

Agentic Framework for Culturally Grounded Visual Story Generation

Anonymous submission

Abstract

Story-based learning is an effective approach for communicating complex concepts, particularly when narratives reflect learners' cultural backgrounds and include visual components. While large language models (LLMs) and text-to-image systems have made automated story generation more accessible, existing methods often struggle with cultural grounding, long-form coherence, and character consistency. In this work, we present an agentic, user-in-the-loop framework for generating culturally aligned visual stories tailored to communities in the Global South. Our system employs a set of coordinated LLM powered agents to construct curriculum guided narratives, develop culturally grounded characters, draft coherent story arcs, and produce visually consistent illustrations. This approach provides fine grained human-AI collaboration, enabling educators, and community facilitators to co-design stories that accurately reflect their learners' lived realities.

Code —

<https://anonymous.4open.science/r/IUI-public-3FC5>

Introduction

Visual storytelling is a powerful medium for education and community engagement. Prior work highlights that learning materials are most effective when they reflect culturally familiar settings, practices, and relationships (Phoebe Eilers and Torres 2021; Ezeife 2006; Bhat et al. 2024). However, current LLMs and diffusion model based storytelling systems often generate generic plots, inconsistent characters, or culturally inaccurate depictions (Migal et al. 2024; Hamna et al. 2025; Khanuja et al. 2024). These limitations disproportionately affect communities in the Global South, where rich cultural variation demands more nuanced and context-aware story generation. Recent developments in agentic frameworks have shown promise for improving long-range coherence and modular control in creative generation workflows (Li et al. 2024; Xia et al. 2025). Building on these advances, we introduce a culturally sensitive, multi-agent narrative framework designed to support community grounded story creation for learning contexts such as financial literacy and healthcare awareness. Our system uses coordinated role-driven LLM powered agents to construct curriculum guided narratives, develop culturally grounded

characters, draft coherent story arcs, and produce visually consistent illustrations. To support cultural accuracy and reduce stereotyping, users can iteratively refine story elements throughout the pipeline. For our preliminary user study in Kerala, India, we adopted an Amar Chitra Katha inspired visual style to match local familiarity. This work contributes a practical tool for culturally aware content creation in low-resource contexts and offers early evidence of its educational impact.

System Design

Our framework generates culturally aligned narratives through a modular architecture comprising two primary components: a text story generation module and a visual story generation module, orchestrated through a set of LLM driven agents as shown in figure 1. The system takes three key inputs: (1) source content specifying the learning goals or instructional topics, (2) demographic information of target learners, and (3) the geographic location that grounds the story. The text module uses a sequence of GPT-4o-powered agents (OpenAI 2024), while the visual module employs a hybrid strategy combines model fine-tuning FLUX.1 dev (Labs 2025) for maintaining a consistent art style with a reference-based workflow that ensures character fidelity.

Text Story Generation

- **Curriculum agent:** Translates high-level educational goals from the source documents into a structured five-module curriculum. It does automatic evaluations the draft for appropriateness for the target audience, reviews the progression of difficulty, then gathers user feedback to produce the final curriculum.
- **Narrative agent:** Using target-user and location data, this agent expands the cultural details incorporating Newmark's markers (Newmark 2010) and structures plot-lines using the Sabido methodology (Barker 2007; Nariman 1993). Using this cultural grounding and the final curriculum, it proposes five distinct narrative approaches, allowing users to choose themes aligned with community needs.
- **Character profile agent:** Generates culturally grounded character profiles covering appearance, relationships, traditions, occupations, and social roles. Users can edit

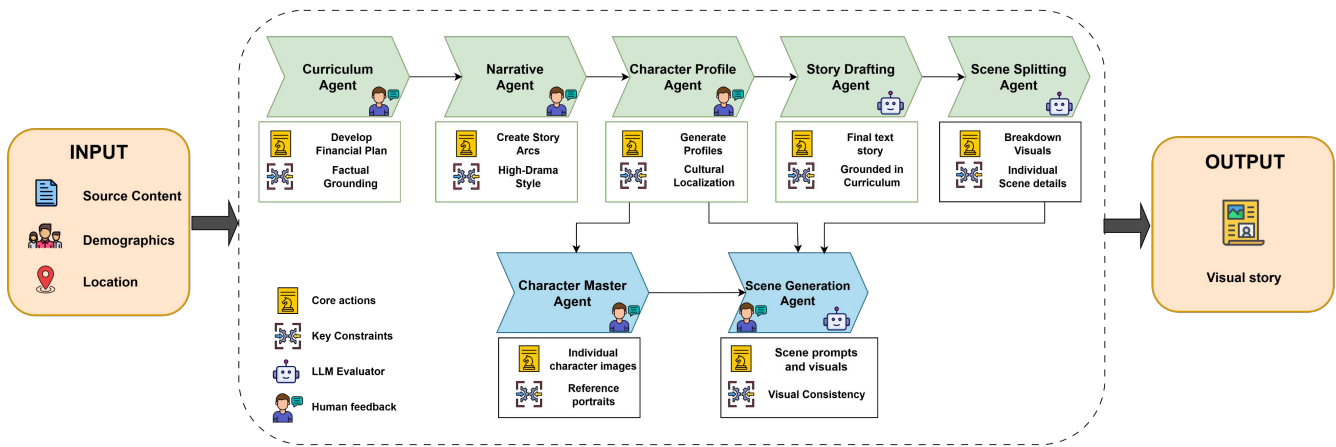


Figure 1: End-to-end architecture of our agentic, user-in-the-loop visual story generation framework. The green modules represent the Text Story Generation pipeline, which transforms the input (source content, demographics, location) into structured narratives and scene descriptions. The blue modules represent the Visual Story Generation pipeline, which converts these textual specifications into character images and scene-level visuals, producing the final visual story.

these details to avoid stereotyping and remains relatable to the intended audience.

- **Story Drafting agent:** Produces coherent long-form narratives around 5,000 words in total, aligned with the curriculum and chosen cultural context.
- **Scene Splitting Agent:** Segments the story into visual scenes with settings, characters, and key plot events based on the overall story arc, pacing, and logical progression of events. The number of scenes per module can be adjusted based on user preference.

Visual Story Generation

- **Character Master Agent:** Generates high-fidelity reference portraits for each character, defining their facial features, attire, and color palette for use throughout visual generation. Each portrait is reviewed and approved by the user before being finalized and stored.
- **Scene Generation Agent:** Creates final illustrations by combining reference portraits with textual prompts describing actions and settings. A vision-language evaluator checks for character consistency, cultural accuracy, and style coherence, refining outputs iteratively.

User Study

We conducted a small-scale evaluation with six adults from Kerala, India, focusing on financial literacy. Participants read a culturally grounded, multi-scene story generated by the system and completed pre- and post-tests measuring understanding of budgeting, saving, and interest. As shown in 2, participants demonstrated a 23% average improvement in post-test scores, with five of six showing gains. Participants also reported that story elements such as household scenes, clothing, and character relationships felt authentic, relatable, and aligned with local experience.

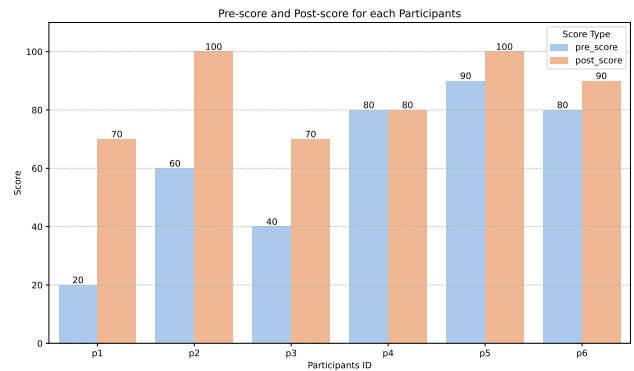


Figure 2: Score of participants in the user study

Conclusion

The framework directly contributes to Global South empowerment by enabling communities to create educational stories that reflect their cultural realities and learning needs. By lowering the barriers to producing accurate, resonant, and locally grounded narratives, it supports more equitable access to AI-driven content creation and strengthens community agency in shaping learning materials. In doing so, it offers a practical pathway for NGOs, educators, and grassroots practitioners to build contextualized, culturally meaningful educational resources that advance inclusive learning outcomes.

References

Barker, K. 2007. Sex, Soap and Social Change – The Sabido Methodology. *AIDSLINK*, (104).
 Bhat, A.; Kulkarni, N.; Husain, S.; Yadavalli, A.; Kaur, J. N.; Shukla, A.; Shelar, M.; and Seshadri, V. 2024. Speaking

in Terms of Money: Financial Knowledge Acquisition via Speech Data Generation. *ACM J. Comput. Sustain. Soc.*, 2(3).

Ezeife, A. N. 2006. Culture-sensitive Mathematics: The Walpole Island Experience.

Hamna; Sudharsan, D.; Seth, A.; Budhiraja, R.; Khullar, D.; Jain, V.; Bali, K.; Vashistha, A.; and Segal, S. 2025. Kahani: Culturally-Nuanced Visual Storytelling Tool for Non-Western Cultures. In *Proceedings of the 2025 ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies, COMPASS '25*, 379–400. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714849.

Khanuja, S.; Ramamoorthy, S.; Song, Y.; and Neubig, G. 2024. An image speaks a thousand words, but can everyone listen? On image transcreation for cultural relevance. arXiv:2404.01247.

Labs, B. F. 2025. FLUX.1-dev. <https://huggingface.co/black-forest-labs/FLUX.1-dev>. 12 billion parameter rectified-flow transformer for image generation.

Li, D.; Sohn, S. S.; Zhang, S.; Chang, C.-J.; and Kapadia, M. 2024. From Words to Worlds: Transforming One-line Prompts into Multi-modal Digital Stories with LLM Agents. In *Proceedings of the 17th ACM SIGGRAPH Conference on Motion, Interaction, and Games, MIG '24*. New York, NY, USA: Association for Computing Machinery. ISBN 9798400710902.

Migal, A.; Seredina, D.; Telnina, L.; Nazarov, N.; Kolmogorova, A.; and Mikhaylovskiy, N. 2024. Overview of Long Story Generation Challenge (LSGC) at INLG 2024. In Mille, S.; and Clinciu, M.-A., eds., *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, 47–53. Tokyo, Japan: Association for Computational Linguistics.

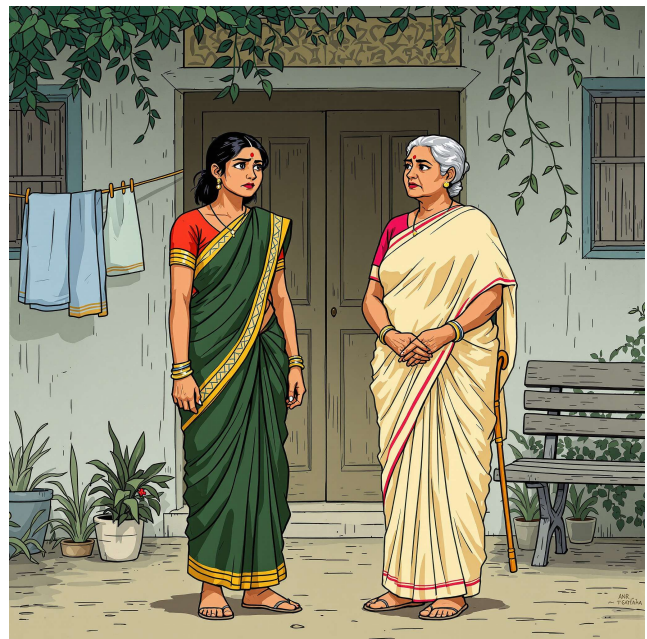
Nariman, H. 1993. *Soap Operas for Social Change: Toward a Methodology for Entertainment-Education Television*. Bloomsbury Academic.

Newmark, P. 2010. *Translation and culture*, 171–182. Peter Lang Verlag.

OpenAI. 2024. GPT-4o (“o” for “omni”) system card. <https://platform.openai.com/docs/models/gpt-4o>.

Phoebe Elers, M. J. D., Steve Elers; and Torres, R. 2021. Applying the Culture-Centered Approach to visual storytelling methods. *Review of Communication*, 21(1): 33–43.

Xia, H.; Peng, H.; Qi, Y.; Xu, B.; Li, J.; Lei, H.; and Wang, X. 2025. StoryWriter: A Multi-Agent Framework for Long Story Generation. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management, CIKM '25*, 6559–6563. New York, NY, USA: Association for Computing Machinery. ISBN 9798400720406.



One evening, as Asha was hanging clothes to dry, she shared her worries with her neighbour, Lakshmi Teacher, who worked at a government school. “Chechi, we are thinking of borrowing money from the local moneylender, but I’ve heard such terrible stories about him. I don’t know what else to do. We have no savings, and our expenses are piling up.” Lakshmi Teacher shook her head. “Asha, borrowing from a moneylender should be your last resort. They charge high interest rates, and you can easily fall into a cycle of debt. Instead, try to manage your expenses and save a little each day. Even if it’s just a few rupees, it adds up over time.”

Example story with Visuals



Taking this advice to heart, Asha started tracking her daily expenses. She switched to buying essentials from the Maveli Store instead of the expensive local shop. She also cut unnecessary expenses, like buying snacks for Raju every day. Instead, she cooked extra food at home and packed it for him. She walked to work instead of taking an auto-rickshaw and saved a small amount every evening.



One day, Asha visited Kalyani and saw her sister looking exhausted. "Chechi, we borrowed Rs. 50,000, and now we have to pay Rs. 70,000 because of interest! Krishnan says we need another Rs. 1,00,000 for his sister's wedding. I don't know what to do." Asha sighed and held Kalyani's hand. "Kalyani, you don't have to follow what others do. Spending lakhs on a wedding when you can barely afford rent is not wise. We need to separate needs from wants. Let's work out a budget together."