# WordWeaver: Interactive System for Facilitating Digital Story Illustration

Maryam Kermanshahani[1], Sajad Shirali-Shahreza[1,*] and Gerald Penn[2,*]

[1]Amirkabir University, No. 424, Hafez Ave., Tehran, 1591634311, Iran
[2]University of Toronto, 4283-40 St. George St., Toronto M5S 2E4, Canada

**Abstract**

WordWeaver simplifies the process of generating illustrations for stories. Users can tag sections of their story text and generate relevant images using the Stable Diffusion model based on these tags. The system enhances the illustration process by allowing flexible character placement (by tagging / labelling) and reducing repetition, thereby increasing speed and lowering time and cost. The system is evaluated through software testing (end-to-end) and user testing (meCUE questionnaire), where human subjects compare two image generation models: tag-based and direct prompt writing. The results indicate that the tag-based method performed better overall.

**Keywords**

Book Cover Art, Image Generation, Prompt Engineering, User Interfaces

## 1. Introduction

Story visualization is one of the most crucial aspects of storytelling[1], which helps to illustrate the content of a story or design a book cover. Although the content of the story is important, the cover image also plays a crucial role in attracting readers[2]. A book cover is designed to give readers clues about the book's content, serving as a form of communication between the author and the reader[3]. Since human perception is largely visual[4], illustrations help improve comprehension and retention while making stories more engaging. Studies suggest that the strongest and most meaningful messages are conveyed through a careful combination of words and images[5].

One of the key challenges in story visualization is selecting or generating an image that matches the story's content. Generating a sequence of coherent and meaningful images from natural language stories requires understanding and reasoning in both language and images[6]. The ideal image should reflect both the author's and the reader's vision of the story. A user can try to use text-to-image generators to create images for a written story, but this usually requires multiple rounds of tuning the generation prompt. Rewriting prompts, especially long ones, to achieve the desired image can be exhausting. To address this, we developed a system in which users can tag sections of text and generate images by selecting tags, eliminating the need to repeatedly write lengthy prompts. Additionally, since each tag's text is stored consistently, the images are more cohesive.

Our project aims to design and implement a system that helps authors generate images for their stories, with a focus on software techniques. This involves creating a user interface where users can tag different sections of text within a potentially long narrative and generate images based on these tags, making the process easier and more enjoyable. The system uses the Stable Diffusion image generation model, running on a private server, to create images from the prompts.

This system can greatly help independent authors speed up the self-publication process and reduce costs. In addition, beyond its benefits for storytellers and writers, it can find applications in other fields.

*[PromptEng] Third International Workshop on Prompt Engineering for Pre-Trained Language Models co-located with the ACM WebConf, April.13, 2026, Dubai, United Arab Emirates*

*Corresponding author.
maryam.kermanshahani1379@gmail.com (M. Kermanshahani); shirali@aut.ac.ir (S. Shirali-Shahreza); gpenn@cs.toronto.edu (G. Penn)
https://seriousxr.ca/maryam-kermanshahani/ (M. Kermanshahani); https://sajad.shirali.ir/ (S. Shirali-Shahreza); http://www.cs.toronto.edu/~gpenn/ (G. Penn)
0009-0004-0741-8562 (M. Kermanshahani); 0000-0003-1654-6385 (S. Shirali-Shahreza); 0000-0003-3553-8305 (G. Penn)

For example, it can be used to enhance educational content for children, or enable social media users to create relevant images corresponding to a given text or story and share them on their pages.

Our aim is to address some of the challenges that authors face when they want to use image generators such as Stable Diffusion:

- Difficulty of setting up the tools for personal use.
- "Learning curve" for how to write proper prompts that obtain acceptable results.
- Producing consistent styles of story illustration.
- Generating images that match a pre-conceived vision of a scene or story.
- A user-friendly interface for novices.

## 2. Related Work

Today, DALL-E, Stable Diffusion, and MidJourney are the leading models for text-to-image conversion[7]. DALL-E, developed by OpenAI, was launched in 2021 and allows users to input phrases and receive corresponding images. Stable Diffusion, released in 2022 by Stability AI, is an open-weight model widely used for image generation. MidJourney, which operates through Discord, works similarly but focuses more on animé-like or surreal images.

While these models have their own strengths and weaknesses, Stable Diffusion often outperforms the others, especially in facial generation. Some of the features of other models, like MidJourney's animé style and DALL-E's focus on single-subject images, contribute to their limitations.

There have been studies on generating book covers using AI. For example, one project gathered a new dataset and trained a model. While it could not produce professional-quality covers, it served as a source of creative inspiration[2]. Another platform, Books-by-AI[1], tries to use GANs to generate book covers.

There are also efforts to generate illustrations for story content. One project used generative models to create stories with minimal user input, focusing on both text and image generation[8]. Another introduced a framework using diffusion and visual memory modules to ensure consistent characters and scenes across illustrations[9]. A third one developed a new sequential story-to-image model called StoryGAN, which enhances image quality and consistency with a deep context encoder[6]. There is also research that has been conducted on prompt engineering. One study created an interactive system called "Promptify" to improve text prompts when using Stable Diffusion[10].

As mentioned above, research on generating images for stories exists, including short text-to-image generation and book cover creation. To our knowledge, however, our project is the first to introduce the idea of tag-generation and usage to simplify the image generation process.

Unlike direct use of image-generation models like Stable Diffusion that requires advanced skills (e.g., proper prompt generation), our system is designed to be more accessible to novice authors. Secondly, our system allows users to generate a sequence of images for a full story, rather than just a single image per phrase, by easily changing the composition of the image. By tagging sections of the text, users can create images quickly without writing long prompts and easily investigate different scene composition. Furthermore, since tags are stored consistently, the resulting images are more cohesive as they compare different combination of tags. The system also combines various tools in one place, making the illustration process faster, easier, and more enjoyable. Through an iterative user interaction, the system helps users achieve results closer to their expectations.

Finally, while previous systems mainly focused on training artificial intelligence models, we have improved story illustration by using existing models, and have instead focused on the user experience.

[1] https://booksby.ai/

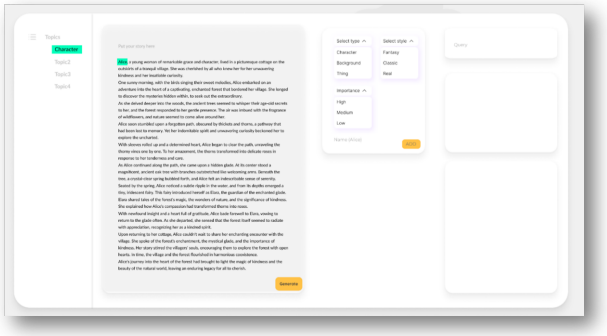

**Figure 1:** Initial proposed designed of WordWeaver in Figma.

## 3. WordWeaver: Crafting Covers and Images for Stories

In this section, we describe the path that we followed to create WordWeaver. We begin by discussing the design and implementation process, outlining the key decisions and techniques used in developing the system. Next, we explain the interface and navigation design and describe how users interact with the system and how usability principles guided the interface design. Then we discuss prompt generation. Together, these components provide a comprehensive overview of how the system was developed.

### 3.1. Design and Implementation

The design phase begins with the creation of user interface designs and wireframes in Figma, as shown in Figure 1. This helps to structure the interaction flow and plan the core features of the system. Building on this foundation, the system includes a front end built with React and Tailwind CSS, and a back end developed using ExpressJS. The back end uses the Stable Diffusion XL model to generate images based on user input.

### 3.2. Interface and Navigation

The WordWeaver interface is organized around five icons positioned on the left-hand side of the screen (Figure 2). Four of these icons correspond to the main functional pages of the system, while the fifth provides access to an interactive tour guide. This structure allows users to easily navigate through different stages of the workflow.

- **Welcome Page**: Provides an overview of the system and its purpose (Figure 2).
- **Image Generation Page**: This is the core functionality of the system. The users will input their story text here (Figure 3). Upon submitting the story, the interface automatically transitions to the actual section for image generation (Figure 4). The user can select tag categories such as *Person*, *Thing*, and *Background*, highlight parts of the story to assign these tags, and choose an image style (e.g., "3D Model," "Analog Film," "Anime"). The system supports multiple tags, and based on the selected tags and styles, it constructs a textual prompt. Clicking the "Generate Image" button sends this prompt to the back end (which is running Stable Diffusion XL model) to produce one or more images, which can be saved for later use.
- **Comparison Page**: This was designed to emulate the interface of Automatic1111, an open-source GUI for Stable Diffusion. It includes a single text-prompt window, the most common approach in image generation UIs. This page enables comparison between WordWeaver's tagging interface and existing prompt-based tools. Figure 5 presents the original Automatic1111 web-based user interface and its functionalities. By contrast, Figure 6 shows our implementation of this interface, developed to facilitate comparison and evaluation. We hide the various parameters that the user can set in the Automatic1111 system to provide a fairer comparison with our system. Showing

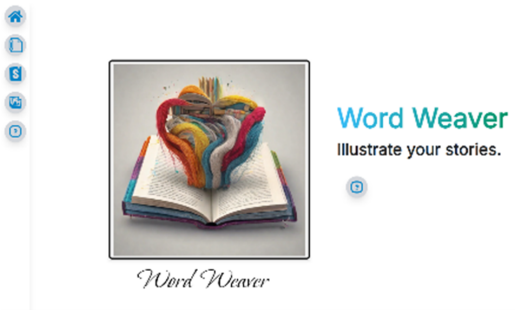


**Figure 2:** Welcome page

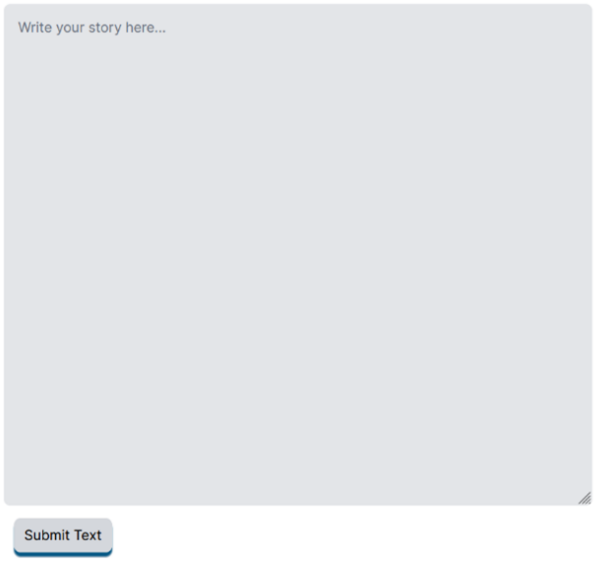


**Figure 3:** Entering text page.

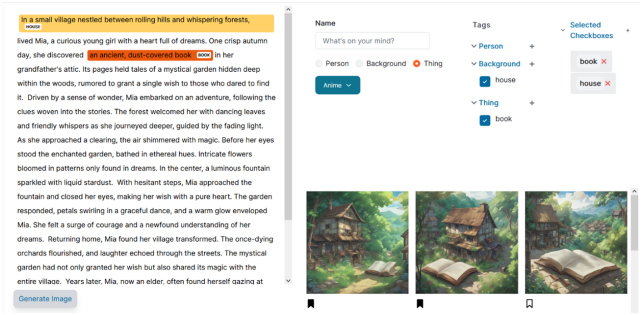


**Figure 4:** Creating tags for concepts.

too many parameters to a novice user can overwhelm them and results in a lower score which is not related to entering a text prompt, in comparison to our tag-based approach.

- **Gallery Page**: This page displays the images saved by the user during the generation (Figure 7). This enables the user to try different combinations of tags, selects those that seem better from each combination, and at the end selects the best images among them.
- **Tour Guide**: The fifth icon activates a built-in tutorial implemented using a pre-built library. It introduces users to each component of the system and supports onboarding by providing contextual guidance throughout the interface.

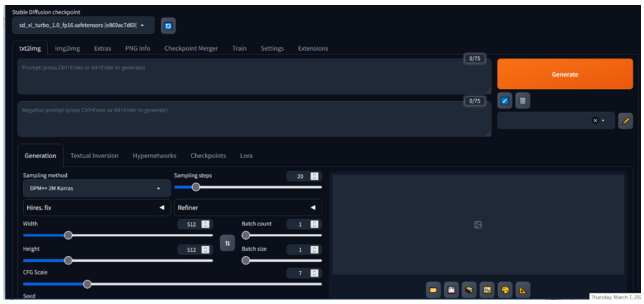


**Figure 5:** The Automatic1111 web-based user interface for Stable Diffusion.

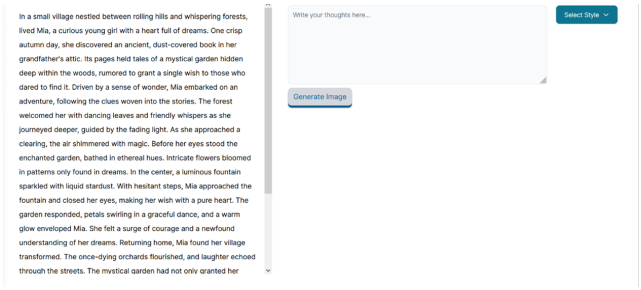


**Figure 6:** Prompt-based Image Generation Option.

**Figure 7:** Gallery.

### 3.3. Prompt Generation

Information is transmitted to the Stable Diffusion XL model through a dynamically generated text prompt. This prompt is constructed by a function that interprets user-defined tags and their corresponding textual descriptions. For instance, a user aiming to generate an image of an historical scene might annotate various components such as “a medieval knight” as Person, “a sword” as Thing, and “a stone castle” as Background, while also specifying Painting as the intended artistic style. Based on these annotations, the system formulates a textual prompt, e.g., “A medieval knight and a sword in a stone castle.” This prompt also includes additional keywords such as selected visual style. It is then submitted to the back end through our defined API to produce the candidate images.

## 4. Evaluation

We evaluate our system by two methods: Software End-to-End testing and a pilot User testing. For End-to-End testing, we used Cypress. Cypress is an end-to-end test automation tool for graphical user interfaces of web applications[11]. We tested the main process of generating images for the story by tagging tests and the tests passed without any errors.

In our pilot testing, we aimed to assess the ease of use and efficiency of the system with actual users. Although this was only the pilot stage (4 colleagues tested our system), the experimental design has

been validated and completed. We compared our new tag-based system with a baseline setup where only a text box is provided for the user to write their own prompts. This baseline mirrors the common setup in text-to-image systems such as Stable Diffusion, DALL·E, and even MidJourney. To this end, as mentioned in the previous chapter, we added another page to the interface that contains only a text box for writing prompts.

We first used the system to generate images for a story by both methods. Then we asked users to fill out a questionnaire. This form is based on the meCUE[12] questionnaire. The meCUE, AttrakDiff, and UEQ questionnaires are among the most well-known for evaluating user experience[13].

## 4.1. meCUE Evaluation

The meCUE questionnaire assesses all the key elements of user experience together. In fact, this questionnaire covers the highest number of common evaluation factors across various UX questionnaires[14]. This standardized questionnaire is based on the Component Model of User Experience (CUE) framework, designed to measure key aspects of user experience (UX) for interactive products.

It consists of four modules:

- **Product perception:**
  - Usefulness, usability.
  - Visual aesthetics, status, engagement.
- **Emotions:** Positive emotions, Negative emotions.
- **Consequences:** Product loyalty, Intention to use.
- **Global:** Overall judgment.

Each module is instrumented with a Likert scale and can be used separately or in combination with other modules. The final version of the questionnaire includes 34 items categorized into 10 scales. The fourth module, which captures the overall judgment about the system, is rated on a scale from -5 to 5, while the other three modules are rated on a 1 to 7 scale: strongly disagree, disagree, somewhat disagree, neutral, somewhat agree, agree, and strongly agree.

After answering the initial demographic questionnaire, each subject first chooses an English story from the internet, generates images for their story, and then fills out the evaluation form twice: once for the system that generates images through label tagging and then for the baseline system. The evaluation data are stored in an Excel file, in a workflow based on the method described in [12].

The results of the pilot evaluation are shown in Table 1. Although our limited pilot size made it impossible to prove conclusions, we list our preliminary observations:

- **Product perception:**
  - Usefulness vs. usability: The image generation page that requires direct entry of a prompt (i.e., the baseline method) scored higher in terms of usability (i.e., easier and more efficient), while the label-tagging system (our proposed system) scored higher in terms of usefulness (a value judgement about the purpose of an affordance). This is mainly because our peers elected to pipe long segments of the story into the generation page as a prompt. The length of the prompt was limited only by the underlying transformer window of the large language model (considering the common technique of splitting very large prompts into 75-token chunks).
  - Visual aesthetics, status, engagement: The label-tagging system (our system) received higher scores in all three categories.
- **Emotions:** Both positive and negative emotions were more prominent for the label-tagging system (our system). Additionally, for both systems, the positive emotions outweighed the negative ones.
- **Consequences:** Our label-tagging system scored higher in both product loyalty and future intent to use.
- **Global:** In the overall evaluation, our label-tagging system also received a slightly higher score.

**Table 1**
Comparison of average muCUE scores between proposed labing interface and baseline text-base prompts.

| Module | Subscale | Our Proposed Labelling Approach | Baseline Text-Based Prompt |
|---|---|---|---|
| Module I | Usefulness | 5.83 | 5.00 |
| | Usability | 3.00 | 3.75 |
| | Visual Aesthetics | 5.08 | 4.75 |
| | Status | 5.08 | 3.92 |
| | Commitment | 3.92 | 3.50 |
| Module II | Positive Emotions | 5.42 | 5.25 |
| | Negative Emotions | 2.79 | 2.04 |
| Module III | Intention to Use | 4.58 | 4.25 |
| | Product Loyalty | 4.58 | 4.17 |
| Module IV | Overall Evaluation | 3.40 | 2.80 |

## 5. Conclusion

The project's goal is to simplify the generation of images for stories. We propose a new system that is based on the idea of tagging the important parts of the story by tags and then use tags to generate images. Using the tagging tool, our system allows users to generate images for specific parts of their stories more easily and faster than previous systems. Our system is being evaluated through both software testing and a pilot user testing.

We compared our system with a baseline option of writing the whole prompt in a text prompt (to mimic the common UI of image generation tools). The baseline system seems to perform slightly worse overall so far, and better only in usability and in eliciting slightly fewer negative emotions. Users seems to prefer the label-tagging system overall, although it has a steeper startup cost in terms of learnability.

The system is not distributed as a ready-made, installable package yet, but we are planning to do that. There are also several avenues for improvement. Enhancing the generated queries could improve user satisfaction, as prompt quality directly impacts AI model's performance. A complete evaluation must still be conducted. As a consequence of this pilot study, perhaps some assessment criteria could be refined to provide more accurate insights.

Furthermore, increasing image consistency across a story —– especially in characters and settings —– is crucial. This is an important and challenging problem that needs more research in the future. New editing tools such as image editing features and resizing options can also be added. Finally, improving the AI model to avoid visual anomalies and generate more aesthetically pleasing images is a key area for future research.

## Acknowledgments

This research was supported by a grant from the Naver Corporation.

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

[1] M. Balabanović, L. L. Chu, G. J. Wolff, Storytelling with digital photographs, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00, Association for Computing Machinery, New York, NY, USA, 2000, pp. 564–571. doi:10.1145/332040.332505.

[2] E. Haque, M. F. K. Khan, M. I. Jubair, J. Anjum, A. Z. Niloy, Book Cover Synthesis from the Summary, in: Proceedings of the 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications, AICCSA '22, 2022, pp. 1–6. doi:10.1109/AICCSA56895.2022.10017541.

[3] W. Zhang, Y. Zheng, T. Miyazono, S. Uchida, B. K. Iwana, Towards Book Cover Design via Layout Graphs, in: J. Lladós, D. Lopresti, S. Uchida (Eds.), Document Analysis and Recognition, ICDAR '21, Springer International Publishing, Cham, 2021, pp. 642–657. doi:10.1007/978-3-030-86334-0_42.

[4] S. Nag Chowdhury, W. Cheng, G. de Melo, S. Razniewski, G. Weikum, Illustrate Your Story: Enriching Text with Images, in: Proceedings of the 13th International Conference on Web Search and Data Mining, WSDM '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 849–852. doi:10.1145/3336191.3371866.

[5] Enjellina, E. V. P. Beyan, A. G. C. Rossy, A Review of AI Image Generator: Influences, Challenges, and Future Prospects for Architectural Field, Journal of Artificial Intelligence in Architecture 2 (2023) 53–65. doi:10.24002/jarina.v2i1.6662.

[6] Y. Li, Z. Gan, Y. Shen, J. Liu, Y. Cheng, Y. Wu, L. Carin, D. Carlson, J. Gao, StoryGAN: A Sequential Conditional GAN for Story Visualization, in: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR '19, 2019, pp. 6322–6331. doi:10.1109/CVPR.2019.00649.

[7] A. Borji, Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2, 2023. doi:10.48550/arXiv.2210.00586. arXiv:2210.00586.

[8] S. Fotedar, K. Vannisselroij, S. Khalil, B. Ploeg, Storytelling AI: A Generative Approach to Story Narration, in: AI4Narratives — Workshop on Artificial Intelligence for Narratives, 2021, pp. 19–22.

[9] T. Rahman, H.-Y. Lee, J. Ren, S. Tulyakov, S. Mahajan, L. Sigal, Make-A-Story: Visual Memory Conditioned Consistent Story Generation, in: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR '23, 2023, pp. 2493–2502. doi:10.1109/CVPR52729.2023.00246.

[10] S. Brade, B. Wang, M. Sousa, S. Oore, T. Grossman, Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1–14. doi:10.1145/3586183.3606725.

[11] M. T. Taky, Automated Testing With Cypress, Bachelor's thesis, VAMK University of Applied Sciences, 2021.

[12] M. Minge, M. Thüring, I. Wagner, Developing and Validating an English Version of the meCUE Questionnaire for Measuring User Experience, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 60 (2016) 2063–2067. doi:10.1177/1541931213601468.

[13] I. Díaz-Oreiro, G. López, L. Quesada, L. A. Guerrero, Standardized Questionnaires for User Experience Evaluation: A Systematic Literature Review, Proceedings 31 (2019) 14. doi:10.3390/proceedings2019031014.

[14] A. Hinderks, Applicability of User Experience and Usability Questionnaires, Journal of Universal Computer Science 25 (2019) 1717–1735.