Research Paper

Paper	Dataset	Application	Introduction	Methodology	Result
FAIRYTAILOR: A MULTIMODAL GENERATIVE FRAMEWORK FOR STORYTELLING 2108.04324.pdf (arxiv.org)	Text datasets include Reddit WritingPrompts (prompt-story pairs) and children's books (500-token extracts with prompts). * Image dataset is Unsplash, chosen for diverse landscapes and objects relevant to fairy tales.	**Education:** **1.Creative Writing Prompts**: FairyTailor can help students overcome writer's block and spark creative storytelling ideas. 2.**Interactive Learning:** Teachers can utilize the platform to create interactive story-based lessons for different age groups. **3.Language Learning:** FairyTailor can be a fun tool for language learners to practice building sentences and narratives	1.Automated story generation is challenging due to open-ended nature and creativity requirements. ● Multimodal content (text and images) can enhance storytelling, especially for young readers. ● Existing systems lack human interaction and focus on unimodal generation. **Related work** ● Previous story generation models struggle with repetition, incoherence, and lack of control over themes. ● * Story visualization systems often require specific text or take too long to generate images ● FairyTailor aims to address these issues by combining human input with multimodal generation.	**1.Benchmark Model:** (details might be limited) ● Training a pre-trained language model (like GPT-2) on the text data. ● Training an image retrieval model on the Unsplash dataset with text descriptions (if noun-based retrieval). ● Defining rules or metrics for re-ranking generated text (readability, positivity, etc.). **2.Final Model:** ● Fine-tuning the pre-trained language model on the story dataset, focusing on metrics like coherence, simplicity, and "tale-like" qualities. This "tale-like" score might involve analyzing common fairy tale elements in the training data. ● Refining the image retrieval process using pre-computed embeddings and style transfer for visual consistency.	● **Positive User Experience:** Users enjoyed the platform and found the high-quality autocomplete suggestions helpful for generating stories. ● **Improved Storytelling:** The final model, compared to the Benchmark model, likely creates more coherent, simpler, and "tale-like" stories through fine-tuning and additional metrics. ● **Valuable Research Tool:** FairyTailor offers researchers a platform to deploy and user-test their story generation models.
2209.06192.pdf (arxiv.org) StoryDALL-E: Adapting Pretrained Text-to-Image Transformers for Story Continuation	The paper introduces a new story continuation dataset called DiDeMoSV, along with using existing datasets PororoSV and FlintstonesSV adapted for this task.	1.**Interactive Storytelling**: Imagine a system where you can start a story with a text prompt and the AI generates visuals to follow along, creating an interactive story experience. 2.**Storyboarding and Concept Art**	This paper introduces a new task called story continuation, which is a more realistic setting for story visualization compared to the existing task. It also proposes a model called StoryDALL-E to adapt large pretrained text-to-image	1.**Story Continuation Task:** This reframes story visualization by providing an initial scene (source image) as input. This allows the model to maintain consistency with characters and setting throughout the story. 2.**StoryDALL-E Model**: This is the proposed model based on a pre-trained text-to-image transformer. Here's what makes it special:	● **Story Continuation Task:** This approach is more realistic for story visualization compared to traditional methods as it allows for unseen characters and settings. ● **StoryDALL-E Model:** This model outperforms a GAN-based model (StoryGANc) on story

		Generation: Storyboards and concept art are visuals used in movies, animation, and games to plan out the narrative and visuals. This technology could help artists and creators generate drafts and explore visual ideas for their stories. **3.Education and Learning:** Stories can be a powerful tool for education. This technology could be used to create interactive educational stories where visuals are automatically generated based on the narrative. **4.Generating Illustrations for Text Content:** Imagine automatically generating illustrations for written content like news articles, books, or social media posts. This could save time and resources for content creators.	transformers for this task. Story continuation provides an initial scene to the model, allowing it to better generalize to unseen characters and settings in new narratives.	• **Global Story Encoder:** This component captures the overall context of the story by processing all the captions provided. • **Retro-fitted Cross-Attention Layers:** These layers allow the model to focus on relevant parts of the source image and use that information for generating subsequent frames in the story. **3.Datasets:** The paper utilizes a combination of new and existing datasets: • A new dataset called DiDeMoSV specifically designed for story continuation. • Two existing story visualization datasets (PororoSV and FlintstonesSV) adapted for story continuation by using the first frame as the source image. **4.Evaluation:** • The model is compared to a GAN-based model (StoryGANc) on the three datasets. StoryDALL-E outperforms StoryGANc in terms of visual quality (FID score) and achieves similar character classification accuracy on some datasets. • Human evaluation also shows that StoryDALL-E generates more visually appealing and relevant stories. **5.DALL-E Mega Integration:** The final version of the paper incorporates the recently released DALL-E Mega, a	continuation datasets in terms of visual quality (FID score) and achieves similar character classification accuracy on some datasets. Human evaluation also shows it generates better stories. • **DALL-E Mega Integration:** Integrating the larger DALL-E Mega model further improves results, achieving up to 3% better FID score. • **Applications:** The approach has potential applications in interactive storytelling, storyboarding, education, illustration generation, and accessibility tools.

				larger and more powerful pre-trained model. This leads to further improvement in results (up to 3% better FID score). **6.Demo System:** An in-browser demo system is made available allowing users to experiment with the mega-StoryDALL-E model trained on the Pororo dataset. Overall, the methodology focuses on adapting pre-trained text-to-image models with task-specific modifications for story continuation. This approach achieves promising results in generating coherent and visually appealing stories.	
Zero-shot Generation of Coherent Storybook from Plain Text Story using Diffusion Models 2302.03900.pdf (arxiv.org)	The research does not explicitly mention using a specific dataset for training or evaluation. The focus is on the zero-shot generation capability of the method, meaning it should work on any story text provided.	The primary application of this research is the generation of storybooks from text descriptions. This could be useful for creating educational materials, children's books, or personalized stories. Additionally, the method's ability to generate coherent image sequences could be applied to other areas such as animation or video game development.	This research paper proposes a new method for generating coherent storybooks from plain text stories, without any training data required. This is achieved by combining large language models (LLMs) and text-guided latent diffusion models. LLMs are used to understand the context of a story and generate prompts for the diffusion models, which then create corresponding images.	**1.Prompt Generation:** An LLM is used to analyze the story text and generate captions describing each scene. These captions are further refined to be suitable for text-to-image models. **2.Initial Image Generation:** The refined captions are used as prompts for a text-guided latent diffusion model, which generates images for each scene. **3.Face Restoration:** Faces in the generated images are identified and enhanced using a separate model. **4.Coherent Identity Injection:** To ensure the main character maintains a consistent appearance, an iterative process refines the images using a textual embedding of the desired identity.	**1.Zero-shot Storybook Generation:** The method successfully generates coherent storybooks from plain text stories, without any prior training on specific image-text pairs. This allows for creating stories with various characters and settings without needing large amounts of labeled data. **2.Consistent Character Appearance:** The iterative identity injection process ensures that the main character maintains a consistent appearance throughout the storybook. This is crucial for maintaining coherency and immersion in the story. **3.Outperforms Baselines:** The proposed method surpasses existing text-guided and image-guided semantic

					image editing techniques in terms of: • **Coherency Preservation:** The main character's appearance remains consistent across different images in the story. • **Background Preservation:** The background elements from the initial image are maintained during the editing process. **4.User Study:** A user study with human participants showed that the generated images achieved high scores in terms of: • **Correspondence:** The images accurately reflect the corresponding scenes described in the text. • **Coherency:** The main character's appearance remains consistent throughout the story. • **Smoothness:** Transitions between foreground elements (characters) and background elements are seamless and natural-looking.
Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation 17 march 2023 2212.11565.pdf (arxiv.org)		**1. Object Editing:** By modifying the text prompt, users can replace, add, or remove objects in the video while maintaining consistency in their movements. **2.Background Change:** The background scenery of the video can be altered while preserving	1. Generated video from text description using pre-trained text to image models. 2. Traditional approach involves training T2V models on large scale video dataset which is expensive. Tune-a-video overcome this by using a single text-video pair for video generation.	**1.One-Shot Video Tuning:** Instead of training a model on massive video datasets, Tune-A-Video leverages a pre-trained text-to-image diffusion model and fine-tunes it on a single video and its corresponding text description. **2.Leveraging Pre-trained Knowledge:** The method utilizes the ability of pre-trained text-to-image models to understand and generate	1.The method outperforms baseline models (CogVideo, Plug-and-Play, Text2LIVE) in terms of generating temporally coherent videos that accurately reflect the edited text prompt. 2.User studies also showed a preference for videos generated by Tune-A-Video compared to the baselines.

		the object's actions. **3.Style Transfer:** Videos can be transformed into various artistic styles (e.g., comic book, Van Gogh) by incorporating style cues in the text prompt. **4.Compatibility with Personalized Models:** Tune-A-Video can be integrated with personalized text-to-image models, allowing users to generate videos in a specific style or featuring a particular subject.		images based on textual descriptions. **3.Sparse Spatio-Temporal Attention:** To capture temporal coherence between video frames, the authors introduce a sparse spatio-temporal attention mechanism that focuses on the first frame and the previous frame, reducing computational complexity. **4.Structure Guidance via DDIM Inversion:** During inference, structural information from the input video is incorporated through DDIM inversion, which helps generate temporally coherent videos.	**Overall, Tune-A-Video presents a promising approach for text-driven video generation and editing. It leverages pre-trained models, requires minimal training data (one video-text pair), and offers flexibility in terms of object manipulation, background changes, and style transfer.**
Full article: Using artificial intelligence in craft education: crafting with text-to-image generative models (tandfonline.com) **Using artificial intelligence in craft education: crafting with text-to-image generative models**		**For Educators:** **1.Developing curriculum:** This research can inform the development of lesson plans and activities that integrate generative AI into the craft education curriculum. Educators can use the findings to explore the potential benefits (idea generation, externalization of ideas) while mitigating the concerns (limited by constraints, de-emphasis of making). **2.Facilitating discussions:** The paper provides talking points for educators to initiate discussions with students about AI, creativity, and the future of craft. They	The primary objective of this research is to shed light on the opportunities and drawbacks of utilizing generative AI as a tool for creative design exploration and learning within the realm of craft education. The insights gleaned from this study can contribute to ongoing discussions about the role of AI in fostering creative learning and inform the development of future pedagogical practices.	1.A workshop was conducted where 15 participants (craft teachers and teacher educators) used generative AI to create images based on text descriptions. 2.Following this, there were discussions about the potential benefits and drawbacks of this technology in craft education.	**Potential benefits** **1.Enhanced Ideation:** Participants saw AI as a tool to spark creative design ideas. It could help visualize unusual concepts or impossible shapes, which could then be further developed into real-world projects. **2.Improved Externalization:** AI could potentially help students articulate vague design ideas by generating visual representations based on their descriptions. This could be especially useful for young children who struggle to express themselves verbally. **3.Analyzing Design Constraints:** AI-generated visualizations could be used to analyze the feasibility of a design idea in terms of materials, skills, and available resources. This could help students make informed decisions during the design process. **Concerns and challenges**

		can use the research to explore topics like AI bias, copyright issues, and the importance of critical thinking in the digital age. **3.Self-reflection:** Craft educators themselves can reflect on their teaching practices in light of this research. They can consider how AI might enhance or hinder student learning and adapt their approaches accordingly. **For Educational Developers and Policymakers:** **1.Informing professional development:** The research can be used to design professional development programs that equip educators with the knowledge and skills to integrate AI effectively into their craft education practices. **2.Shaping educational policy:** The findings can inform policy discussions about the role of AI in education and the importance of addressing potential challenges like bias and the de-emphasis of hands-on learning. **For Developers of Generative AI Tools:** **1.Designing for**			**1.Limited by Design Constraints:** AI-generated designs might not consider real-world limitations like material availability, student skill level, or ergonomics. This could lead to frustration if students attempt to create designs that are impossible to execute. **2.De-emphasis of Making:** The embodied experience of working with materials and using craft skills might be lost if students rely heavily on digital design tools. **3.Assessment Challenges:** The use of AI in design raises questions about how to assess creative learning and design skills accurately. **4.Unforeseen Tensions** ● **Bias:** AI-generated images are based on the data they are trained on, which can lead to perpetuating existing biases. This raises concerns about limiting creativity and reinforcing stereotypes. ● **Copyright:** There are questions about copyright infringement as AI uses existing artwork to generate new images. ● **Black-boxing Creativity:** The process by which AI generates images is opaque, making it difficult to understand how creativity is involved. This raises questions about human agency in the design process. ● **Behavior Engineering:** AI-generated images with personalized styles could be used for marketing and manipulating user behavior. ● **Misinformation and Cultural Memory:** AI-generated images could

		education: Developers can use the research to understand the specific needs and concerns of educators using generative AI in craft education. This can guide the development of user-friendly tools with features that address limitations and promote creative learning.			blur the lines between real and fake, potentially impacting trust in information and altering cultural memory.
StorVi (Story Visualization): A Text-to-Image Conversion https://www.ijfcc.org/papers/328-CS3009.pdf	The paper doesn't explicitly mention the dataset used. It likely consists of a collection of text stories and corresponding images for characters, objects, and environments.	**1.Improve comprehension:** Visualizing the story elements can make it easier for children to understand the narrative and follow along. **2.Enhance enjoyment:** Engaging visuals can make storytelling more interactive and fun for children, increasing their interest and motivation. **3.Develop creative thinking:** By seeing the story come to life visually, children might be inspired to use their own imagination and creativity to further visualize the story.	This research paper proposes a system named StorVi (Story Visualization) that converts text stories into 2D scene images. The target audience for this system is children ages 4-7 and the goal is to aid them in visualizing stories through pictures. The researchers acknowledge the importance of storytelling and visualization in children's development and believe this system can be a helpful tool.	breakdown of the methods used: **1.Part-of-Speech (POS) Tagging:** This identifies the grammatical function of each word in the story (noun, verb, adjective, etc.). **2.Classification Algorithm:** Categorizes elements within the story such as characters, actions, objects, locations, and environment. **3.Simple Co-reference Resolution Algorithm:** Identifies instances where pronouns refer to previously mentioned characters. **4.Brute Force String Matching Algorithm:** Finds corresponding image matches for characters and environments within a database. **5.Degeneralization:** If a general term is encountered (e.g., furniture), the system searches for a more specific instance within the same category (e.g., chair). **6.Textualization:** If no image match is found, the system displays the word itself as text. **7.Spatial Rule-Based Algorithm:** Positions characters and objects within the scene based on spatial prepositions used in the story (e.g., under, on, behind).	The researchers evaluated StorVi through user surveys with teachers and parents of children ages 4-7. The results indicated: **1.Usability:** Teachers and parents agreed the system was usable. **2.User-Friendliness:** Both groups agreed the system was user-friendly. **3.Content of Generated Frames:** Teachers agreed the generated images were good, while parents rated them as fair.

Make-A-Story: Visual Memory Conditioned Consistent Story Generation Make-a-Story: Visual Memory Conditioned Consistent Story Generation (thecvf.com)	**FlintstonesSV** [16]: This dataset contains story descriptions with named entities (characters) replaced by pronouns for a more challenging reference resolution task. It includes: ● 20,132 training stories ● 2,071 validation stories ● 2,309 test stories ● 7 main characters ● 323 backgrounds **PororoSV** [30]: This dataset contains stories originally written with references to characters by pronouns. The authors modified the dataset similarly to FlintstonesSV for reference resolution	This method is useful for generating a sequence of illustrative image frames with coherent semantics given a sequence of sentences.	1.There has been a recent surge in impressive generative models that can produce high-quality images based on descriptions. 2.Current models rely on descriptions that clearly describe scenes and actors. 3.This method is not suitable for complex tasks like story visualization, where characters and backgrounds need to be consistent across scenes. 4.This paper addresses these challenges by proposing a novel autoregressive diffusion-based framework with a visual memory module.	1.The approach uses an autoregressive structure to generate temporally consistent stories. 2.It builds upon diffusion models to generate high-quality images and learns the generative conditional distribution of visual stories. 3.Latent Diffusion Models are employed to improve efficiency when dealing with high-dimensional data. 4.A memory attention mechanism is introduced to ensure consistency and smooth story progression. This mechanism resolves ambiguous references using visual memory.	1.The model can generate high-quality frames that are consistent with the story description. 2.It can also model appropriate correspondences between characters and backgrounds. 3.The model outperforms previous state-of-the-art methods in terms of generating frames with high visual quality and consistency.