smoothly. In contrast, the hybrid approach excelled at adaptation by combining the tonal control of custom prompts with semantic safeguards to preserve meaning. This SubGPT+Hybrid model showed advanced translation ability, faithfully conveying the core of the comedy even as witty dialogue between characters moved beyond a single interpretation and into entertainment that resonates across cultures.

While professional human subtitlers represent the gold standard for translating comedy entertainment material, their expertise requires tremendous investments of time and skill. Despite performance, relying solely on human expertise has limitations in efficiency, adaptability, and access. Generative tools like SubGPT offer a compelling middle ground, balancing quality with scalability. Within a measurable scope, advanced MT can uphold baseline sitcom enjoyment with fault-tolerant output efficient for many commercial contexts. Real-time generated draft subtitles lower costs and lead times while retaining comprehension. Furthermore, post-processing refinement allows for creative control to balance resource expenses. SubGPT-human hybrid leveraging balances strengths while mitigating limitations, pointing towards more accessible, adaptable, and durable entertainment translation infrastructure through automated workflow assimilation.

5. Conclusion

SubGPT's development and evaluation shed light on the future of AVT, where the synergy between advanced language models and human expertise will likely become the norm. The tool's ability to handle diverse linguistic challenges while maintaining cultural relevance demonstrates the potential for AI to simplify the subtitling process. However, the varying performance across different genres and the need for prompt customization highlight the importance of continued research and development in this field. As the demand for localized content grows, tools like SubGPT will play a crucial role in making AV content more accessible to global audiences. The integration of large language models in subtitling workflows has the potential to significantly reduce the time and cost associated with traditional translation methods, leading to a more diverse range of content being made available to viewers worldwide and fostering greater cultural exchange and understanding.

Moreover, the hybrid approach combining automated translation with human post-editing opens up new possibilities for collaboration between AI and human translators. Rather than replacing human expertise, tools like SubGPT can augment the work of professional subtitlers, allowing them to focus on the creative aspects of translation while the AI handles the more repetitive and time-consuming tasks. Looking ahead, the use of SubGPT in the entertainment industry could pave the way for similar applications in other domains, such as education, news media, and corporate communications. As the technology continues to evolve and improve, it is likely that we will see an increasing number of industries adopting AI-powered translation tools to bridge linguistic and cultural gaps.

The future of AVT is rapidly evolving, and tools like SubGPT are at the forefront of this transformation. As AI continues to advance, it is essential for researchers and developers to prioritize the development of transparent, ethical, and accountable systems that can adapt to the diverse needs of users worldwide. This requires ongoing collaboration between machine learning experts, linguists, and cultural scholars to ensure that these tools are not only technologically sophisticated but also culturally sensitive and socially responsible. By embracing this multidisciplinary approach, we can harness the power of AI to create a more inclusive and equitable global community, where linguistic and cultural diversity are celebrated as strengths rather than barriers. Finally, the success of SubGPT and similar tools will depend on their ability to balance the benefits of automation with the irreplaceable value of human creativity, empathy, and understanding.

Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD
Hussein Abu-Rayyash: https://orcid.org/0000-0002-9695-4030
Ali F. Al-Ramadan: https://orcid.org/0009-0003-5383-464X
Tool Link: https://github.com/aalramadan/SubGPT

References
[1] Abu R, H., & Haider, A. (2022). Construction of a Parallel Corpus of English-Arabic Movie Subtitles: A Genuine Source for Audiovisual Translators. *The International Journal of Humanities Education*, *21*(1), 21–37. https://doi.org/10.18848/2327-0063/CGP/v21i01/21-37
[2] Artegiani, I., & Kapsaskis, D. (2014). Template files: Asset or anathema? A qualitative analysis of the subtitles of *The Sopranos*. *Perspectives*, *22*(3), 419–436. https://doi.org/10.1080/0907676X.2013.833642
[3] Athanasiadi, R. (2017). Exploring the potential of machine translation and other language assistive tools in subtitling: A new era? In *Audiovisual Translation: Research and Use* (29–49). Peter Lang Bern. https://library.oapen.org/bitstream/handle/20.500.12657/50109/9783631776407.pdf?sequence=1#page=31
[4] Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Proceedings of the Acl Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65–72. https://aclanthology.org/W05-0909.pdf
[5] Bielsa, E., & Bassnett, S. (2008). *Translation in global news*. Routledge. https://books.google.com/books?hl=en&lr=&id=XL59AgAAQBAJ&oi=fnd&pg=PP1&dq=Bassnett,+S.+(2014).+Translation+in+Global+News.+Routledge.&ots=0oO_t7d3k6&sig=E1icRgMeOF6pYjEnrJBOf0EfP5c

[6]   Bolaños-García-Escribano, A., & Díaz-Cintas, J. (2020). The Cloud Turn in Audiovisual Translation. In Ł. Bogucki & M. Deckert (Eds.), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (519–544). Springer International Publishing. https://doi.org/10.1007/978-3-030-42105-2_26

[7]   Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual Translation: Subtitling. Manchester: St*. Jerome publishing.

[8]   Gambier, Y. (2018). *Reception Studies and Audiovisual Translation*. 1–365.

[9]   Gambier, Y., & Di-Giovanni, E. (2018). Reception studies and audiovisual translation. *Reception Studies and Audiovisual Translation*, 1–365.

[10]   Gao, Y., Wang, R., & Hou, F. (2023). *How to Design Translation Prompts for ChatGPT: An Empirical Study* (arXiv:2304.02182). arXiv. http://arxiv.org/abs/2304.02182

[11]   García-Escribano, A. B. (2023). Editing machine-generated subtitle templates: A situated subtitler training experience. *Cadernos de Tradução*, *43*, e93050.

[12]   Georgakopoulou, P. (2006). Subtitling and globalisation. *The Journal of Specialised Translation*, *6*, 115–120.

[13]   Georgakopoulou, P. (2019). Template Files: The Holy Grail of Subtitling. *Journal of Audiovisual Translation*, *2*(2), 137–160.

[14]   Gu, W. (2023). *Linguistically Informed ChatGPT Prompts to Enhance Japanese-Chinese Machine Translation: A Case Study on Attributive Clauses* (arXiv:2303.15587). arXiv. http://arxiv.org/abs/2303.15587

[15]   Han, L., Jones, G. J. F., & Smeaton, A. F. (2021). *Translation Quality Assessment: A Brief Survey on Manual and Automatic Methods* (arXiv:2105.03311). arXiv. http://arxiv.org/abs/2105.03311

[16]   Hepburn, T. (2021, January 8). Global Streaming Subs Are Projected to Hit 2 Billion by 2025. *Cord Cutters News*. https://cordcuttersnews.com/global-streaming-subs-are-projected-to-hit-2-billion-by-2025/

[17]   Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). *Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine* (arXiv:2301.08745). arXiv. http://arxiv.org/abs/2301.08745

[18]   Kheiri, K., & Karimi, H. (2023). *SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning* (arXiv:2307.10234). arXiv. http://arxiv.org/abs/2307.10234

[19]   Krejtz, I., Szarkowska, A., & Krejtz, K. (2013). *The effects of shot changes on eye movements in subtitling*. https://depot.ceon.pl/handle/123456789/9102

[20]   Lin, P., & Xie, C. (2020). Translation Quality Assessment: Past and Present: Juliane House London: Routledge, 2015, x+160 pp. *Australian Journal of Linguistics*, *40*(1), 130–132. https://doi.org/10.1080/07268602.2018.1562856

[21]   Linde, Z. de, & Kay, N. (1999). *The semiotics of subtitling. Manchester: St*. Jerome Publishing.

[22]   Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. http://arxiv.org/abs/1301.3781

[23]   Moslem, Y., Haque, R., Kelleher, J. D., & Way, A. (2023). *Adaptive Machine Translation with Large Language Models* (arXiv:2301.13294). arXiv. http://arxiv.org/abs/2301.13294

[24]   Nikolić, K. (2015). The Pros and Cons of Using Templates in Subtitling. In R. B. Piñero & J. D. Cintas (Eds.), *Audiovisual Translation in a Global Context* (pp. 192–202). Palgrave Macmillan UK. https://doi.org/10.1057/9781137552891_11

[25]   O'hagan, M. (2005). Multidimensional translation: A game plan for audiovisual translation in the age of GILT. *Challenges of Multidimensional Translation*, 76–87.

[26]   Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. https://aclanthology.org/P02-1040.pdf

[27]   Pedersen, J. (2007). Cultural Interchangeability: The Effects of Substituting Cultural References in Subtitling. *Perspectives*, *15*(1), 30–48. https://doi.org/10.2167/pst003.0

[28]   Pym, A. (1992). Translation error analysis and the interface with language teaching. *The Teaching of Translation*, 279–288.

[29]   Saadany, H., & Orasan, C. (2021). BLEU, METEOR, BERTScore: Evaluation of Metrics Performance in Assessing Critical Translation Errors in Sentiment-oriented Text. *Proceedings of the Translation and Interpreting Technology Online Conference TRITON 2021*, 48–56. https://doi.org/10.26615/978-954-452-071-7_006

[30]   Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PloS One*, *13*(6), e0199331.

[31]   Yamada, M. (2023). *Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability*. https://arxiv.org/ftp/arxiv/papers/2308/2308.01391.pdf