StoryGuard: Ethical Monitoring of AI-Generated Narratives for Children

1st Prakhar Langer
Dept. School of Computer Engineering
Manipal Institute of Technology
Bengaluru, India
prakhar.mitblr2022@learner.manipal.edu

2nd Gauri Kalnoor
Dept. School of Computer Engineering
Manipal Institute of Technology
Bengaluru, India
gauri.kalnoor@manipal.edu

3rd Vishnu Srinivasa Murthy Yarlagadda
Dept. School of Computer Engineering
Manipal Institute of Technology
Bengaluru, India
vishnu.murthy@manipal.edu

4th Shailaja A Akkur
Dept. of Management Studies
Nitte Meenakshi Institute of Tech
Bengaluru, India
shailaja.anilkumar@nmit.ac.in

5th Rajeshwari R R
Dept. name of organization
Dr. Ambedkar Institute of Technology
Bengaluru, India
rajeshwari.mba@drait.edu.in

6th Manish Motghare Jr.
Dept. of Research and Development
Nagpur University & Raisoni College of Engineering and Technology
Nagpur, India
motgharemanish@gmail.com

2025 11th International Conference on Education and Technology (ICET) | 979-8-3315-4548-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICET67257.2025.11290752

Abstract—**Making sure AI generated content is morally sound has become crucial as it appears more and more in entertainment and educational materials, especially for young audiences. The comprehensive pipeline for identifying and addressing ethical transgressions in AI generated children's stories, StoryGuard, is presented in this paper. With Gemini 1.5 Flash serving as the foundation for both creation and assessment, our method compares narratives to three fundamental criteria: safety, inclusion, and age appropriateness. Our framework provides a Retrieval-Augmented Generation (RAG) system to facilitate querying over a curated dataset of ethically compliant stories, as well as automated story rewriting to enforce ethical standards in addition to detecting violations. In addition to identifying objectionable content, the system helps create standards for the ethical application of AI in the classroom.**

Keywords—ai-generated content, automated story rewriting, ethical ai in education, ethical standards, gemini 1.5 flash, retrieval-augmented generation (RAG)

I. INTRODUCTION

The way machines comprehend and produce language that is similar to that of humans has been completely transformed by recent developments in Large Language Models (LLMs), such as OpenAI's GPT series [1], Google's Gemini [2], Meta's LLaMA [3], and Anthropic's Claude [4]. These models demonstrate a unprecedented capacity to generate text that is logical, rich in context, and human-like. They are trained on a range of online corpora and built with billions of parameters. At their core, LLMs use transformer-based architectures that use attention mechanisms to predict the next token in a series based on context. They can therefore write code, write summaries, write poetry, translate languages, and even answer questions. The probabilistic sampling methods (beam search, nucleus sampling, greedy decoding) that underpin LLMs' generative capabilities allow them to produce text that is both creative and unexpected while still appearing contextually relevant. Although these models are proficient in language, they are not naturally aware of safety, ethics, or context sensitivity, especially when it comes to material intended for vulnerable audiences like children.

LLMs can do a lot of things, like make up stories for fun and to learn. One can change the age, theme, emotion, reading level, and even the language of stories that AI makes for kids [5]. These tools can be used for a lot of things, like smart learning platforms, bedtime story generators, adaptive reading companions, and ways to help kids be creative in school. NovelAI, Storybird, and Read Along are just a few of the many platforms that are using AI more and more to make reading more fun and interesting for kids [6].

This new idea has both good and bad points [7]. Automation makes it possible to create stories on a large scale. However,it also raises serious ethical questions about the content that is being made. Stories that are culturally biased, violent, stereotypical, or linguistically inappropriate could harm children's psychological development, reinforce harmful norms, or exclude underrepresented groups. Furthermore, the absence of human supervision in large-scale content generation systems intensifies these concerns. In light of these issues, there is an urgent demand for mechanisms capable of identifying, evaluating, and addressing ethical breaches in AI-produced content for children.

Despite the importance of this issue, the majority of LLMs safety research focuses on toxicity filtering, adversarial prompting, or general bias detection without paying particular attention to narrative ethics for kids. As a solution, this work suggests StoryGuard–an end-to-end framework made to track and improve the morality of AI-generated children's stories. The system is built on three key pillars:

- Generation: The work produces a varied collection of children's stories using Gemini 1.5 Flash that are centered around well-known themes like friendship, kindness, courage, and honesty. Both morally acceptable and dubious content are purposefully included for the sake of research validity.
- Evaluation: Every story is evaluated according to three fundamental ethical standards which include age appropriateness (does the language and themes sound appropriate for children), inclusivity (avoiding stereotypes and cultural/gender bias), and safety (there should not be any presence of violence or harm).
- Intervention and Retrieval: To ensure compliance, AI generated stories that don't pass ethical checks and have low scores are automatically rewritten by the intervention of LLMs. The LangChain framework and FAISS are used to create a Retrieval-Augmented Generation system that enables users to monitor or query the carefully curated ethical story database in natural language.

The field of AI-driven storytelling has evolved rapidly with the emergence of large language models (LLMs), prompting a variety of explorations into creativity, user perception, ethical implications, and collaborative frameworks. At the core of these investigations is the question of whether AI can exhibit creativity and evoke genuine cognitive or emotional engagement. Reseach delves into readers' emotional and cognitive responses to ChatGPT-generated narratives, questioning the originality and perceived creativity of AI-authored stories [8]. In parallel, [6] evaluates the role of AI in narrative persuasion and transport, highlighting the importance of coherence and personalization in influencing the belief and engagement of the reader.

Several works expand this inquiry into more structured frameworks. Two research works both introduce frameworks that use LLMs to maintain narrative coherence, genre fidelity, and emotional flow across genres like fantasy and mystery [9, 10]. These works align with our goal of modular story control, especially during the rewriting and personalization stages. Similarly, [11] provides an empirical comparison of Human Vs. AI storytelling, revealing the nuanced creativity trade-offs and advantages of hybrid approaches, a concept that supports the human-in-the-loop component proposed in our system.

The role of modality and user collaboration is explored in works like [5] and [12], where multimodal and visual storytelling is introduced, enriching narratives beyond pure text. This is complemented by co-creative tools such as TaleBrush [13], Saga[14], and Wordcraft [15], which allow users to iteratively co-author stories with LLMs, balancing creative control and automation. These studies underscore the relevance of user agency and iterative feedback in enhancing engagement a key principle in our RAG-based QA interface.

On the ethical frontier, multiple works examine both the dangers and safeguards of AI-generated content. A few research works investigate the misuse of LLMs in misinformation generation, stressing the need for detection and moderation frameworks [16, 17]. Further, Doshi et al. warns about homogenization of content, calling for mechanisms to preserve diversity in AI outputs [18]. The proposed ethicality identification and rewriting modules are directly aligned with these concerns, aiming to filter out harmful narratives while maintaining semantic richness.

From a practical standpoint, educational applications of LLMs are gaining traction. There is a showcase of AI's potential in personalized learning and language education through storytelling [19, 20]. Likewise, [21] addresses child-centered design for visual storytelling tools, reinforcing the importance of age appropriate and cognitively-aligned narrative systems. These findings validate our emphasis on ethical story-telling tailored to children's developmental needs.

On the technological front, [22] provides a comprehensive survey on Retrieval-Augmented Generation (RAG) systems, which underpins our interactive QA module. Additionally, [23] introduces Fabula a narrative intelligence tool using RAG for structured and unstructured data a concept parallel to our use of RAG for interpretability in children's stories.

At the verdict, [24] offers a historical perspective by categorizing story generation methods into symbolic, statistical, and neural approaches, establishing the research trajectory that culminates in today's LLM-driven systems. Together, these works reveal an evolving landscape of AI storytelling marked by a convergence of creativity, ethics, personalization, and user interaction. Our work builds on this foundation by proposing an integrated pipeline that not only generates and ethically evaluates AI-authored stories but also allows users to interpret and interact with content in a safe, meaningful way.

The main contributions that this research makes are: a collection of carefully selected children's stories generated by AI with ethical considerations noted. An innovative loop of assessment and rewriting that uses the intervention of Gemini 1.5 Flash to identify and fix ethical errors. introduction of a question-answer-based RAG interface over the finished, ethically reviewed stories dataset, enabling interactive content analysis and user-led questioning. Proposal of story design guidelines for safe and inclusive AI narratives in educational settings. Additionally, we introduce two unique features not found in prior works: Automatic ethical rewriting of unsafe or biased stories. A searchable QA module allowing questions over ethically rewritten stories, serving as a tool for educators or curriculum developers. This paper is structured as follows: Section II reviews previous studies on the ethical application of LLM, toxicity and bias in generation, AI in education, and story assessment frameworks. Section III explains the entire procedure, which includes generating stories, evaluating them on specific metrics and generating scores accordingly, rewriting only the unethical stories, and developing RAG systems on the final dataset containing all the ethical stories. Section IV displays the results of the ethical assessments provides graphical.

II. Method

In this work, we propose a comprehensive multi-stage pipeline aimed at generating, evaluating, ethically correcting, and interactively explaining AI-generated stories for children. The methodology is divided into four distinct but, interlinked phases. Each phase is designed to address a specific objective while contributing to the overall system performance. The steps are modular, data-driven, and scalable, ensuring adaptability for diverse use cases.

A. Unfiltered Story Generation

The goal of this initial stage is to utilise an LLMs capabilities to generate children's stories. The model can generate a variety of results, including stories that might be offensive or dangerous, since there are no safety or ethical restrictions. The LLM predicts one token at a time while creating a story S in an autoregressive fashion given a natural language prompt P.

$$S = LLM(P) \tag{1}$$

$$P(S|P) = \prod_{t=1}^{T} P(w_t|\, w_{<t}, P) \tag{2}$$

Where w_t is the token generated at position t, and T is the total sequence length. This step establishes the baseline behavior of unconstrained language models in story generation.

B. Ethicality Identification Module

This stage categorizes whether a generated story is in compliance with the ethical standards using certain specific metrics like age appropriateness, bias, and safety.

- The ethical status of the story level is determined by aggregating the predictions of the sentence level.
- The story S is split into sentences $\{s_1, s_2, \ldots, s_n\}$.
- To assess the ethicality of the stories, each sentence of the story is run through a classification model f.

$$f(s_i) = y_i \in \{E, U\} \qquad (3)$$

Where, E indicates an ethical sentence and U an unethical one.

$$E^{th}score(S) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{f(s_i) = E\} \qquad (4)$$

A threshold $\theta \in [0,1]$ is applied. If $E^{th}\, Score(S) < \theta$ the story is flagged for ethical rewriting. For example, $\theta = 0.8$.

C. Ethical Rewriting Engine

To modify ethically flawed stories to align with ethical storytelling principles while preserving core elements like plot and tone:

- The story S is rewritten using a conditional LLM.
- A set of constraints C is applied, consisting of ethical guidelines.

$$\hat{S} = g(S|C) \qquad (5)$$

Where, $\hat{S}$ is the rewritten ethical story and g is the transformation function. For Optimizing this further,

$$\hat{S} = \underset{\tilde{S}}{\operatorname{argmax}}[\lambda_1 . E^{th}Score(\tilde{S}) + \lambda_2 . \mathrm{Sim}(S, \tilde{S})] \qquad (6)$$

- Sim denotes the semantic similarity (e.g., cosine similarity).
- λ_1 and λ_2 control the trade-off between ethical compliance and semantic retention.

D. RAG-based Question Answering Interface

To enhance the interpretability of the story by allowing children or guardians to ask questions about the ethical content of the story and receive grounded responses, a RAG-based question answering interface is necessary.

- Segment and embed the rewritten story $\hat{S}$ using a sentence transformer.
- Given a query q, retrieve the k most relevant segments $R = \{r_1, \ldots, r_k\}$ using cosine similarity.
- Feed R and q into a generation model to produce the answer A.

$$E_q = \mathrm{Enc}(q) \qquad (7)$$

$$\mathrm{sim}(E_q, E_{r_i}) \quad \forall i \qquad (8)$$

$$R = \mathrm{Top}K_i \left(\mathrm{sim}(E_q, E_{r_i})\right) \qquad (9)$$

$$A = LLM\ (q|R) \qquad (10)$$

Here, A is the response generated based on the retrieved context R.

This multistage pipeline ensures that LLM generated stories not only uphold ethical standards but also support meaningful, transparent interaction with users through grounded question answering.

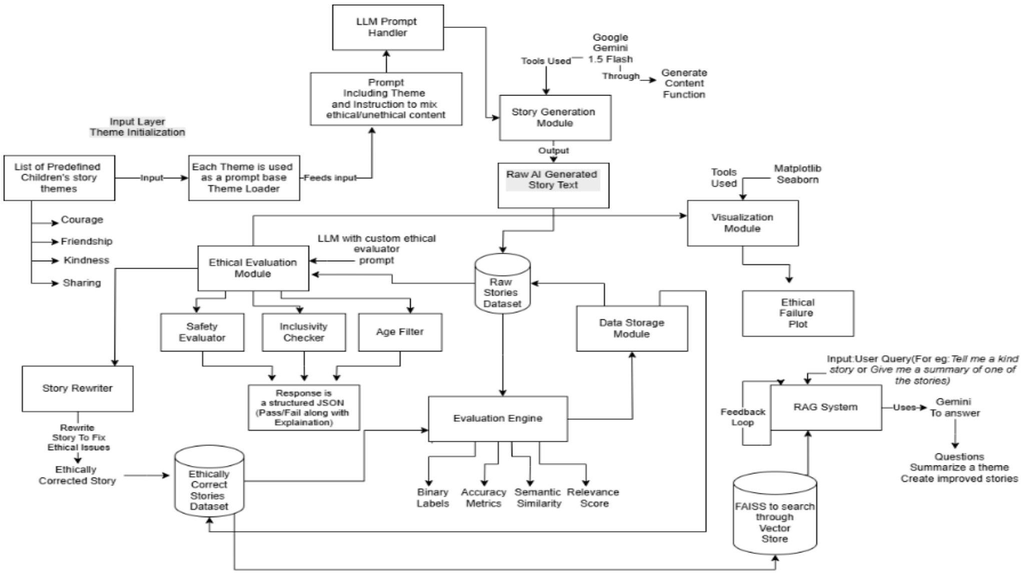
LLM Prompt Handler
Prompt Including Theme and Instruction to mix ethical/unethical content
Tools Used
Google Gemini 1.5 Flash
Through
Generate Content Function
Story Generation Module
Output
Raw AI Generated Story Text
Input Layer Theme Initialization
List of Predefined Children's story themes
Input
Each Theme is used as a prompt base Theme Loader
Feeds input
Courage
Friendship
Kindness
Sharing
Tools Used
Matplotlib Seaborn
Visualization Module
Ethical Failure Plot
LLM with custom ethical evaluator prompt
Ethical Evaluation Module
Raw Stories Dataset
Safety Evaluator
Inclusivity Checker
Age Filter
Data Storage Module
Story Rewriter
Response is a structured JSON (Pass/Fail along with Explaination)
Evaluation Engine
Rewrite Story To Fix Ethical Issues
Ethically Corrected Story
Ethically Correct Stories Dataset
Binary Labels
Accuracy Metrics
Semantic Similarity
Relevance Score
Input:User Query(For eg: Tell me a kind story or Give me a summary of one of the stories)
Feedback Loop
RAG System
Uses
Gemini To answer
Questions
Summarize a theme
Create improved stories
FAISS to search through Vector Store

Fig. 1. Proposed Methodology for Story Generation

III. Result And Discussion

In this research, we implemented an extensive evaluation pipeline for children's stories that are generated by AI, focussing on the three core principles of ethicality: Safety, Inclusivity, and Age Appropriateness. Using Google's Gemini 1.5 Flash model, we generated diverse stories on topics such as friendship, courage, and honesty, intentionally mixing both ethically sound and problematic narratives to simulate real-world variability in AI content generation. The flowchart in Fig. 1 depicts the representation of the system architecture of StoryGuard, an extensive pipeline designed to generate, evaluate, and improve children's stories based on ethical and thematic considerations. The procedure begins with a predefined list of themes such as Courage, Friendship, Kindness, and Sharing. These themes are processed through the Theme Loader, which generates prompts using a selected theme with instructions to blend ethical and unethical elements.

These prompts are then processed by the LLM Prompt Handler, which interfaces with a large language model, specifically Google Gemini 1.5 Flash, via the Story Generation Module to create raw AI-generated stories. The output stories are stored in a Raw Stories Dataset and passed to the Ethical Evaluation Module, which utilizes the LLM with a custom evaluation prompt to assess each story's content. This module integrates three critical evaluators: a Safety Evaluator, Inclusivity Checker, and Age Filter, which together produce a structured JSON output indicating pass/fail status with explanations. Stories failing the evaluation are sent to a Story Rewriter Module, which rewrites the content to address ethical concerns, resulting in a collection of Ethically Corrected Stories.

These ethically corrected and raw stories are then evaluated by the Evaluation Engine, which calculates binary labels, accuracy metrics, semantic similarity, and relevance scores. The data is also logged in the Data Storage Module for persistence and further use. A Visualization Module, using tools like Matplotlib and Seaborn, provides insights into ethical failures across themes and stories via plots and analytics. To make the system interactive and user-centric, a Retrieval-Augmented Generation (RAG) system is integrated. It uses a FAISS-based Vector Store to retrieve relevant stories based on user queries like "Tell me a kind story" or "Summarize a theme". The retrieved content is then enhanced or summarized using Gemini, with a Feedback Loop ensuring ongoing improvement of story quality and system performance. This architecture enables StoryGuard to function as an end-to-end solution for generating safe, inclusive, and age-appropriate children's stories guided by both AI and human-centered ethical considerations.

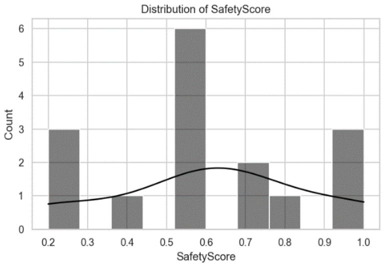

Fig. 2. Safety Score

The histogram and KDE plot in Fig. 2 show how the generated stories scored in terms of safety, based on predefined evaluation metrics. The scores range from 0.2 to 1.0, with a noticeable peak around 0.6. This indicates that a majority of stories were moderately safe, with a few extremely safe stories (scores near 1.0) and a smaller number falling in the low safety range (below 0.4). This reflects the balance between intentionally including some violent themes for research and ensuring overall content control.

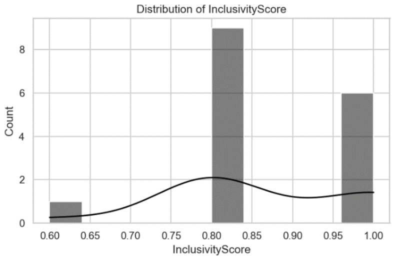

Fig. 3. Inclusivity Score

The inclusivity scores from Figure 3 mostly cluster around 0.8 and 1.0, suggesting a strong performance in generating inclusive content across themes. The Gemini model appears to produce content that is highly inclusive. Very few stories are rated poorly in inclusivity, which is a desirable outcome for children's content. This shows strong ethical language modeling in terms of avoiding exclusionary or biased representations. unethical elements.

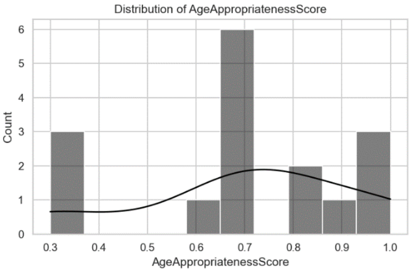

Fig. 4. Age Appropriateness Score

Scores from Figure 4 are mostly centered between 0.6 and 1.0, indicating a good level of age appropriateness. There is a small portion with scores around 0.3, indicating some stories were flagged as inappropriate. This validates the model's ability to adjust tone and complexity, but also highlights outliers likely due to the mixed inclusion of violent elements as per research needs. This balance must be carefully reviewed when deploying such content for actual use with children.

Ethicality scores from Figure 5 generally show a smooth bell-shaped distribution with the majority of data points between 0.6 and 1.0, peaking near 0.7. This suggests that most stories adhered to basic ethical storytelling standards. However, variability exists due to the experimental design (mixing violent and non- violent themes). Higher scores indicate moral story-telling, while the few lower scores highlight potential concerns worth reviewing. Each generated story underwent a structured evaluation using a dedicated prompt template to assess its adherence to ethical standards. The evaluation classified each story on a pass/fail basis for the three criteria. If any criterion failed, the story was automatically rewritten using the same LLM to enforce ethical compliance. To quantify the pipeline's performance, we calculated multiple classification metrics: accuracy, precision, recall, and F1 score by comparing the model's classification outcomes against a predefined gold standard (assuming an ideal ethically acceptable story would pass all three tests).

These metrics provide a holistic view of the system's reliability in identifying and correcting ethical violations.

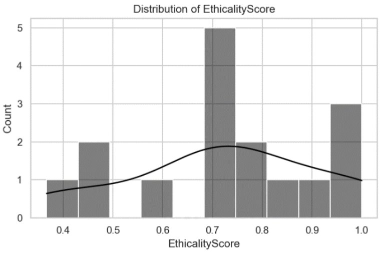

Fig. 5. Ethicality Scores

Additionally, to ensure the rewritten stories preserved the original narrative context, we computed semantic similarity scores using a transformer-based sentence embedding model (all-MiniLM-L6-v2). The majority of rewritten stories maintained high semantic proximity to their originals, indicating ethical correction did not significantly distort narrative meaning. For future integration into a Retrieval-Augmented Generation (RAG) pipeline, we also introduced a relevance scoring module, which estimated how contextually aligned the story was to its intended theme or query. Though placeholder scores were used in this iteration, this component sets the groundwork for incorporating real-time knowledge grounding into future deployments.

IV. Conclusion

This paper presented a comprehensive, multi-stage pipeline for the ethical generation and interpretation of AI-created children's stories. The system integrates unconstrained story generation, sentence-level ethical evaluation, constraint-based rewriting, and a RAG-driven interactive question-answering interface. The approach ensures content safety, preserves narrative coherence, and supports user transparency through explainable AI. While the pipeline is modular and scalable, limitations remain in handling edge-case ethical nuances and preserving deeper creative elements. Future enhancements will focus on adaptive learning and improved retrieval mechanisms to further optimize ethical alignment and user engagement. Supporting multilingual, culturally aware stories and incorporating human feedback can enhance global relevance and real-world performance.

References

[1] E. L. Hill-Yardin, M. R. Hutchinson, R. Laycock, and S. J. Spencer, "A Chat (GPT) about the future of scientific publishing," Brain, Behavior, and Immunity, vol. 110, pp. 152–154, May 2023, doi: 10.1016/j.bbi.2023.02.022.

[2] M. Imran and N. Almusharraf. Google gemini as a next generation ai educational tool: a review of emerging educational technology. *Smart Learning Environments*, 11(1):22, 2024, doi: https://doi.org/10.1186/s40561-024-00310-z.

[3] M. Masalkhi, J. Ong, E. Waisberg, N. Zaman, P. Sarker, A. G. Lee, and A. Tavakkoli. A side-by-side evaluation of llama 2 by meta with chatgpt and its application in ophthalmology. *Eye*, 38(10):1789–1792, 2024, doi: https://doi.org/10.1038/s41433-024-02972-y.

[4] C. Nguyen, D. Carrion, and M. K. Badawy, "Comparative Performance of Anthropic Claude and OpenAI GPT Models in Basic Radiological Imaging Tasks," Journal of Medical Imaging and Radiation Oncology, vol. 69, no. 4, pp. 431–439, Apr. 2025, doi: 10.1111/1754-9485.13858.

[5] M. Gulsoy, V. Kokach, B. Kocaçınar, and F. P. Akbulut, "AI-Driven Contextual Story Creation with Integrated Text and Visual Generation," 2024 Innovations in Intelligent Systems and Applications onference (ASYU), pp. 1–6, Oct. 2024, doi: 10.1109/asyu62119.2024.10757014.

[6] H. Chu and S. Liu, "Can AI tell good stories? Narrative Transportation and Persuasion with ChatGPT," Apr. 2023, doi: 10.31234/osf.io/c3549.

[7] Y. Wang and M. Kreminski, "Can LLMs Generate Good Stories? Insights and Challenges from a Narrative Planning Perspective," 2025 IEEE Conference on Games (CoG), pp. 1–8, Aug. 2025, doi: 10.1109/cog64752.2025.11114137.

[8] P. A. Reed. *Is ChatGPT Creative? Cognitive-Affective Responses to AI-Generated Stories*. PhD thesis, Fielding Graduate University, 2023.

[9] N. Sharma, P. Karwasra, P. Sharma, and M. A. Tahir, "AI Based Story Generation," Pattern Recognition, pp. 32–47, Nov. 2024, doi: 10.1007/978-3-031-78128-5_3.

[10] Edirlei Soares de Lima, Margot ME Neggers, and Antonio L Furtado. Multigenre ai-powered story com- position. *arXiv preprint arXiv:2405.06685*, 2024, doi: https://doi.org/10.48550/arXiv.2405.06685.

[11] N. Beguš, "Experimental narratives: A comparison of human crowdsourced storytelling and AI storytelling," Humanities and Social Sciences Communications, vol. 11, no. 1, Oct. 2024, doi: 10.1057/s41599-024-03868-8.

[12] V. N. Antony and C.-M. Huang, "ID.8: Co-Creating Visual Stories with Generative AI," ACM Transactions on Interactive Intelligent Systems, vol. 14, no. 3, pp. 1–29, Aug. 2024, doi: 10.1145/3672277.

[13] J. J. Y. Chung, W. Kim, K. M. Yoo, H. Lee, E. Adar, and M. Chang, "TaleBrush: Sketching Stories with Generative Pretrained Language Models," CHI Conference on Human Factors in Computing Systems, pp. 1–19, Apr. 2022, doi: 10.1145/3491102.3501819.

[14] H. Shakeri, C. Neustaedter, and S. DiPaola, "SAGA: Collaborative Storytelling with GPT-3," Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing, pp. 163–166, Oct. 2021, doi: 10.1145/3462204.3481771.

[15] A. Yuan, A. Coenen, E. Reif, and D. Ippolito, "Wordcraft: Story Writing With Large Language Models," 27th International Conference on Intelligent User Interfaces, pp. 841–852, Mar. 2022, doi: 10.1145/3490099.3511105.

[16] S. Kreps, R. M. McCain, and M. Brundage, "All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation," Journal of Experimental Political Science, vol. 9, no. 1, pp. 104–117, Nov. 2020, doi: 10.1017/xps.2020.37.

[17] J. Zhou, Y. Zhang, Q. Luo, A. G. Parker, and M. De Choudhury, "Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions," Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–20, Apr. 2023, doi: 10.1145/3544548.3581318.

[18] A. R. Doshi and O. P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," Science Advances, vol. 10, no. 28, Jul. 2024, doi: 10.1126/sciadv.adn5290.

[19] P. Pataranutaporn et al., "AI-generated characters for supporting personalized learning and well-being," Nature Machine Intelligence, vol. 3, no. 12, pp. 1013–1022, Dec. 2021, doi: 10.1038/s42256-021-00417-9.

[20] E. Bonner, R. Lege, and E. Frazier, "Large Language Model-Based Artificial Intelligence In The Language Classroom: Practical Ideas For Teaching," Teaching English With Technology, vol. 2023, no. 1, 2023, doi: 10.56297/bkam1691/wieo1749.

[21] A. Han and Z. Cai, "Design implications of generative AI systems for visual storytelling for young learners," Proceedings of the 22nd Annual ACM Interaction Design and Children Conference, pp. 470–474, Jun. 2023, doi: 10.1145/3585088.3593867.

[22] P. Zhao, H. Zhang, Q. Yu, Z. Wang, Y. Geng, F. Fu, L. Yang, W. Zhang, J. Jiang, and B. Cui. Retrieval-augmented generation for ai-generated content: A survey. *arXiv preprint arXiv:2402.19473*, 2024, doi: https://doi.org/10.48550/arXiv.2402.19473.

[23] P. Ranade and A. Joshi, "FABULA: Intelligence Report Generation Using Retrieval-Augmented Narrative Construction," Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, pp. 603–610, Nov. 2023, doi: 10.1145/3625007.3627505

[24] A. I. Alhussain and A. M. Azmi, "Automatic Story Generation," ACM Computing Surveys, vol. 54, no. 5, pp. 1–38, May 2021, doi: 10.1145/3453156.