

# Automatic Generation of Illustrations for Children's Fairy Tales From Text: AI-Powered App Based on DALL-E 3

Irina Kogai

Dubai, United Arab Emirates

email: ikogai.dev@gmail.com

**Abstract**—This paper explores the capabilities of DALL-E 3, an advanced artificial intelligence model, to automatically generate illustrations for children's fairy tales. The primary objective is to assess the relevance of the images produced by this Artificial Intelligence (AI), examining how well these illustrations can complement and enhance the storytelling experience for young audiences. To achieve this, we used Toloka.ai, a crowdsourcing platform where a diverse pool of evaluators assessed the relevance of AI-generated illustrations to specific excerpts of fairy tales. This methodology allowed for a broad-based, democratized approach to understanding how illustrations created by AI are perceived in terms of accuracy, engagement, and emotional resonance with the source material. Our findings reveal that DALL-E 3 demonstrates a significant potential in creating visually appealing, contextually appropriate, and emotionally resonant illustrations that can rival traditional artistic methods, with a relevance rate of 62%. However, the conducted research shows that images generated by DALL-E 3 are often irrelevant and inconsistent, which makes it hard to use a system like this in a fully automated fashion without human supervision. This research contributes to the emerging field of AI in creative industries, offering insights for developers, educators, and storytellers seeking innovative approaches to enrich children's literary experiences.

**Keywords**—DALL-E 3; children's fairy tales; automatic illustration generation; AI-generated illustrations; visual storytelling.

## I. INTRODUCTION

In the realm of children's literature, illustrations play a pivotal role in bringing stories to life, offering visual narratives that complement and enhance the textual content. Traditionally, this task has been the domain of human artists, whose creativity and interpretation add depth and emotion to the tales. However, the advent of AI technologies, particularly in the field of generative art, presents novel opportunities and challenges for the creation of such illustrations.

The development of AI models like DALL-E 3 by OpenAI signifies a breakthrough in our ability to generate detailed, context-specific images from textual descriptions. This technology harbors the potential to revolutionize the way illustrations are produced for children's fairy tales, making it possible to automate the creation of artwork that resonates with the narratives and characters of these stories. The implications for storytelling, publishing, and educational content are profound, offering new avenues for engagement and creativity.

Despite the promise, the application of AI in creating illustrations for children's literature raises questions regarding the quality, relevance, and emotional connection of the generated images. Can an AI like DALL-E 3 understand and interpret fairy tales well enough to produce illustrations that capture the essence of these stories? How do these AI-generated

illustrations compare to those created by human artists in terms of appealing to children's imaginations and understanding? Will they relate to the story or not?

To address these questions, our study investigates the capability of DALL-E 3 to generate illustrations for children's fairy tales from text descriptions. We used Toloka.ai, a crowdsourcing platform, to assess the relevance of the AI-generated illustrations.

This paper aims to contribute to the emerging discourse on the intersection of AI and creative arts, specifically within the context of children's literature. By exploring the potential of DALL-E 3 to create meaningful and engaging illustrations, we seek to understand the role that AI can play in enriching storytelling and whether it can serve as a viable tool for authors and publishers in the future.

In Section 2, the paper explores related works in the field of automatic generation of illustrations for children's fairy tales from text. Section 3 outlines the methodology adopted in the development of an AI-powered application designed to automatically generate illustrations for children's fairy tales. Section 4 details the implementation of the AI-powered application, including the technical architecture and the integration of the Generative Pre-trained Transformer 4 (GPT-4) and DALL-E 3. Section 5 presents the performance evaluation of the solution, highlighting the relevance of the images. Finally, Section 6 concludes the paper with a summary of the findings and implications for the future of AI in children's literature.

## II. RELATED WORKS

The domain of automatic generation of illustrations for children's fairy tales from text represent a fascinating intersection of Natural Language Processing (NLP), computer vision, and creative AI technologies. This section reviews the related works that have contributed to the development of AI-powered applications, particularly focusing on technologies akin to DALL-E 3 for generating illustrations from textual content. The foundation for generating illustrations from text lies within the advancements of text-to-image synthesis. Reed et al. [1] were among the pioneers to explore this domain by conditioning deep convolutional Generative Adversarial Networks (GANs) on text features, enabling the generation of realistic images from descriptive text. Their work paved the way for subsequent innovations in image synthesis from textual descriptions. GANs have undergone significant evolution since their introduction by Goodfellow et al. [2]. The development of variants such as Conditional GANs (cGANs) and, later, Big GANs [3], has drastically

improved the quality and resolution of generated images, making them more applicable for creating detailed illustrations. Parallel to the advancements in GANs, progress in language models, particularly transformer-based architectures like Bidirectional Encoder Representations from Transformers (BERT) [4] and GPT [5], has significantly improved the understanding and generation of natural language. DALL-E, introduced by OpenAI [6], marked a significant milestone in text-to-image generation by effectively combining the prowess of GPT-4 [7] with image synthesis capabilities. Its successor, DALL-E 2 [8], and the latest, DALL-E 3 [9], have further refined this approach, offering unprecedented quality and creativity in generating images from textual descriptions. These models have demonstrated a remarkable ability to understand complex descriptions and generate coherent and contextually relevant illustrations, making them ideal for creating children’s book illustrations. Several studies have explored the application of AI in children’s literature, particularly in generating illustrations. The work of Zakraoui et al. on “Visualizing Children Stories with Generated Image Sequences” [10] represents a significant contribution to the field, demonstrating the potential of using AI to create sequences of images that narrate children’s stories. This approach underscores the evolving landscape of AI applications in children’s literature focusing on the ability to visually narrate stories through a series of generated images, enriching the storytelling experience.

### III. METHOD

This section outlines the methodology adopted in the development of an AI-powered application designed to automatically generate illustrations for children’s fairy tales. The process leverages two cutting-edge AI technologies: GPT-4 for story generation and DALL-E 3 for the creation of corresponding illustrations. The integration of these technologies facilitates the end-to-end generation of illustrated fairy tales, from textual content creation to visual representation. The initial phase of the methodology involves generating the textual content of the fairy tales. For this purpose, OpenAI’s GPT-4, an advanced iteration of the Generative Pre-trained Transformer models are utilized. GPT-4’s ability to understand and generate human-like text makes it an ideal tool for crafting creative and engaging stories. The process is initiated by providing GPT-4 with a seed prompt, which outlines the desired theme or elements to be included in the fairy tale. The model then generates a story based on this prompt, ensuring originality and thematic relevance. The generated stories are structured to include clear narrative elements such as setting, characters, conflict, and resolution, making them suitable for children’s literature. Following the creation of the textual content, DALL-E 3 is employed to generate illustrations corresponding to specific parts of the fairy tales. DALL-E 3, a state-of-the-art text-to-image generation model developed by OpenAI is capable of creating detailed and contextually relevant images from textual descriptions. This capability is leveraged to translate selected excerpts of the generated fairy tales into visual illustrations. The generated stories and corresponding

illustrations are integrated into a cohesive fairy tale book format. This integration involves pairing each illustration with the relevant portion of the text, ensuring a logical and seamless narrative flow. Following the initial integration, a refinement process is undertaken. This process includes reviewing the coherence between the text and illustrations, the narrative flow, and the overall aesthetic appeal. Adjustments may involve regenerating certain illustrations with modified descriptions to better capture the intended scene or revising parts of the text for clarity or impact. The methodology described herein represents a novel approach to leveraging AI for the creative task of generating illustrated children’s fairy tales. By combining the linguistic capabilities of GPT-4 with the visual creativity of DALL-E 3, this work contributes to the expanding field of AI-assisted content creation, opening new possibilities for storytelling and illustration.

### IV. IMPLEMENTATION

The implementation of the AI-powered application for generating illustrated children’s fairy tales involved the development of a frontend interface using TypeScript with React [11] and a backend server utilizing Node.js [12]. This section details the technical architecture, the integration of the GPT-4 and DALL-E 3 APIs for content and illustration generation and the overall workflow of the application. The application architecture is divided into two main components: the frontend and the backend. The frontend serves as the user interface, allowing users to input initial prompts for story generation and to interact with the generated content. It is developed using TypeScript, a statically typed superset of JavaScript, which provides enhanced development experience and maintainability through strong typing and object-oriented features. The frontend communicates with the backend server via RESTful APIs to request story generation and illustration services. The core functionality of the application relies on the integration with OpenAI’s GPT-4 and DALL-E 3 APIs. The process begins with the backend server receiving a prompt from the frontend, which it forwards to the GPT-4 API to generate the fairy tale text. The code of the feature creating fairy tales and illustrations is shown in Figure 1. The GPT-4 API is called with a specific prompt to control the length, style, and thematic elements of the generated story, ensuring it aligns with the user’s input and is suitable for a children’s fairy tale. We ask the GPT-4 to return the response in JavaScript Object Notation (JSON) format. The decision to utilize JSON objects for structuring the fairy tale text was driven by the need for a flexible and easily parseable format that could accommodate the dynamic nature of story generation and illustration.

For creating a fairy tale, we use the following prompt:

```
const NEWPROMPT = Generate a fairy tale in the genre of SELECTEDGENRE featuring a CHARACTERGENDER protagonist named CHARACTERNAME. Generate a response in JSON format!!! Each paragraph of the fairy tale should be encapsulated as a value within a distinct key-value pair, where keys are sequential numbers starting from 1 and it is always a number and value is always a string. The tale should
```

```

const COMPLETION = await openai.chat.
  completions.create({
    messages: [{ role: "system",
      content: NEWPROMPT }],
    model: "gpt-4-1106-preview",
    response_format: { type: "
      json_object" }
  });
const STORY = JSON.parse(completion?.
  choices[0]?.message.content)

```

Figure 1. Creating a fairy tale

incorporate `ADDITIONALPROMPT` to enrich the storyline. Ensure that each value contains a mini-narrative of 3 to 5 sentences, facilitating their use as prompts for image generation. Structure the JSON object so that it is easily parseable, with each paragraph kept under 2000 characters and segmented into individual sentences as separate key-value pairs. This format will support direct use of the sentences as prompts for creating corresponding images. Please maintain a coherent and engaging narrative flow throughout the fairy tale, adhering to the specified genre and character details. Do not include any explicit content, and ensure that the tale is suitable for children. Do not use words that are flagged by OpenAI content policies. Use words up that a 12 year old would understand.

The design of the prompts was guided by the need to generate fairy tales that are imaginative, culturally sensitive, and age-appropriate. The prompt structure was carefully crafted to ensure the generated stories and illustrations would captivate children's interest while adhering to ethical and content standards.

The JavaScript Object Notation (JSON) object is structured to facilitate the extraction of individual paragraphs, each of which serves as a prompt. The generated fairy tale is embedded in the following prompt which was developed with the intention to create visuals that complement the story's narrative, emphasizing a vibrant, engaging, and wholesome environment:

```

const ILLUSTRATIONPROMPT = Create a vibrant and
engaging illustration suitable for a children's book, capturing
the essence of an adventurous story. The scene should vivid
STORY, bringing to life the protagonist journey with rich,
colorful imagery that appeals to young readers. The illustration
should be wholesome, filled with elements of nature such as
trees, flowers, and perhaps friendly forest animals, to comple-
ment the story's theme. It is crucial that the image is free from
any form of violence, explicit content, or text, ensuring it is
appropriate for a children's audience and adheres to publishing
standards for educational and entertaining material. The art
style should be warm and inviting, with a focus on creating
an immersive experience for children to imagine themselves
alongside the protagonist on its journey.

```

The directive to include elements of nature and ensure the

```

openai.images.generate({
  model: "dall-e-3",
  ILLUSTRATIONPROMPT,
  size: "1024x1024",
  style: 'natural',
}).then(image => image?.data[0]?.url).
  catch(error => {
    console.error('Error generating image
      for prompt ${i}:', error);
    return null;
  });

```

Figure 2. Function for creating a fairy tale and generating illustrations

absence of violence or explicit content in the prompt reflects our commitment to fostering a safe and positive reading environment for children. The choice of a warm and inviting art style was made to encourage children's imagination and identification with the story's protagonist and their adventures.

Then, we create illustrations using the following code in Figure 2.

Once the story is generated, the backend server identifies key scenes or elements within the text that would benefit from illustration. For each of these, a descriptive text snippet is extracted or formulated and sent to the DALL-E 3 API to generate corresponding illustrations. The DALL-E 3 API's capability to create high-quality images from textual descriptions allows for the generation of visually appealing and contextually relevant illustrations for the fairy tale. The user initiates the story creation process via the frontend interface by providing a seed prompt or theme for the fairy tale. The frontend then communicates with the backend server, which orchestrates the calls to the GPT-4 and DALL-E 3 APIs. The generated story and illustrations are returned to the frontend, where they are displayed to the user in an interactive, book-like format. Users have the option to regenerate specific parts of the story or illustrations if they desire alterations. The frontend interface provides tools for users to select parts of the story for regeneration and to submit new prompts or descriptions for illustrations, enhancing the interactive and customizable nature of the application.

## V. PERFORMANCE

To evaluate the performance and relevance of the AI-generated illustrations to the specific excerpts of fairy tales, an extensive evaluation was conducted using Toloka.ai, a crowdsourcing platform known for its diverse pool of evaluators. The primary objective of the evaluation was to determine the extent to which the AI-generated illustrations accurately reflected the content and spirit of the fairy tale excerpts they were associated with. To achieve this, a task framework was designed for evaluators on Toloka.ai, guiding them through a systematic assessment process.

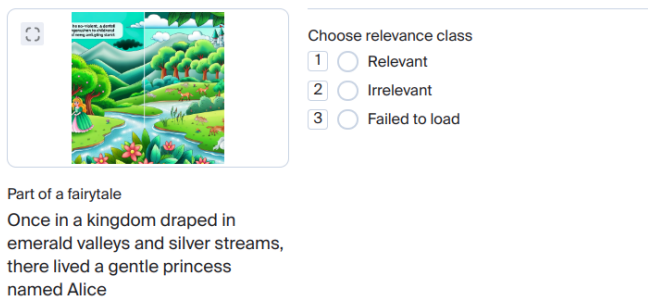


Figure 3. Evaluation sample

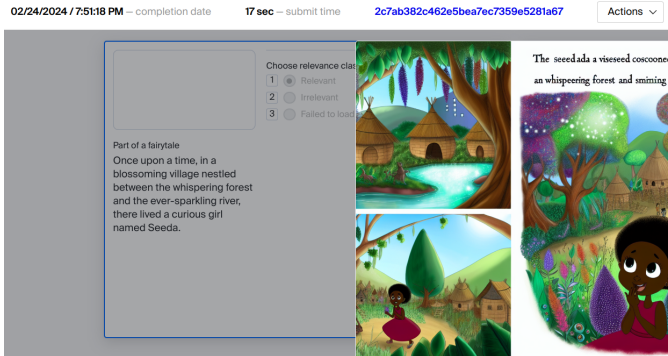


Figure 4. Response sample 1

Evaluators were presented with an AI-generated illustration alongside a specific excerpt from a fairy tale as in Figure 3. Their task was to judge the relevance of the illustration to the text, considering whether the visual elements accurately represented the narrative details and mood of the story segment. The task instructions provided to evaluators were as follows:

“Examine the illustration and Title: Evaluators were instructed to start by observing the illustration and reading its accompanying title, both of which were intended to encapsulate a segment of a fairy tale.

Make your decision: Based on their examination, evaluators were asked to classify the illustration as either “Relevant” or “Irrelevant” to the part of the fairy tale it was supposed to depict.

Adjust image size if necessary: An option to enlarge the illustration was available to ensure evaluators could closely inspect the image for detailed assessment.”

The criteria for relevance were clearly defined:

“Relevant: The illustration was considered relevant if it directly corresponded with the narrative details of the fairy tale segment, meaning all key elements mentioned or implied in the text were visually represented.

Irrelevant: An illustration was marked as irrelevant if it significantly diverged from the fairy tale segment by depicting unrelated elements, omitting critical details, or if the title and content of the illustration did not align with the narrative segment under evaluation.”

Figure 4 and Figure 5 are representing examples of the

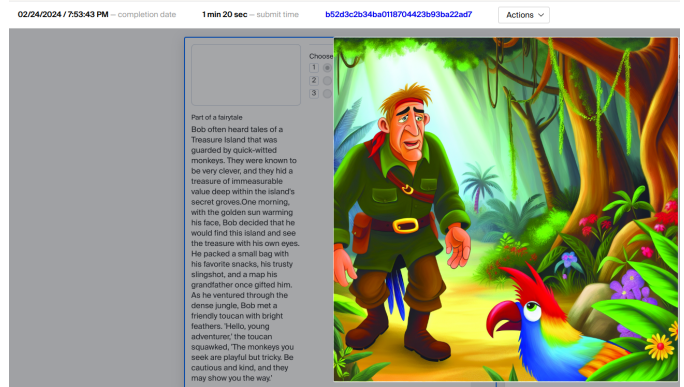


Figure 5. Response sample 2

evaluation’s results. Involving 134 evaluators, the study covered 473 distinct evaluations, offering a broad perspective on the effectiveness of AI in producing contextually appropriate visuals. According to the collected data, 295 evaluations deemed the illustrations relevant, indicating a successful match between the illustration and the narrative content of the fairy tale excerpt. Conversely, 178 evaluations found the illustrations irrelevant, pointing to a discrepancy between the visual representation and the story segment.

These findings offer a nuanced view of the AI’s performance, showcasing a substantial success rate in generating illustrations that align well with the fairy tales’ narrative elements. However, the presence of a significant number of illustrations judged as irrelevant also highlights areas where the AI’s understanding and interpretation of text can be further improved. The insights gained from this extensive feedback loop are invaluable for future enhancements of AI models, aiming to increase the precision and relevance of generated illustrations in subsequent fairy tale projects. This evaluation not only demonstrates the AI’s potential in creating visually engaging content for children’s literature but also emphasizes the importance of continuous refinement in AI applications to better serve creative storytelling.

## VI. CONCLUSION

This paper presented an innovative approach to enriching children’s fairy tales with AI-generated illustrations, leveraging the capabilities of GPT-4 for text generation and DALL-E 3 for visual content creation. Through the integration of advanced AI technologies, we have developed an application that automates the process of creating illustrated fairy tales, offering a novel interactive experience for both children and adults alike.

Our methodology involved using the GPT-4 API to generate fairy tale narratives based on user prompts and employing the DALL-E 3 API to produce illustrations that visually complement specific excerpts from these tales. The implementation showcased the seamless collaboration between frontend technologies like TypeScript and backend services powered by Node.js, culminating in a user-friendly platform that bridges

the gap between traditional storytelling and modern technological innovations. However, there are still possibilities for further improvements. Future plans include:

- to explore innovative methods that allow the selection of the text excerpts to create illustrations autonomously;
- to incorporate additional quality metrics into our evaluation process, including visual appeal and emotional resonance;
- to compare existing solutions such as Midjourney and DreamStudio for generating illustrated children's fairy tales.

The performance assessment carried out through the crowdsourcing platform Toloka.ai, yielded significant observations regarding the relevance and precision of the AI-generated illustrations. With a relevance rate of 62.3%, as indicated by 295 out of 473 evaluations affirming the alignment between illustrations and their corresponding fairy tale excerpts, these findings illuminate the capability of AI in generating visually engaging and contextually fitting illustrations for literary compositions. However, this evaluation also spotlighted areas necessitating enhancement, particularly the need to improve the AI's interpretation of complex narrative nuances aimed at diminishing the occurrence of illustrations deemed irrelevant.

In conclusion, this project demonstrates the feasibility of using AI to automate the generation of illustrated children's fairy tales. The findings from this study underscore the importance of continuous refinement in AI model training and the potential for future research to further enhance the storytelling experience. By pushing the boundaries of what is possible with AI, we can provide enriching and immersive literary experiences that captivate the imagination of readers of all ages.

#### REFERENCES

- [1] S. Reed *et al.*, "Generative adversarial text to image synthesis," in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan and K. Q. Weinberger, Eds., ser. Proceedings of Machine Learning Research, vol. 48, New York, New York, USA: PMLR, 2016, pp. 1060–1069. [Online]. Available: <https://proceedings.mlr.press/v48/reed16.html> [retrieved: Mar. 2024].
- [2] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, MIT Press, 2014, pp. 2672–2680. [retrieved: Mar. 2024].
- [3] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018. [retrieved: Feb. 2024].
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018. [retrieved: Feb. 2024].
- [5] A. Radford *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019. [retrieved: Feb. 2024].
- [6] M. D. M. Reddy, M. S. M. Basha, M. M. C. Hari, and M. N. Penchalaiah, "Dall-e: Creating images from text," *UGC Care Group I Journal*, vol. 8, no. 14, pp. 71–75, 2021.
- [7] J. Achiam *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. [retrieved: Feb. 2024].
- [8] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, vol. 1, no. 2, p. 3, 2022.
- [9] J. Betker *et al.*, "Improving image generation with better captions," *Computer Science*. <https://cdn.openai.com/papers/dall-e-3.pdf>, vol. 2, no. 3, p. 8, 2023.
- [10] J. Zakraoui, M. Saleh, S. Al-Maadeed, J. M. Alja'am, and M. S. Abou El-Seoud, "Visualizing children stories with generated image sequences," in *Visions and Concepts for Education 4.0: Proceedings of the 9th International Conference on Interactive Collaborative and Blended Learning (ICBL2020)*, Springer, 2021, pp. 512–519.
- [11] Jordan Walke, *React*, version 18, 2022. [Online]. Available: <https://react.dev> [retrieved: Mar. 2024].
- [12] Ryan Dahl, *Node.js*, version 18.19.0, 2023. [Online]. Available: <https://nodejs.org> [retrieved: Mar. 2024].