# Enhancing Collaborative Storytelling for People with Dementia through AI-Based Media Generation

**Rares Burghelea**

**Supervisor(s): Paul Raingeard de la Blétière, Mark Neerincx**

**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Rares Burghelea
Final project course: CSE3000 Research Project
Thesis committee: Mark Neerincx, Paul Raingeard de la Blétière, Christoph Lofi

## Abstract

Storytelling is a powerful non-pharmacological intervention in care for People with Dementia (PwD), offering the opportunity of self-expression, emotional connection and identity preservation. However, creation of multimedia material to accompany such stories usually requires trained experts and a considerable effort, which limits scalability and personalization. This paper presents a modular system using Artificial Intelligence (AI) which, together with a social robot, aims to transform collaboratively written stories into personalized images and songs. Using a local Large Language Model (LLM) for prompt generation and external diffusion models for media synthesis, the system also supports an interactive feedback system in which users can refine the output. Because of ethical constraints related to data confidentiality, fully autonomous music generation was not implemented, however, the system allows users to generate music manually with any external AI model, automatically detecting the song and playing it in the storytelling interaction flow. The system was tested using simulated stories on consumer-grade hardware, as such confirming the feasibility of media generation in real time, while also respecting confidentiality. The results show strong alignment between the story content and generated visuals, supporting the system's potential to enrich dementia storytelling experiences in an ethical and accessible way.


## 1 Introduction

Dementia is an umbrella term describing a range of neuro-degenerative conditions characterized by progressive cognitive decline, including impairments in memory, communication, and daily functioning [4][6][10]. As global populations age, the prevalence of dementia is rising sharply, with over 55 million people affected worldwide and numbers projected to triple by 2050 [12][10]. While pharmacological treatments offer limited relief, non-pharmacological interventions, such as reminiscence therapy, art therapy, music therapy and storytelling, have been proven to provide many benefits, such as well-being, better communication and more emotional connection for People with Dementia (PwD)[2][11][4].

Among these, collaborative storytelling has emerged as a meaningful, failure-free activity, where PwD can express themselves and create a story, promoting social engagement, collaboration and identity preservation, despite cognitive challenges [3][5]. However, implementing such storytelling activities at scale remains resource-intensive, requiring trained professionals, considerable preparation and a significant amount of time for the media generation [9]. As a result, many care settings face such difficulties when trying to offer personalized, significant and dynamic storytelling activities, despite the many known benefits.

Recent advances in artificial intelligence (AI) and social robotics present a promising opportunity to address these challenges by automating and enhancing storytelling activities for people with dementia. Socially assistive robots have been explored in dementia care to support companionship, cognitive stimulation, and reminiscence therapy, with encouraging results for improving mood and engagement [8]. However, their potential to facilitate collaborative storytellin, particularly through the integration of AI-generated multimedia, remains underexplored. Large language models (LLMs) and generative AI tools, including text-to-image diffusion models and AI-based music generators, now make it feasible to automatically create personalized images and songs from a textual story description [10][13][1]. These technologies offer new possibilities for real-time, interactive, and multi-modal storytelling experiences tailored to the individual's narrative and feedback, reducing technical barriers and manual workload.

Music, in particular, has shown strong therapeutic value in dementia care. Studies have found that listening to familiar or emotionally resonant music can stimulate memory, reduce agitation, and promote mood regulation in PwD, even in advanced stages of cognitive decline. Unlike visual media, music can be passively consumed and still elicit strong emotional responses, making it a valuable tool for non-verbal engagement. Therefore, enabling the automatic generation of personalized music from user-created stories holds unique potential to deepen the emotional and therapeutic impact of collaborative storytelling interventions.

Despite growing evidence for the benefits of creative storytelling, current approaches face significant limitations in practice. Programs such as TimeSlips and other creative expression interventions have been shown to improve engagement, mood, and communication in people with dementia [3][11], yet they typically require trained facilitators, in-person group settings, and considerable preparation time [9]. Additionally, producing multimedia materials to accompany stories, such as images or music, often involves complex tools like video editing software, taking weeks of effort from skilled professionals [9]. While programs like TimeSlips use pre-existing images to prompt imagination, such static content limits personalization and may not reflect the unique narrative created by the participants. These resource demands limit the scalability and personalization of storytelling interventions in many care settings. As a result, there is a need for more accessible, efficient, and adaptive storytelling solutions that can generate meaningful multimedia content in real time and reduce reliance on manual production.

This project forms part of a broader research initiative involving the development of an integrated social robot system for dementia care. Within this joint effort, individual contributions focus on different modules such as GUI development, story facilitation, and evaluation. The present work specifically addresses the automatic generation of

personalized images and music from collaboratively created stories, forming the media generation component of the overall system.

This research focuses on the development of an AI-based system for generating personalized music from collaboratively created stories involving a person with dementia and a conversation partner. The central research question guiding this work is: *How can AI generate personalized songs from a collaboratively created story and refine the output based on participant feedback?* Involving participants in the refinement process is essential for ensuring that the generated outputs are emotionally resonant, contextually appropriate, and aligned with the storytelling goals. To answer this question, the project investigates methods for extracting key narrative elements from the story input, selecting suitable AI models for music generation, and designing mechanisms to incorporate user feedback to iteratively improve the musical output. By integrating generative AI with an accessible feedback loop, this work aims to contribute to more personalized, engaging, and meaningful storytelling experiences for people with dementia. The remainder of this paper describes the background, methodology, and evaluation plan for this system.

## 2 Related Work

This section reviews previous work on storytelling interventions for people with dementia (PwD), as well as the application of generative AI models in the creation of multimedia content such as images and music. Additionally, the role of feedback and personalization is examined in relation to maintaining the emotional relevance of generated outputs within dementia care settings.

### 2.1 Creative Storytelling for Dementia Care

Creative storytelling has emerged as a valuable non-pharmacological intervention in dementia care, offering opportunities for PwD to express themselves, maintain identity, and foster social engagement despite cognitive decline. While TimeSlips [3] and similar storytelling programs [2][4] have demonstrated measurable psychological and social benefits, their effectiveness is deeply reliant on experienced facilitators. This dependency introduces two essential limiting factors: (1) the emotional depth of the activity heavily depends on the facilitators' skills [11] and (2) the content remains mostly generic due to the preselected prompts [5]. These limitations suggest that despite the model being effective, it lacks scalability and adaptability for personal narrative expression, an area where AI-generated media could offer an innovative alternative.

### 2.2 AI and Multimedia Generation for Storytelling

Recent advances in generative AI have made it feasible to dynamically produce personalized media, such as images and music, from textual descriptions. Text-to-image diffusion models (e.g., Stable Diffusion, DALL·E 2) and text-to-music systems (e.g., MusicLM, Bark) offer powerful tools for enriching narrative experiences [13][1]. These tools dramatically reduce the time, effort, and expertise needed to produce customized content and open new opportunities for accessibility and personalization in care settings.

However, in dementia-specific applications, such generative systems remain underutilized. Rios-Rincón et al. [9] review existing digital storytelling tools and note that most rely on static multimedia or manual input, lacking adaptive personalization. Furthermore, existing tools often fail to integrate feedback from users in real-time, which is critical for maintaining emotional and contextual relevance.

### 2.3 Feedback, Personalization, and Ethical AI Use

Emotional congruence is central to effective storytelling with PwD. The AI-generated output must be not only coherent but also emotionally appropriate. This underscores the need for feedback loops in which caregivers or PwD can guide the refinement of generated media. Previous work [10] emphasizes the potential of large language models (LLMs) to support co-creative storytelling, but also cautions against ethical pitfalls, particularly regarding data privacy and model transparency.

To address these challenges, the system integrates a feedback mechanism that allows the user to iteratively refine the generated content according to their preferences. Previous studies rarely include mechanisms that allow real-time personalization or feedback-based refinement [9], which limits the emotional adaptability and user autonomy in these types of systems. Moreover, the ethical trade-offs in relying on commercial AI APIs are often under-examined, in spite of their major implications in trust, data ownership and sustainability in sensitive care context [10]. These gaps directly motivate the central question in the proposed research: How can AI images and songs be generated in an emotionally resonant way, while preserving privacy and being responsive to participant input.

## 3 Methodology

This project implements a multi-modal AI system that transforms collaboratively created stories into personalized images and songs. The system operates in three stages: (1) story input and processing, (2) AI-driven media generation, and (3) refinement via user feedback. It forms the media generation component of a broader research project, in which different group members develop complementary modules. In particular, this system is designed to integrate with a dementia-friendly user interface, which facilitates story writing and interaction in an accessible way. The architecture includes a web-based interface, a Flask backend with WebSocket support, and external generative models accessed via Hugging Face APIs and a local large language model (LLM). Below, the methodology is further described in more detail.

#### Story Input and Interface

Stories are collaboratively written in real-time by a person with dementia, their conversation partner, and the robot, using a keyboard interface. At this stage, the story is typed into a text box, while the robot participates by speaking, asking guiding questions, or making suggestions with human approval. While the implementation of the story interaction logic is outside the scope of this paper and handled by a sep-

arate project, the story text produced is used as the sole input to the media generation modules described below.

**Image Generation**
Upon receiving a story, the backend initiates a two-step image generation pipeline. First, it uses the `ollama` library to query a locally hosted LLM (Gemma 2) with a system prompt [7] designed to extract vivid visual details from the story (see Appendix A). This results in a visually descriptive text prompt. The prompt is then forwarded to Hugging Face's inference API[1] to generate an image using the Stable Diffusion 3 Medium model [13][2]. The resulting image is stored locally and returned to the frontend in base64-encoded format for immediate display.

**Music Generation**
While the system architecture was designed to support automatic music generation from story descriptions, this feature was not fully integrated due to technical and ethical constraints. Specifically, while several state-of-the-art text-to-music models exist, such as MusicLM [1] or proprietary services like Suno, they are either unavailable for public or academic use, or require cloud-based access to external servers.

A core design principle of this project was to minimize data exposure, particularly given the sensitivity of working with people with dementia. As noted in prior work on AI in dementia care [10][9], systems that transmit user-generated content, such as speech or personal narratives, to third-party APIs must take special care to avoid privacy breaches, misuse, or unintended retention. To uphold these principles, the use of local, self-hostable models was prioritized, and direct API-based generation was not implemented.

Instead, the current system allows users or caregivers to generate music manually using an external tool of their choice. To support this, the system generates a detailed music prompt, based on the story's emotional tone, style, and instrumentation, which the user can copy and paste directly into their preferred AI music generation service. Once the resulting song is placed in a designated folder, the system automatically detects it and plays it as part of the storytelling flow. This approach enables integration of high-quality music generation while respecting privacy requirements. The architecture remains modular, and a local music generation model, should one become available with sufficient quality, could easily be integrated in future iterations. Ultimately, future work should explore privacy-preserving pipelines for generating emotionally resonant music in real-time within sensitive care contexts.

**Feedback and Iterative Refinement**
Users can provide free-form feedback on the generated multimedia through a dedicated text field. When feedback is submitted, the system uses both the original story and the feedback to generate a revised visual or audio prompt. This is accomplished by prompting the Gemma 2 model with an instruction that preserves the core content of the original scene while integrating the suggested changes. The prompt engineering strategy follows a consistent structure: it first reminds the model of its prior visual or musical interpretation, then instructs it to adapt that output using the new feedback. This strategy ensures continuity in the generated content while allowing targeted adjustments. The refined prompt is then used to regenerate the multimedia via the same models, and the updated media is returned to the front-end and displayed to the user. Full example prompts, including initial and feedback-refined variants, are included in Appendix A.

The dialogue system also manages interaction flow through predefined phases (e.g., greeting, storytelling, and music playback), which help structure the robot's behavior and ensure context-aware responses.

### 3.1 Alternative Approaches Considered

Several alternative approaches were considered before settling on the final system configuration. For prompt generation, multiple large language models were tested, including Gemma 2 (local), Gemini (cloud-based), ChatGPT-3, and ChatGPT-4. These models were compared across three dimensions: generation speed, semantic consistency between the story and prompt, and overall prompt quality. Commercial models such as ChatGPT-4 consistently produced more elaborate and coherent prompts with lower latency. However, their use posed ethical concerns due to their reliance on cloud infrastructure, data retention policies, and the lack of transparency regarding model behavior. These limitations, coupled with subscription costs and potential violations of data minimization principles, rendered them unsuitable for use in this project, particularly given the sensitive nature of dementia-related storytelling. A more detailed discussion of these ethical considerations is provided in Section 6.

## 4 System Architecture

### 4.1 Overview

The system is a multi-modal AI service that collaborates with a social robot to support storytelling and media generation for people with dementia. It is implemented as a separate application that communicates with the Navel Robot platform via networked API endpoints. The external system handles story interpretation, prompt generation, and multimedia synthesis, while the robot remains responsible for vocal interaction, dialogue management, and playback of generated media.

Users interact with the robot through a shared interface, typically a touchscreen tablet, that allows them to type or select story elements, listen to the robot's suggestions, and provide feedback on generated content. Once a short narrative is collaboratively created, the robot triggers a call to the external system, sending the story text.

On the backend, a Flask-based server receives the story and initiates the media generation pipeline. A local large language model (Gemma 2), accessed through the Ollama API, transforms the narrative into a detailed visual prompt and a descriptive musical prompt. These prompts are produced using predefined templates that guide the LLM to extract relevant visual or musical elements, as described in Appendix A. The visual prompt is sent to a Hugging Face-hosted Stable Diffusion model to generate an image. For music, due to ethical and technical constraints, the system does not generate

[1] https://huggingface.co/inference-api

[2] `stabilityai/stable-diffusion-3-medium-diffusers`

audio directly. Instead, users or caregivers manually generate a song using a third-party tool of their choice (e.g., Suno), and the system automatically detects and plays the resulting audio file within the interaction flow (see Section 3). This design preserves privacy by avoiding the transmission of sensitive data to cloud-based services while maintaining flexibility for future integration of local music generation models.

The generated image and song are returned to the robot interface. The robot then vocalizes a response and displays the image while playing the melody. Optionally, the user or caregiver may provide free-form feedback, which is used to regenerate improved content through the same pipeline. This feedback loop supports iterative personalization and fosters a sense of co-creative authorship.

Figure 1 illustrates the system architecture and data flow between the robot, user interface, AI backend, and media generation services.

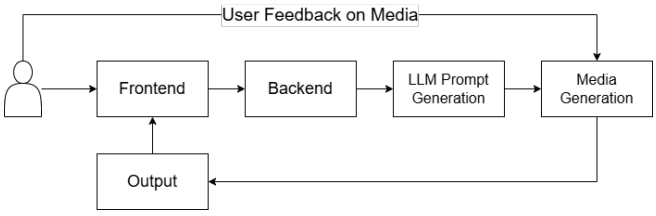


Figure 1: System architecture showing the interaction between user, backend, LLM prompt engine, and media generation models.

# 5 Evaluation

This project aimed to explore the technical feasibility of using local language models for prompt engineering in combination with external generative models to create personalized multimedia based on collaboratively authored stories. Given the exploratory and systems-oriented nature of this work, as well as ethical constraints surrounding the inclusion of real users with dementia, a qualitative evaluation was conducted based on simulated story sessions, focusing on three key capabilities: multimedia relevance, generation latency, and feedback-driven refinement.

## 5.1 Methodology

Four fictional short stories were designed mimicking the kind of emotionally expressive and visually descriptive narratives commonly observed in dementia storytelling sessions. Each story involved simple, evocative elements, such as a pet's birthday, a walk through a snowy village, or a seaside memory, chosen to evaluate the system's ability to extract meaningful prompts and produce fitting media.

The evaluation consisted of three main components:

- **Media Accuracy:** For each story, the generated image was compared against the story's key elements, e.g., setting, mood, characters, and objects, assessing semantic alignment and emotional congruence.
- **Feedback Responsiveness:** User-like feedback was submitted (e.g., "make the sky more dramatic" or "add another person in the scene") and reviewed whether the updated image reflected the refinement while preserving the core story context.
- **System Performance:** The average generation time was recorded for images via the Hugging Face inference API and noted resource usage across runs.

## 5.2 Results and Insights

**Image Generation.** Across all test cases, the system produced images that were thematically and visually consistent with the input stories. Core story details, such as specific objects ("a red dress", "two golden balloons"), environments ("foggy lake", "snowy hilltop"), and emotions ("quiet joy", "loneliness")—were captured with reasonable fidelity in the generated images. The system tended to preserve atmosphere and structure reliably, though occasional hallucinations (e.g., extra characters or missing accessories) occurred in 1 out of 4 images. Average image generation latency was 56.3 seconds per story, consistent across runs. Example images illustrating initial outputs and feedback-driven refinements are provided in Appendix B.

**Feedback Refinement.** The feedback loop proved effective in adjusting generated outputs. For example, when the story described "a boy watching birds from the hill" and feedback was added to "include more birds in the sky," the regenerated image successfully increased the bird count while maintaining similar composition. Semantic edits (e.g., changing weather or mood) were generally well-integrated. However, extremely detailed or abstract feedback occasionally produced minimal changes, suggesting a need for more precise prompt conditioning.

Across the four stories, two rounds of feedback were tested per story, targeting both low-level visual features (e.g., adding objects or changing colors) and high-level semantic traits (e.g., adjusting mood or emotional tone). The system consistently responded to direct, concrete feedback such as object presence, size, or number, while abstract feedback (e.g., "make it feel more peaceful") produced more subtle or ambiguous changes. This suggests the system is more reliable at integrating tangible scene modifications than interpreting nuanced emotional refinements—an important consideration when designing feedback prompts for end-users.

**Key Lessons.** These findings support the system's viability for use in dementia storytelling applications, particularly for producing emotionally aligned and story-consistent images. The local LLM-based prompt generation enabled controllable, privacy-conscious media creation, while the feedback mechanism provided a low-barrier way to iteratively personalize output. The results also highlight that user-facing prompt quality, not just model architecture, plays a critical role in controlling output fidelity.

In future deployments, tighter integration between prompt structure, user feedback types, and model capabilities could further improve controllability and reduce generation errors. While the music generation component was not fully automatic, the architecture demonstrated that manual integration with user-generated music is feasible and modular

## 6 Responsible Research

This project engages with a vulnerable population, people with dementia (PwD), and therefore raises important ethical and technical considerations related to data protection, user agency, and emotional safety.

### Data Minimization and Privacy

One of the central ethical concerns in digital dementia care is the protection of sensitive personal information. Prior research has consistently shown that both PwD and their supporters value data agency above all else, including privacy, transparency, and the ability to delete data on demand [10]. To respect these preferences, the system was explicitly designed to minimize data exposure. No user-generated content (e.g., story text, feedback, or interaction logs) is transmitted to third-party servers without strict data-handling guarantees.

In particular, music generation was not fully implemented due to this ethical constraint. While high-quality models like MusicLM or Suno exist, they require cloud-based access and offer no guarantee of secure or transparent data handling. Instead, only local models (Gemma 2 via Ollama) were used, and cloud APIs were deliberately avoided. This choice ensures that storytelling data, often rich with personal or emotionally resonant content, remains private and does not leave the device or local server.

### Accuracy and Relevance of AI Output

Accuracy is particularly critical in the context of assistive technologies for PwD. Inaccurate or incongruent content can cause confusion, emotional distress, or disengagement. The literature emphasizes that emotionally mismatched outputs, whether due to hallucinations in LLMs or overly generic imagery, can undermine identity and storytelling coherence [9][10]. To address this, the system integrates a user feedback loop that allows iterative refinement of the generated image (and ideally music), thereby promoting emotional congruence and alignment with the user's intent.

### User-Centered Design and Co-Creation

Another key principle in the ethical development of dementia care technologies is co-design. Successful systems must be tailored to the lived experience of PwD and their caregivers. Prior work has found that co-development helps avoid designing solutions that are either inaccessible or irrelevant [10]. The system embeds co-creation not just as a design philosophy but as an operational feature: users shape the story, guide the media generation through feedback, and co-author the outcome. This reinforces agency and helps avoid patronizing or one-size-fits-all interactions.

### 6.1 Reproducibility Considerations

Reproducibility was addressed through modular design, fixed prompt templates, and documented dependencies. The system consistently uses the same local LLM (Gemma 2 via Ollama) and Stable Diffusion 3 Medium model through Hugging Face's API. While image outputs may vary due to the random nature of generative models and lack of seed control in external APIs, prompts and story inputs are fully version-controlled. Environment details, including hardware requirements and model versions, are specified in the project's documentation.

All components, prompt generation, feedback refinement, media playback, are decoupled and testable independently, improving traceability and reproducibility. Example prompts and outputs are provided in Appendix A. While cloud-hosted model updates may affect output consistency over time, the system supports migration to local inference for full control in future work.

This setup reflects a conscious trade-off between feasibility and replicability, ensuring that core processes remain transparent, repeatable, and extensible.

## 7 Limitations

While this project demonstrates the feasibility of AI-assisted multimedia storytelling in dementia care, several limitations remain. First, the system was not evaluated with real users due to ethical and logistical constraints. All testing was conducted using simulated stories and self-evaluation. Second, although the architecture supports music generation, high-quality local text-to-music models were not available, and privacy concerns prevented the use of commercial APIs. Instead, music was generated manually by the user using an external AI tool. Third, while visual prompts generally yielded semantically relevant images, occasional mismatches or hallucinations occurred—particularly when abstract or overly detailed prompts were used. These issues reflect the current trade-offs in using small, privacy-preserving models and highlight the need for more robust prompt conditioning and model control mechanisms.

## 8 Discussion

This project set out to explore whether AI-generated multimedia could be meaningfully integrated into collaborative storytelling activities for people with dementia. The implementation shows that real-time image generation and responsive feedback mechanisms are technically feasible using lightweight, locally hosted models. Moreover, the architecture supports privacy-preserving deployment, an essential condition for sensitive settings such as dementia care.

The evaluation demonstrates that even with small, locally run language models, it is possible to generate semantically aligned prompts that drive high-quality image outputs. However, the system is sensitive to the structure and clarity of the input story. When user stories lack visual cues or contain abstract concepts, the resulting images may be less coherent or relevant. Additionally, while the feedback loop reliably captured low-level visual adjustments (e.g., color, object presence), it struggled with more abstract or emotional refinements. These limitations point to an important lesson: system performance hinges not only on model choice but also on how end-users are guided to structure stories and feedback.

The manual music integration workflow, driven by privacy concerns, demonstrates a trade-off between automation and ethical design. Rather than using powerful cloud-based services that risk storing sensitive narrative content, the system

prompts users to generate music via external tools and play it locally. This highlights a recurring theme: in dementia care, technical ambition must often be tempered by data conservancy. Future iterations should continue to explore privacy-preserving generative pipelines, including the use of differential privacy techniques or locally fine-tuned models trained on domain-specific data.

The project also confirms the value of modular, human-in-the-loop design. By separating the prompt generation, media creation, and feedback loop, the system remains adaptable to future model upgrades or deployment settings. These design choices support extensibility, new modalities (e.g., slideshows, evolving scenes, musical mood transitions) could be incorporated with minimal architectural changes. Similarly, robot dialogue could be grounded more deeply in narrative context, especially if future versions support turn-based storytelling.

Finally, while no real user testing was conducted, the system structure was informed by existing co-creation practices in dementia care. The emphasis on feedback and personalization aligns with literature showing the emotional and cognitive value of identity reinforcement and storytelling engagement for people with dementia. With appropriate ethical protocols and evaluation designs, future studies should assess user perceptions, accessibility, emotional outcomes, and caregiver support.

## 9 Conclusion

This project demonstrates the feasibility of using generative AI to support collaborative storytelling for people with dementia, with a particular focus on image and music generation from short narratives. By combining a local large language model, external generative services, and an interactive feedback loop, the system enables dynamic multimedia creation while respecting critical privacy and ethical considerations.

Although the system has not yet been evaluated with real users and music generation remains limited, the architecture provides a modular and extensible foundation for future development. The results suggest that even lightweight, locally hosted models can support expressive, personalized interactions that align with the goals of dementia care: maintaining identity, fostering engagement, and enabling joyful moments of connection.

Ultimately, this work offers a promising step toward responsible, human-centered AI systems that enrich the lives of people with cognitive impairments through co-creative technology.

## A Prompt Templates

The following predefined prompt templates are used in this project to extract either visual or musical descriptions from a collaboratively written story. These prompts are passed to a local large language model (Gemma 2) before media generation.

### STORY_TO_PROMPT

*You are a visual imagination assistant. I will give you a short, evocative story told by a person with dementia. Your task is to extract the key visual elements and generate a detailed, concrete, and visually rich description that can be used as a prompt for a text-to-image diffusion model. Focus on the visual setting, people, objects, and emotions. Keep it grounded in reality unless the story includes imaginative or surreal elements—then preserve their poetic tone. Describe the scene in one paragraph. Include specific objects, characters, colors, lighting, mood, and any background details that would help visualize the scene. Do not quote or reference the original story in your output. Story:*

### STORY_TO_MUSIC_PROMPT

*You are a musical imagination assistant. I will give you a short, evocative story told by a person with dementia. Your task is to extract the mood, tone, and key emotions from the story and generate a descriptive music prompt. The output should describe the genre, tempo, instruments, and overall mood for a music generation model. Keep it poetic, emotional, and precise. Do not quote or reference the original story in your output. Just describe the music that matches the atmosphere of the story.*

## B Example Outputs

**Figure 1:** Generated image based on a longer story that included the line: *"A little girl in a red dress stood at the edge of the foggy lake, tossing breadcrumbs to the ducks as the morning mist danced around her."* The system correctly captured the main scene elements, atmosphere, and emotional tone.

**Figure 2:** Initial image based on a longer story that included the line: *"A boy watching birds from the hill under a cloudy sky."* The scene lacks sufficient bird presence despite the prompt.

**Figure 3:** Regenerated image after feedback: *"Add more birds in the sky."* The update correctly increases bird count while preserving composition and tone.

## C Use of AI Assistance for the Writing Process

AI tools were used solely to improve the readability, clarity, formatting, and grammar of the text. No ideas, arguments, or content were generated by AI; all contributions were conceived and written by the author. All AI-assisted edits were manually reviewed and verified for accuracy and alignment with the intended meaning.

The assistance involved limited, task-specific prompting. Typical prompts included:

- `Please reformulate this paragraph to make it easier to read and correct any grammatical errors within it. [... Redacted due to length]`
- `What could logical subsections be for this chapter: [... Redacted due to length]`

## References


[1] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325*, 2023.

[2] Anne Basting. Creative storytelling and self-expression among people with dementia. In *Thinking about dementia: Culture, loss, and the anthropology of senility*, pages 180–194. Rutgers University Press, 2006.

[3] Thomas Fritsch, Jung Kwak, Stacey Grant, Josh Lang, Rhonda R Montgomery, and Anne D Basting. Impact of timeslips, a creative expression intervention program, on nursing home residents with dementia and their caregivers. *The Gerontologist*, 49(1):117–127, 2009.

[4] Barbara J Harmer and Martin Orrell. What is meaningful activity for people with dementia living in care homes? a comparison of the views of older people with dementia, staff and family carers. *Aging & Mental Health*, 12(5):548–558, 2008.

[5] Seoyoun Kim, Kyong Hee Chee, and Olga Gerhart. Generativity in creative storytelling: Evidence from a dementia care community. *Innovation in Aging*, 4(2):1–7, 2020.

[6] Jinlong Ma, Qian Wang, Yanmei Lang, Shi Lv, Yuzhen Xu, and Baojian Wei. Effectiveness of creative story therapy for dementia: a systematic review and meta-analysis. *European Journal of Medical Research*, 28(1):342, 2023.

[7] Ollama. Ollama: Run large language models locally. https://ollama.com, 2024. Accessed: 2025-06-10.

[8] Paul Raingeard de la Bletiere. A musical robot for people with dementia. In *Proceedings of the 26th International Conference on Multimodal Interaction*, pages 602–606. ACM, 2024.

[9] Adriana Maria Rios Rincon, Antonio Miguel Cruz, Christine Daum, Noelannah Neubauer, Aidan Comeau, and Lili Liu. Digital storytelling in older adults with typical aging, and with mild cognitive impairment or dementia: A systematic literature review. *Journal of Applied Gerontology*, 41(3):867–880, 2021.

[10] Matthias S Treder, Sojin Lee, and Kamen A Tsvetanov. Introduction to large language models (llms) for dementia care and research. *Frontiers in Dementia*, 3:1385303, 2024.

[11] Andrea A Vigliotti, Victoria M Chinchilli, and Daniel R George. Evaluating the benefits of the timeslips creative storytelling program for persons with varying degrees of dementia severity. *American Journal of Alzheimer's Disease & Other Dementias*, 34(3):163–170, 2019.

[12] World Health Organization. Dementia: Key facts, 2023. Retrieved from https://www.who.int/news-room/fact-sheets/detail/dementia.

[13] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, and Junmo Kim. Text-to-image diffusion models in generative ai: A survey. *Preprint submitted to Elsevier*, 2024. arXiv:2303.07909.