© June 2026 | IJIRT | Volume 13 Issue 1 | ISSN: 2349-6002

Dialogue Generation for Dramas from Transcripts Using Large Language Models

Shreeya Daga[1], Shraddha Gangurde[2], Samali Rajderkar[3], Mangesh Gosavi[4], Prof. Vishal Jaiswal[5]
[1,2,3,4,5]Department of Information Technology, Pune Institute of Computer Technology, Pune, India

***Abstract*—High-volume, long-running daily soap television series pose an especially difficult task for large language models (LLMs). Daily soaps require an LLM to generate production-ready dialogue continuously, with the added constraint of maintaining strict episodic continuity and deep consistency in character behaviors throughout multiple episodes. In this survey, we will provide a cohesive framework for using LLMs to coauthor serialized television. We will also examine current advances in LLMs for creative writing for our television framework. The main focus of the survey will be on the technical challenges faced by LLMs when creating scripts for serialized television, including generating large amounts of dialogue (10-15 pages per episode), maintaining coherent narrative structure through hundreds of connected scenes, and faithful representation of each character's voice during multi-character dialogue scenes. Three essential elements for creating scalable scripted dialog have been identified in our research. First, hierarchical context management for the maintenance of long-term story arc; second, multiagent orchestration for the incorporation of production metadata such as music, sfx, and camera cues; third, human-in-the-loop editing environments that support both non-destructive revision and version control. Our results indicate a paradigmatic shift from one-shot prompt engineering to a modular, context-aware architecture that treats character profiles, narrative history, and technical cues as structured input representations. The major implication of our research is that fine-grained LLM orchestration (i.e., combining long-context modelling, adaptive memory, and collaborative interfaces) is necessary for developing these models from experimental script generators to reliable, high throughput creative partners capable of supporting the demands of character-driven television storytelling on a perpetual basis.**

***Index Terms*— Dialogue Generation, Large Language Models, Narrative Continuity, Serialized Television, Script Generation,**

I. INTRODUCTION

There is no other industry (in terms of creating large amounts of content) than television serials. Unlike film/television episodic series, where there are breaks in seasons for both writers and producers, and feature films, which have long development periods (and longer production times), television soap operas create a tremendous amount of material to produce on an almost constant schedule. Soap operas usually require a demanding writing schedule. Writers create about 5 to 6 halfhour-long episodes each week, thus resulting in nearly 250 to 300 episodes per year, in which episode focuses on dialogueheavy scripts that are typically 10 to 15 pages long. The stories often revolve around a large cast of characters. Their relationships and storylines evolve and connect over many episodes, sometimes lasting months or even multiple years. Therefore, this extreme volume of content production places significant restrictions on television soap writing teams: they will burn out; they may produce varying degrees of quality; and most importantly, maintain character voice and story line continuity through every episode produced, in some cases, over 250-300 episodes.

The rise of large language models (LLMs) like GPT-3, GPT4, and Claude, etc, has transformed how we think about creative writing due to the ease and convenience provided by them [1], [2]. These models can generate very coherent text, follow context in conversations [3], [4], and change their writing style through fine-tuning or prompt design. This makes them powerful tools for writers and storytellers [1], [2]. Researchers have been able to apply LLMs to create short stories, one-turn dialog systems [5], and one-act plays [6], yet they have not addressed the unique demands of creating daily soap content, which presents a combination of technical challenges that current systems are unable to meet.

Compared to other forms of narrative creation, daily soap scriptwriting is vastly different in five critical areas:

- Scale and Continuity: The writers are required to write between 3000–4500 pages of script per year while ensuring perfect continuity, where a character's offhand comment in episode 47 may become pivotal to the plot in episode 89. All current LLMs, due to their constraint of a context window of 2000–128,000 tokens, experience difficulty maintaining consistency even within a single episode (of approximately ten pages), and certainly in tracking the many narrative lines that unfold over multiple episodes, weeks, or months [7].
- Character Consistency at Scale: A daily soap typically features 10–20 recurring characters, each of whom must have unique and distinctive ways of speaking, express a wide range of emotions, and exhibit various quirks of behavior and thought patterns. Maintaining such diversity across thousands of lines of dialogue far exceeds what static character profiles can achieve [3], [8].
- Production Metadata: In addition to dialogue, a script must also include camera directions, sound effects, musical cues [9], and temporal annotations that transform textual content into production-ready documentation [10], [11].
- Cultural and Regional References: Daily soaps, particularly in entertainment industries, are deeply embedded with cultural elements such as joint family structures, religious celebrations, and linguistic codeswitching, which western-trained models often fail to reproduce accurately [12], [13].
- Human-in-the-Loop Requirements: Showrunners require sophisticated collaborative interfaces that enable creative control through version management, selective regeneration, and non-destructive editing workflows [1], [2], [6].

The results of this survey make three key contributions to both researchers and practitioners:

- A Systematic Review of a Comprehensive Nature: This study is the first to systematically evaluate the application of Large Language Models (LLMs) for scriptwriting through the lens of large-scale, serialized episode production. It also examines the convergence of issues associated with producing each episode of a daily soap, addressing challenges in continuity, coherence, and narrative scalability.
- A Critical Evaluation Methodology: This work provides a comparative evaluation of existing systems based on the production requirements of daily soap episodes, rather than merely cataloging current tools. The proposed framework outlines a roadmap for future progress through comparative analysis tables, gap identification, and architecture blueprints.
- An Interdisciplinary Linkage: This survey synthesizes research across multiple domains, Natural Language Generation, Human-Computer Interaction [1], [2], Computational Creativity [6], and Cultural AI [12], demonstrating how applying LLMs to daily soap script generation reveals both the limitations and opportunities of current LLM technologies.

II. LITERATURE REVIEW AND COMPARATIVE

Analysis

A. Filmagent:A multi-Agent framework for end-to-enf filmautomation in virtual 3D spaces [10], [11]

The paper introduces FILMAGENT, a multi-agent system powered by large language models designed to automate the entire film production process in 3D virtual spaces. Each agent takes on roles such as director, screenwriter, actor, or cinematographer, collaboratively generating ideas, scripts, and camera setups. Using strategies like Critique-Correct-Verify and Debate-Judge, the system refines creative outputs and reduces hallucinations. Human evaluations rated FILMAGENT's videos highly for coherence and quality.

B. Dialogue Generation Conditional on Predefined Stories:Preliminary Results [14]

This paper aims to enhance dialogue generation by shaping conversation according to a scripted storyline. Rather than producing dongles or nonsensical responses, the model supplies conversation in line with the whole narrative or subject. It takes a two-stage approach, essentially predicting storyrelevant keywords and then generating dialogue about them, to promote better continuity in the narrative and characters. Experiments demonstrate that this approach leads to dialogues that are more natural and better connected to the story.

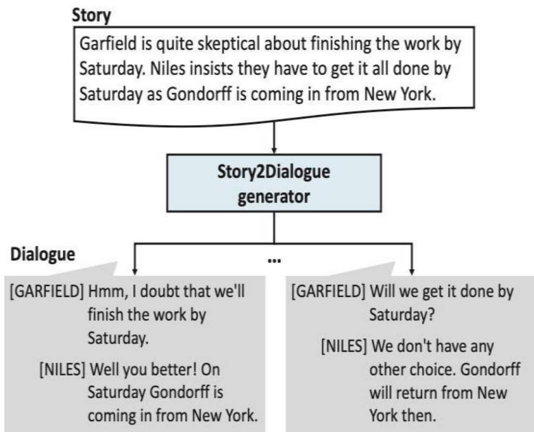

Fig. 1. Schematic diagram of Story2Dialogue (reprinted from [14], under the Creative Commons CC BY 4.0 license).

C. TEMDG: A Multi-Round Dialogue Generation Model Incorporating Topic and Historical Information [4]

This paper introduces TEMDG, a dialogue generation model that improves conversation coherence by combining topic understanding with historical dialogue context. It uses the Biterm Topic Model (BTM) to extract underlying topics and applies a hierarchical attention mechanism to encode long conversational histories. By fusing topic and context information through a multi-attention decoding process, TEMDG produces responses that are more natural, contextually consistent, and diverse. Experimental results on the DailyDialog dataset show significant improvements in topic relevance and fluency.

D. Co-Writing Screenplays and Theatre Scripts with Lan-guage Models: Evaluation by Industry Professionals [6]

This paper investigates the potential contribution of Large Language Models (LLMs) to be used as a co-writer for screenwriting, providing linguistic buffs with both story building and dialogue generation support. The paper investigates the creative balance of user control and AI contribution, through human–AI collaboration experiments.

Results suggest that LLMs are effective at fostering creativity and productivity but struggle with maintaining character coherence over consecutive chapters or across long-arc plot development.

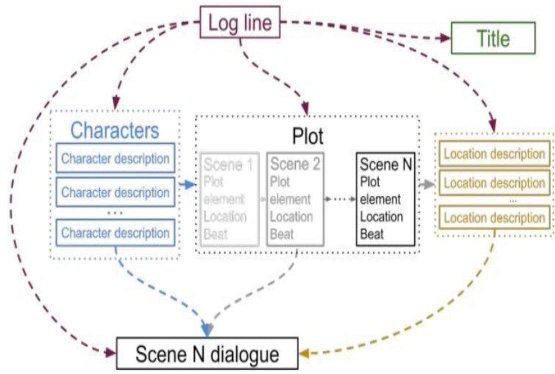

Fig. 2. Dramatron's Hierarchical Coherent Story Gener- ation. Dramatron starts from a log line to generate a title and characters. Characters generated are used as prompts to generate a sequence of scene summaries in the plot. Descriptions are subsequently generated for each unique location. Finally, these elements are all combined to generate dialogue for each scene. The arrows in the figure indicate how text generated is used to construct prompts for further LLM text generation. (reprinted from [6], under the Creative Commons CC BY 4.0 license).

E. Enhancing Multi-Person Dialogue with Large LanguageModels: A Structured Approach to Natural Communication [8]

This paper presents a clear framework aimed at making multi-person dialogue generation with Large Language Models (LLMs) more realistic and engaging. It seeks to make group conversations sound smoother and more natural by teaching the model to handle different speaker roles, manage when each person speaks, and understand the context of the discussion. To maintain the flow of conversation and prevent unnecessary repetition, the system uses organized prompts and a clever method for structuring context. Overall, experiments demonstrate that this approach helps produce dialogues that feel more coherent, lifelike, and logically consistent.

F. Dialogue Rewriting via Skeleton-Guided Generation [15]

This paper proposes a new way to rewrite dialogues using a skeleton-guided generation approach. The idea is to first extract a simplified "skeleton" from the original dialogue, capturing only the key intent and structure. Then, it expands this skeleton into a more fluent, complete response. This method helps reduce redundancy and keeps the conversation natural and coherent. Experiments show it performs better than standard rewriting models in preserving meaning while improving readability.

G. ScriptWriter: Narrative-Guided Script Generation [16],[17]

This paper presents ScriptWriter, a model designed to generate scripts that follow a clear and logical storyline. Instead of producing isolated dialogues or scenes, ScriptWriter focuses on maintaining narrative consistency, guiding each scene to connect smoothly with the next. It uses a narrative planning module to structure events before writing the actual script, which helps preserve character motivation and plot flow. Experimental results show noticeable improvements in coherence and creativity.

H. L3Cube-HindBERT and DevBERT: Pre-Trained BERT Transformer models for Devanagari based Hindi and Marathi Languages [12]

This paper introduces HindBERT and DevBERT, transformer-based language models trained specifically for Hindi and Marathi, which both use the Devanagari script. The authors found that existing multilingual BERT models underperform on these languages, so they trained monolingual and dual-language models using large Hindi and Marathi corpora. These models achieved better accuracy in text classification and named entity recognition tasks compared to multilingual ones. The paper also extends this work to other Indian languages.

III. METHODOLOGIES AND PREVALENT TECHNOLOGIES

Research in dialogue and script generation has advanced toward complex, context-sensitive, and collaborative methods. The current work shows a mix of multi-agent coordination, long-context modeling, narrative control, and human-AI teamwork. This section summarizes these methods and the technologies that support them.

A. Multi-Agent and Role-Oriented Architectures [10], [11],[20]

Recent methods focus on breaking tasks into parts using multi-agent systems, where different language model instances take on specific roles like director, writer, or editor [10], [11]. Instead of depending on one generative process, these systems use coordination protocols, like critique-correct or debatejudge cycles, to improve creative results and ensure logical flow. This agent-based setup allows for workflow automation, divides creative tasks, and offers a scalable way to produce scripts at a high level. Multi-agent systems have become a key approach for complete creative pipelines. Common technologies: agent orchestration layers, message-passing protocols, role conditioned LLMs, and task scheduling interfaces.

B. Contextual and Memory-Aware Dialogue Generation [4],[7]

Another key trend is the use of memory systems that help maintain dialogue continuity over longer conversations. Current systems model short-, medium-, and long-term conversation links using multi-level attention networks or hierarchical encoders. Plus, topic modeling techniques, such as biterm and latent topic models, help keep thematic unity throughout dialogue turns or episodes [4]. This method allows LLMs to maintain the flow of the narrative and the consistency

Table I Comparative Summary of Methodologies

Ref.	Core Technique(s)	Primary Goal / Contribution	Stated Limitation / Con
[10], [11]	Multi-Agent LLM Systems	Designed to divide complex dialogue generation tasks among several specialized agents that work together under shared communication rules. This approach allows coordination of writing, planning, and design within one automated framework.	Demands significant computational power and precise synchronization among agents; combining their outputs can be challenging.

[14], [16], [17]	Narrative/Story-Conditioned Generation	Focuses on producing dialogue that fits naturally within an existing storyline or plot structure, ensuring that scenes progress logically and characters remain consistent throughout.	Depends heavily on the quality of the story outline; restricts creative variation and may mirror the source material too closely.
[15]	Structure/Template-Based Generation	Builds dialogue through predefined outlines or templates, giving developers more control over tone, pacing, and structure while keeping formatting uniform.	Can feel repetitive or predictable; creating effective templates takes effort and limits spontaneous expression.
[1], [2], [6]	Human–AI Co-Creative Systems	Encourages a back-and-forth process between human writers and language models, where the system refines ideas, suggests phrasing, and assists in creative exploration.	Collaboration requires time and user skill; results depend strongly on human input and may lack consistency.
[4], [8]	Context-Aware Multi-Turn Dialogue	Aims to maintain flow and relevance in long conversations by remembering prior exchanges and tracking topics across multiple turns.	Limited by model memory and processing range; performance drops as context grows in length or complexity.
[3], [9], [18], [19]	Modality-Specific Generation	Extends dialogue generation into other forms such as speech or emotionally expressive text, aligning rhythm, tone, and mood for a more natural feel.	Requires extensive data and computation; evaluating emotional or acoustic quality remains subjective.
[12], [13]	Language-Specific Foundation Models	Trains models directly on regional languages and scripts to capture local expressions, grammar patterns, and cultural context more accurately.	Less effective outside the target language; high training costs and data demands limit scalability.
[3], [20]	Transformation and Adaptation Models	Converts dialogue between formats or tones—such as written to spoken or informal to formal—allowing reuse across different media and styles.	May alter meaning or lose nuance; maintaining tone and quality during conversion is difficult.

Note – LLM = Large Language Model; Prosody = Speech rhythm and intonation; Cross-modal = Text-to-speech or text-to-music generation; HindBERT/DevBERT = Indian language–specific Transformer models.of characters in long forms like serialized television scripts. Common technologies: hierarchical attention models, context fusion encoders, topic modeling modules, and memory networks.

C. Narrative-Conditioned and Structure-Guided Generation[14]–[17]

A notable group of methods aims to keep generated dialogues in line with set storylines or narrative frameworks. In this case, the generative process follows story graphs, plot outlines, or keyword cues to ensure that dialogues match the narrative progression and emotional tempo [14], [16], [17]. This often uses two-stage generation processes—first pulling out narrative hints, then expanding those into character dialogues. Structureguided methods, like skeleton-based generation [15], improve control by allowing for partial rewrites or regenerations without disrupting the overall narrative flow [15]. Common technologies: encoder-decoder transformers (BART, T5), dependency-based narrative extractors, scaffoldguided decoding mechanisms.

D. Human-AI Co-Creative Frameworks [1], [2], [6]

Current methods also include active human participation in the generation process. Instead of completely automated systems, these allow for feedback loops where writers can guide, critique, or adjust outputs [2], [6]. The aim is to balance creative freedom with AI support, encouraging teamwork that resembles professional writing environments [1]. These systems usually depend on prompt conditioning, real-time feedback tools, and reinforcement learning with human assessment to keep creative intentions while boosting productivity. Common technologies: interactive co-writing platforms, RLHF models, controlled text generation frameworks, and adaptive prompt optimization.

E. Cross-Modal and Multilingual Methodologies [9], [12],[13], [18], [19]

As research grows in diverse cultural and linguistic settings, methods are increasingly focused on multimodal generation and adapting to local languages. Cross-modal systems combine text generation with music, audio, or visual data, using memory", enabling the model to store, recall, and reason about sentiment conditioning or scene-level embeddings to align events, decisions, and relationships spanning the entire series multiple creative elements [9], [18], [19]. Additionally, [7]. Tackling this "forgetting problem" is fundamental to language- achieving real narrative continuity in extended creative

Table II Research Gap Analysis by Methodology

Gap Category	Affected Methodologies	Specific Issues	Potential Solutions
Long-term coherence	Multi-agent, Persona, Conditioning	Narrative drift over extended content	Hierarchical planning, explicit memory modules, consistency verification layers
Evaluation standardization	All methodologies	Subjective quality assessment, lack of unified metrics	Domain expert panels, multidimensional rubrics, automatedhuman hybrid evaluation
Computational efficiency	Multi-agent, Domain-specific	Resource-intensive training and inference	Model distillation, efficient architectures, caching strategies
Cross-modal consistency	Multimodal integration	Synchronization between text, speech, visual	Unified representation spaces, joint training objectives
Cultural adaptation	Language localization	Beyond linguistic to cultural storytelling patterns	Culture-aware training data, regional narrative structure modeling
Professional integration	All methodologies	Workflow compatibility, learning curves	User-centered design, incremental adoption pathways, training programs

Note – This table summarizes methodological research gaps and their potential mitigation strategies for LLM-based dialogue generation and co-creative systems.specific transformer models made for Devanagari scripts and other regional languages support culturally relevant creative outputs [12], [13]. These developments help create contextually genuine content, especially for regional entertainment industries like Hindi and Marathi television. Common technologies: multilingual transformers (BERT, ALBERT adaptations), cross-modal encoders, sentiment-conditioned embedding layers, and speech synthesis models.

IV. RESEARCH GAPS ANALYSIS

Despite rapid advancements in the creativity of LLMs [1], [2], [6], there remain several challenges that need to be addressed before these LLMs can match the demands of television production. This section outlines some urgent research gaps, and highlights promising avenues for future investigation, for anyone aiming to build truly co-creative AI for long-form narrative contexts.

A. Persistent Episodic Memory and Narrative Planning

One of the most persistent obligations for today's LLM-based story generators is memory: while models are increasingly adept at producing individual scenes or short episodes [14], they often lose the continuity when tasked with maintaining storylines that unfold across dozens or even hundreds of episodes [6]. Critical details slip through the cracks, character backstories are forgotten, plotlines contradict themselves, and long-term arcs unravel. Relying on ever-larger context windows is not enough. What's needed are novel memory architectures, systems that function almost like an "episodic collaborations.

B. Authentic Cultural Grounding

Although there are efforts in order to make AI creative tools more inclusive [12], [13], most models still show the biases of Western-focused training data. This often leads to scripts that don't satisfy non-Western storytelling traditions' unique rhythms, norms, and emotional depth [12]. Genuine cultural understanding requires sensitivity towards how communities' express humor, handle social dynamics, and share meaning through dialogue. Future systems must focus on cultural adjustments [13]. This is essential for capturing the richness of the art of storytelling, instead of limiting creative diversity to a single global standard.

C. Unified Multimodal Synthesis and Continuity Tracking

Professional scriptwriting requires more than just captivating conversation [3]. What's needed are systems that can generate dialogue along with producing notes, blending lines with camera directions, soundtrack suggestions according to mood [9], and scene-setting cues [10], [11]. Yet, a significant gap persists in maintaining continuity: there is still no such mechanism for automatically checking that characters are logically consistent, or that visual details and narrative timelines align across episodes. In the film and TV industries, human continuity supervisors vigilantly track these aspects; automating such checks could be revolutionary, freeing writers and producers to focus on creativity. The development of intelligent virtual continuity tools, ones that flag inconsistencies and safeguard story coherence, remains an open and exciting challenge [10], [11].

D. Fine-Grained Controllability and Workflow Integration Most current AI writing assistants remain far too inflexible for professional work [1], [2], [6]. They're often perfectionist at brainstorming, but fall short and lack expertise when it comes to allowing detailed, incremental revisions to specific minor details. Adjusting the tone of a scene, modifying character emotions, or refining plot pacing typically requires rewriting prompts from scratch. To serve real creative teams, future systems must support undoable edits, collaborative version control, and non-linear workflows, allowing writers to experiment freely, explore multiple directions, and adjust narrative elements without losing prior work and continuity of the prompt [1], [2], [6]. The goal is not just AI as an "assistant," but as a responsive co-author, seamlessly fitting into established creative processes rather than dictating rigid use patterns.

V. CONCLUSION

This survey has navigated the rapidly evolving landscape of dialogue and script generation, a domain energized by the capabilities of Large Language Models (LLMs). The collective body of work reviewed spanning more than twenty recent papers, from foundational models to complex multi-agent systems underscores a clear trajectory: the shift from merely generating coherent text to engineering contextually rich and dramatically compelling narratives.

Our review highlights key areas of intense focus. The challenge of maintaining narrative control and long-range coherence has led to structured approaches like narrative guided generation and the exploration of interactive drama solutions. Simultaneously, the cinematic production pipeline is being revolutionized by multi-agent frameworks, such as FilmAgent, which automate everything from scriptwriting to cinematography. The papers also reveal a necessary push beyond text, with frameworks like Speech Dialogue Factory tackling the critical leap to natural, human-like spoken dialogue, complete with paralinguistic elements.

Yet, crucial challenges persist. The issue of reliability in multi-turn conversations remains a significant concern, as evidenced by studies showing LLMs can" get lost" in context degradation. Furthermore, a parallel effort is evident in democratizing access through models tailored for low-resource languages, such as DevBERT and Paramanu. While human-AI co-creation studies confirm LLMs' role as a powerful" second mind" for creative tasks, future research must continue to bridge the gap between technical fluency and genuine dramatic quality, leveraging emotional conditioning and robust topic maintenance to achieve truly compelling, human-level storytelling. The work on conditional dialogue generation will prove vital in ensuring that technical rigor meets artistic narrative intent.

REFERENCES

[1] T. Chakrabarty *et al.*, "Creativity support in the age of large language models: An empirical study involving emerging writers," in *Proc. Creativity and Cognition (C&C '24)*, Chicago, IL, USA, Jun. 2024.

[2] Q. Wan *et al.*, "It felt like having a second mind: Investigating human-AI co-creativity in prewriting with large language models," *arXiv preprint* arXiv:2307.10811, 2024.

[3] K. Mitsui, Y. Hono, and K. Sawada, "Towards human-like spoken dialogue generation

between AI agents from written dialogue," *arXiv preprint* arXiv:2310.01088, 2023.

[4] L. Yin *et al*., "TEMDG: A multi-round dialogue generation model incorporating topic and historical information," *IEEE Access*, vol. 11, pp. 1–13, 2023, doi: 10.1109/ACCESS.2024.3492009.

[5] L. Cai and S. Ji, "An efficient policy gradient method for conditional dialogue generation," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Beijing, China, 2019, pp. 209–217.

[6] P. W. Mirowski, K. W. Mathewson, R. Evans, and J. Pittman, "Cowriting screenplays and theatre scripts with language models: Evaluation by industry professionals," in *Proc. ACM CHI Conf. Human Factors Comput. Syst. (CHI '23)*, Hamburg, Germany, Apr. 2023.

[7] P. Laban, H. Hayashi, J. Neville, and Y. Zhou, "LLMs Get Lost in Multi-Turn Conversation," *arXiv preprint* arXiv:2505.06120, 2025.

[8] T. Murogaki and T. Nishimura, "Enhancing multi-person dialogue with large language models: A structured approach to natural communication," in *Proc. 8th Int. Conf. Interactive Digital Storytelling*, 2024, pp. 1–14, doi: 10.1145/3663384.3663398.

[9] V. Veerendranath *et al*., "ScripTONES: Sentiment-conditioned music generation for movie scripts," in *Proc. ACM Conf.*, 2023.

[10] Z. Xu *et al*., "FilmAgent: Automating virtual film production through a multi-agent collaborative framework," *arXiv preprint* arXiv:2402.01680, 2024.

[11] Z. Li *et al*., "FILMAGENT: A multi-agent framework for end-to-end film automation in virtual 3D spaces," *arXiv preprint* arXiv:2501.12909, 2025.

[12] R. Joshi, "L3Cube-HindBERT and DevBERT: Pre-trained BERT transformer models for Devanagari-based Hindi and Marathi languages," *arXiv preprint* arXiv:2211.11418, 2023.

[13] M. Niyogi, "Paramanu: A family of novel efficient generative foundation language models for Indian languages," unpublished.

[14] C. Miyazaki, "Dialogue generation conditional on predefined stories: Preliminary results," *IEEE Access*, vol. 11, pp. 1–11, 2023, doi: 10.1109/ACCESS.2023.3303440.

[15] C. Xin *et al*., "Dialogue rewriting via skeleton-guided generation," in *Proc. 37th AAAI Conf. Artif. Intell. (AAAI-23)*, Washington, DC, USA, 2023.

[16] Y. Zhu *et al*., "Leveraging narrative to generate movie script," unpublished.

[17] Y. Zhu *et al*., "ScriptWriter: Narrative-guided script generation," *arXiv preprint* arXiv:2005.10331, 2020.

[18] M. Wang *et al*., "Speech DialogueFactory: A framework for natural speech dialogue generation," in *Proc. Interspeech 2025*, Rotterdam, The Netherlands, Aug. 2025.

[19] M. Wang *et al*., "SpeechDialogueFactory: Generating high-quality speech dialogue data to accelerate your speech-LLM development," *arXiv preprint* arXiv:2503.23848, 2025.

[20] W. Wu *et al*., "From role-play to drama-interaction: An LLM solution," *arXiv preprint* arXiv:2405.14231, 2024.

[21] S. R. Cox and A. Abdul, "Prompting a large language model to generate diverse motivational messages: A comparison with human-written messages," in *Proc. ACM Conf.*, 2023.