Doctoral Dissertation

# Exploring the use of text-based image generation for creative writing

**Azuaje Suarez Gamar Ivan**
Program of Information Science and Engineering
Graduate School of Science and Technology
Nara Institute of Science and Technology

Supervisor: Professor Eiji Aramaki
Social Computing Lab. (Division of Information Science)

Submitted on May 29, 2024

A Doctoral Dissertation
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Azuaje Suarez Gamar Ivan

Thesis Committee:

| | |
|---|---|
| Supervisor | Eiji Aramaki |
| | (Professor, Division of Information Science) |
| | Kiyoshi Kiyokawa |
| | (Professor, Division of Information Science) |
| | Shoko Wakamiya |
| | (Associate Professor, Division of Information Science) |
| | Shuntaro Yada |
| | (Assistant Professor, Division of Information Science) |

# Exploring the use of text-based image generation for creative writing[*]

Azuaje Suarez Gamar Ivan

## Abstract

Text-to-image generation is a rapidly growing field in artificial intelligence that leverages deep learning models, particularly Generative Adversarial Networks (GANs), to transform textual descriptions into visual representations. This dissertation explores the application of text-to-image generation in two distinct creative domains aimed at enhancing creativity and emotional well-being.

Firstly, we introduce Visualyre, a tool designed to assist musicians in creating album art. This application employs GANs and Style Transfer to produce unique visuals that reflect the themes, moods, and identities embedded within the musicians' song lyrics and audio files. A user study involving 35 amateur and independent musicians highlights Visualyre's effectiveness in providing a valuable resource for artists, especially those with limited financial resources or design skills, to enhance their musical projects with personalised album artwork.

Secondly, we introduce StoryWriter, a GAN-based tool that explores the potential of real-time visual feedback to downregulate negative emotions during fictional writing exercises. Although user studies show mixed results, they reveal the potential of StoryWriter to improve emotional outcomes.

These studies illustrate the potential of using AI as a creative tool and as a way to support mental well-being. Visualyre aids musicians in creating personalised album art, while StoryWriter enhances the writing experience while improving emotional outcomes. Future work can further refine these applications, such as expanding the range

[*] Doctoral Dissertation, Graduate School of Science and Technology, Nara Institute of Science and Technology, May 29, 2024.

of supported moods, improving the quality of the generated images, and exploring the long-term effects of using visual feedback for emotional regulation.

**Keywords:**

Computer Vision (CV), Text-To-Image Generation, Neural Style Transfer (NST), Human-Computer Interaction (HCI), Thematic Analysis, Mood detection, Emotion Regulation

# Contents

# List of Figures

# List of Tables

# 1 Introduction

## 1.1 Motivation

In recent years, advancements in artificial intelligence (AI) have revolutionized creative expression, enabling machines to generate text, images, and music in response to human input [5–7]. Among these, text-to-image models have unlocked new avenues for artistic exploration and innovation, blurring the lines between human and machine creativity. This dissertation investigates the potential of such tools in two key domains: assisting musicians in creating personalized artwork for their music and facilitating emotional regulation through therapeutic writing.

The inspiration for this research lies in the growing recognition of AI's potential to transform creative practices and foster emotional well-being. While AI has traditionally been seen as a tool for automation, recent breakthroughs in generative models have opened up unprecedented possibilities for artistic expression and self-discovery. Text-to-image generation tools, in particular, bridge the gap between language and visual representation, empowering individuals to externalize their inner worlds in novel ways.

For instance, the rise of online streaming platforms has increased the demand for personalized album art, especially among independent musicians who may lack the resources or skills to create it themselves. However, current methods for creating album art rely on websites that offer limited image selections or require design skills and software that musicians may not have. AI-powered text-to-image generation models can address this gap by allowing musicians to create unique and personalized album art tailored to their music. To this end, we introduce Visualyre, a tool designed to assist musicians in creating album art. This application employs Generative Adversarial Networks (GANs) [8] and Style Transfer [9] to produce unique visuals that reflect the themes, moods, and identities embedded within the musicians' song lyrics and audio files.

Additionally, the therapeutic benefits of expressive writing for emotional regulation are well-documented, but existing methods may have limitations and potential negative effects. Research has shown that while expressive writing can be beneficial, it can also exacerbate negative emotions in unsupervised settings. AI-powered tools like text-to-image generation models can address these challenges by offering novel solutions for artistic expression and emotional support. For example, these models can generate real-time visual feedback during writing exercises, potentially enhancing the therapeutic experience and providing a form of positive distraction. To explore this potential, we introduce StoryWriter, a GAN-based tool that explores the potential of real-time visual feedback to downregulate negative emotions during fictional writing exercises.

This dissertation aims to address the challenges of aligning AI-generated artwork with the user's creative vision and emotional intent, focusing on two key areas: (1) assisting musicians in generating personalized artwork tailored to their music and (2) facilitating emotional regulation through therapeutic writing. By investigating how these AI tools can be designed and implemented, this research seeks to understand their potential to enhance human creativity and emotional well-being rather than replace or diminish them. The overall architecture for the tools developed for this dissertation, Visualyre and StoryWriter, is shown in Figure 1.

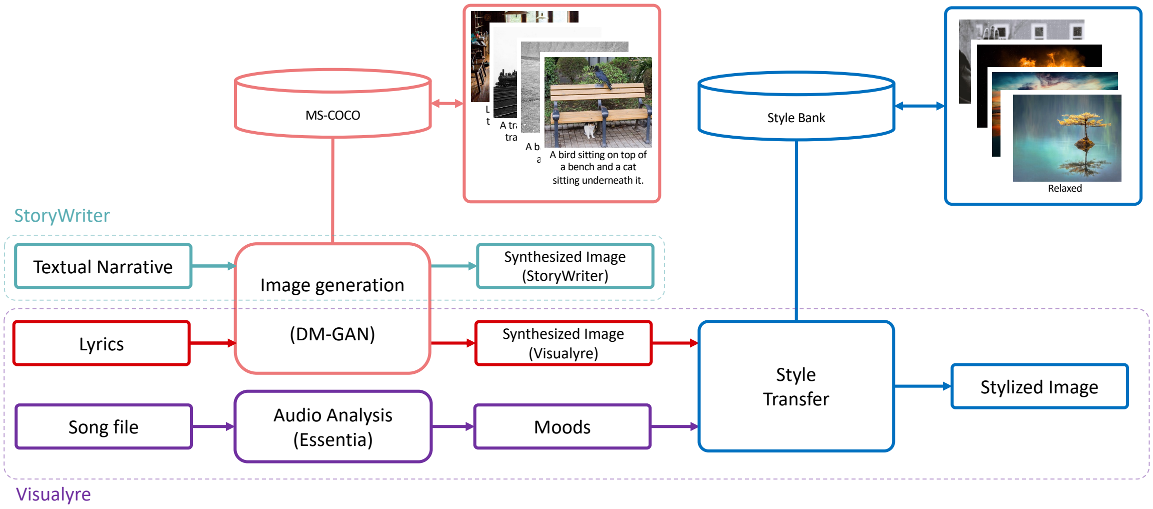

MS-COCO
A bird sitting on top of a bench and a cat sitting underneath it.
Style Bank
Relaxed
StoryWriter
Textual Narrative
Image generation
(DM-GAN)
Synthesized Image (StoryWriter)
Lyrics
Synthesized Image (Visualyre)
Style Transfer
Stylized Image
Song file
Audio Analysis (Essentia)
Moods
Visualyre

Figure 1: Overall architecture for Visualyre and StoryWriter.

## 1.2 Objectives

This dissertation aims to address the following objectives:

1. Investigate the effectiveness of text-to-image generation in generating album art that reflects the moods and themes tailored to the musician's song lyrics and audio files (Chapter 3). This is done by introducing Visualyre, which has the following objectives:

    (a) Evaluate the usability and intuitiveness of Visualyre for musicians.

    (b) Assess the effectiveness of text-to-image generation and Style Transfer in visualizing lyrics and capturing the mood of the song.

    (c) Gauge the overall effectiveness of Visualyre for musicians and the need for such a tool.

2. Explore the potential of real-time visual feedback in downregulating negative emotion through a fictional writing exercise (Chapter 4). This is done by introducing StoryWriter, which has the following objectives:

    (a) Quantify the emotional effect of StoryWriter in downregulating negative emotions.

    (b) Examine how users perceive the experience of using StoryWriter and the mechanisms involved in positive distraction.

    (c) Assess the impact of real-time visual feedback on engagement and emotional outcomes during the writing process.

## 1.3 Outline

The remainder of the chapters of this dissertation are structured as follows. Chapter 2 provides a background on text-to-image generation models, detailing their different architectures and the latest advancements. Chapter 3 presents Visualyre, a tool designed to assist musicians in generating album artwork tailored to their music based on lyrics and audio analysis. Chapter 4 explores the use of StoryWriter, an application that leverages text-to-image generation to provide real-time visual feedback, to explore its potential in supporting emotional regulation through expressive writing. Chapter 5 summarises both studies, acknowledging the limitations of this dissertation and proposing directions for future research.

# 2 Background

This chapter provides a background on text-to-image generation and the advancements of these technologies over the last year. We provide an overview of the different approaches to image generation, an explanation of the main architectures used by the models that adopt text-to-image generation, and the latest advancements that allowed these models to achieve the high quality they have nowadays.

## 2.1 Image Generation

Image generation is a technique that uses deep learning to synthesize visual content from scratch. According to how these models generate content, they can be classified into two different approaches: unconditional and conditional generation.

Unconditional generation generates images based on patterns learned from the training data without any guidance. Although the generation is random, its use allows us to understand the latent space behind these models. For instance, Deep Convolutional Generative Adversarial Network (DCGAN) [10] uses vector arithmetic to combine different attributes of the image, while StyleGAN [11] provides more precise control that allows changing specific attributes (e.g. hair colour, skin tone) without affecting the other aspects of the image.

Conditional generation uses additional information, such as labels and text prompts, to guide the image generation process. Constraining the synthesis to a textual input is known as text-to-image generation, which allows linking natural language descriptions with visual representations. Text-to-image generation models primarily rely on two different architectures, Generative Adversarial Networks (GANs) and Diffusion Models (DMs), which are explained below.

## 2.2 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) [8] are largely responsible for the emergence of text-to-image generation due to their unique architecture and training process. GANs consist of two neural networks, a generator and a discriminator, that are trained simultaneously. A generator takes random noise as input to try to generate images that resemble those of the training data. A discriminator takes an input image and attempts to classify the image as true or false. These networks are trained simultaneously in an adversarial game, where the generator tries to fool the discriminator by producing realistic images while the discriminator becomes better at distinguishing real (those in the training dataset) images from fake (those generated by the generator) ones. Both networks improve throughout training, resulting in a network where the generator creates more realistic images and the discriminator more skillfully identifies false images. Figure 2 illustrates the basic architecture of GAN and the training process.

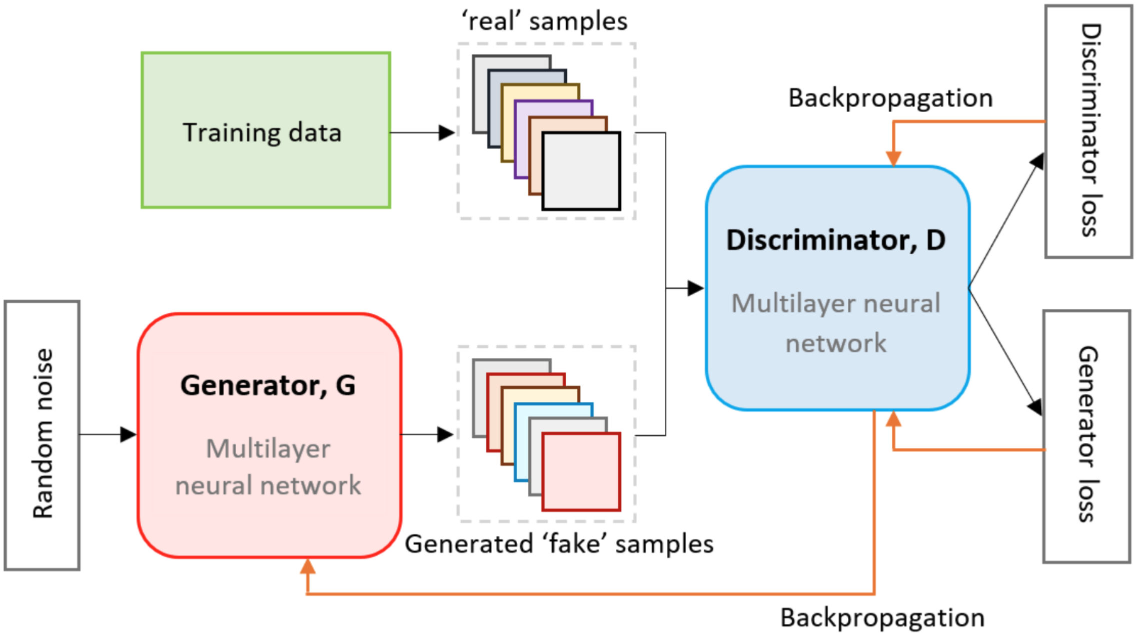


Figure 2: Architecture of GAN models, both networks are trained simultaneously in an adversarial game. Adapted from: [1].

This network architecture can be adapted for a text-to-image generation task. In this scenario, the generator is conditioned to a textual input, where the initial random noise is combined with the textual information, enabling a generation process that produces images relevant to the given text. The discriminator, on the other hand, is trained to distinguish between real images that match the given text and fake images produced by the generator. Over the years, multiple GAN-based text-to-image generation models have been developed, introducing features that improve the quality of the generated images.

- **StackGAN [12].** This model employs two GAN networks in a two-stage process. The first stage sketches the basic shape and colours, creating a low-resolution image. The second stage refines the output from the first stage, generating a high-resolution image with photo-realistic details. An improved model, StackGAN++ [13], introduces a conditioning augmentation technique that encourages the model to focus on specific details in the text. This feature allows the model to achieve a more stable training behaviour and generate higher-resolution images due to its three-stage process.

- **AttnGAN [14].** This model addresses the issue of generating images with fine details and complex relationships between objects by incorporating an attention mechanism. This mechanism allows the generator to focus on relevant words in the text description while generating different parts of the image. This attention mechanism enables the model to understand better the semantic relationships between different elements in the text and translate them into visually coherent and detailed representations.

- **DM-GAN (Dynamic Memory GAN) [2].** This model addresses the challenge of generating complex scenes with multiple objects and relationships by incorporating a dynamic memory module. This module allows the model to store and retrieve information from previous generations, improving the quality of the images generated in the subsequent stages. This approach mitigates a common weakness in multi-stage generative models, where the quality of the final image is heavily dependent on the quality of the initial generations. Figure 3 shows the architecture of DM-GAN.

- **VQGAN (Vector Quantized Generative Adversarial Network) [15].** This model introduces a novel approach to image generation by employing vector quantisation. It represents images as a composition of discrete codes, each corresponding to a learned ”visual token” from a codebook. This discrete representation not only enables efficient compression but also facilitates the manipulation and generation of images using discrete models like transformers. By combining this vector quantisation approach with the generative power of adversarial training, VQGAN achieves impressive results in producing high-quality, diverse, and detailed images from various input modalities, such as text prompts or sketches. The discrete nature of the latent space also allows for more precise control over the image generation process, making it easier to explore variations and manipulate specific aspects of the generated images. Furthermore, VQGAN can be seamlessly integrated with CLIP (Contrastive Language-Image Pre-training) [16] to enable text-to-image generation [17]. This combination allows for iterative refinement of the generated images based on the semantic understanding of the text prompt provided by CLIP, resulting in images that are both visually appealing and semantically aligned with the given text.

## 2.3 Diffusion Models

Training two networks competing against each other can result in instability, as it requires a delicate balance between the generator and the discriminator [18]. This

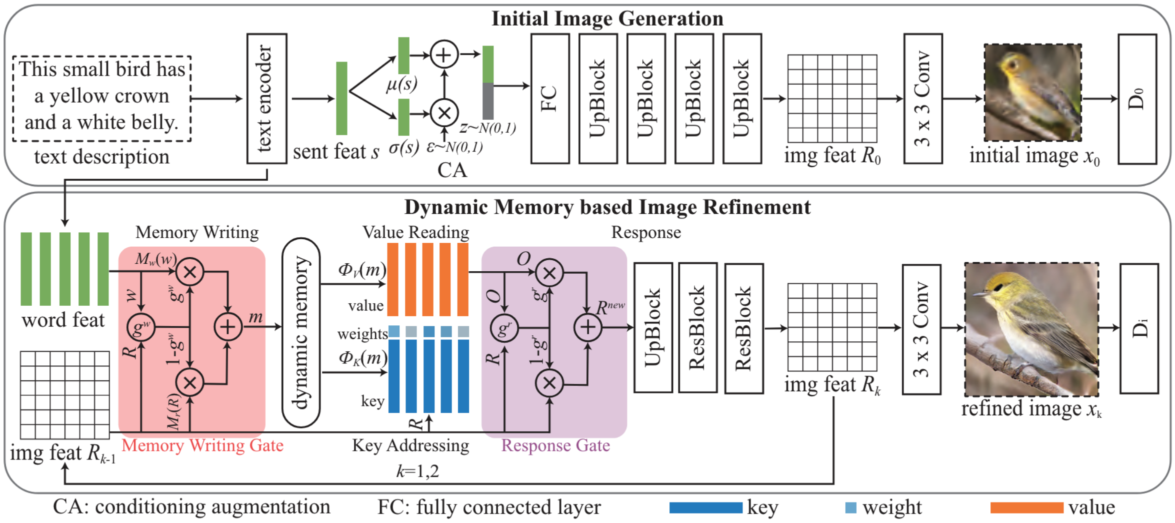


Figure 3: DM-GAN architecture for text-to-image generation. Adapted from: [2].

balance is even more challenging in hierarchical GAN models, where multiple generators and discriminators are involved in a complex interplay. The instability can manifest in several ways: non-convergence, where neither network reaches an optimal solution; mode collapse, where the generator produces a limited variety of outputs, failing to capture the full diversity of the data; vanishing gradients, where the discriminator becomes so adept at identifying fake images, hindering the training progress of the generator.

Diffusion models (DMs) [19] use a different strategy for training when compared to GANs. They progressively degrade an image by introducing random perturbations until it becomes indistinguishable from pure noise. Subsequently, the model learns to reverse this degradation process, gradually reconstructing the original image from the noisy input, leading to a more controlled and stable training process scalable to larger datasets and more complex architectures. Figure 4 showcases the diffusion process. In a text-to-image generation setting, the reverse diffusion process is guided by a text prompt, enabling the model to generate images that align with the meaning and stylistic information from the text. Some notable diffusion models are as follows.

- **GLIDE (Guided Language to Image Diffusion for Generation and Editing) [20]**, a 3.5 billion parameter diffusion model that generates and edits photorealistic images based on text descriptions. This model leverages two guidance techniques: CLIP guidance and classifier-free guidance. The latter produces superior results in terms of photorealism and caption similarity. This model has some limitations, such as struggling with unusual prompts and its relatively slow sampling speed compared to GANs.

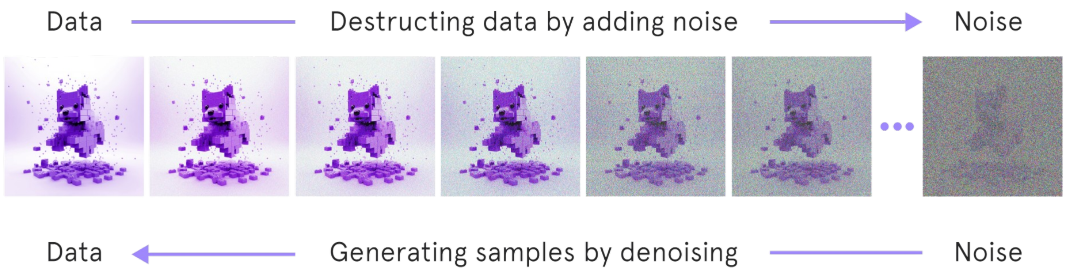

Data
Destructing data by adding noise
Noise
Data
Generating samples by denoising
Noise

Figure 4: Diffusion process, the model adds noise to degrade an image to learn how to generate new images from a noisy input. Adapted from: [3].

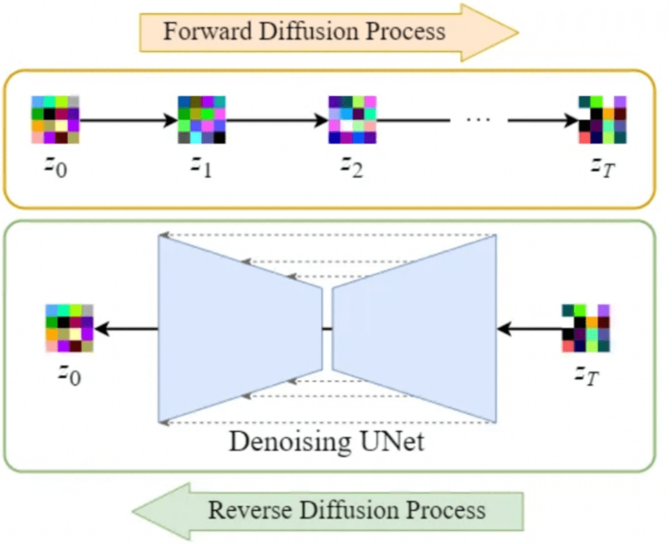


Figure 5: Latent diffusion process, the model applies the diffusion process to the latent space rather than the original image. Adapted from: [4].

- **DALL-E** [21], a 12-billion parameter text-to-image generation model based on an autoregressive transformer that generates images with high-fidelity in a zero-shot setting. DALL-E 2 [22] significantly improved image quality and fidelity to text prompts with 3.5 billion parameters by employing diffusion models and CLIP to better align images and text. The latest iteration, DALL-E 3 [23] improves over the previous model by employing an image captioner to recaption the training dataset. The use of synthetic captions that are highly descriptive allows the model to understand and generate images with more detail and complexity.

- **Stable Diffusion [24]:** Stable Diffusion is a text-to-image AI model that utilises a latent diffusion model (LDM) architecture to generate high-quality images from text prompts. It contains three main components: a Variational Autoencoder (VAE) to compress images into a lower-dimensional latent space, a U-Net to iteratively denoise and refine the latent representation based on text prompts, and a text encoder (usually CLIP) to convert text prompts into embeddings to allow the model to synthesizes images from sentences. This architecture enables the model to generate high-quality images efficiently by operating in the compressed latent space rather than directly manipulating pixel-level data, Figure 5.

# 3 Visualyre: Multimodal Album Art Generation for Independent Musicians

## 3.1 Introduction

Although music is primarily an auditory experience, users' first impressions of unfamiliar artists' work are largely influenced by associated marketing materials, including images and visual branding. For the most part, cover images, such as album art, are designed to embody the artist's aesthetics [25], and they are frequently used in online music discovery and streaming platforms, such as Spotify, Pandora, Last.fm, and Deezer, to attract new listeners and provide users with a multisensory, crossmodal experience congruous with the meanings, moods, and ambience of the songs [26]. Historically, album artworks have mirrored the musical trends and identities of their musicians. For example, in the 1960s, the Beatles used an album cover reminiscent of contemporary visual art to match the experimental direction of the music in their iconic album *Sgt. Pepper's Lonely Hearts Club Band* [25].

In recent years, the meteoric rise of online streaming platforms has made it easier for independent musicians to broadcast their music to listeners around the world. At the same time, this places the burden of marketing and other non-music tasks to the musician [27]. This may be especially for independent musicians, who might lack the financial resources, time, or design experience necessary to commission or create suitable album artwork. Given the important role that such artwork plays in communicating the themes of the album's songs, procuring suitable images is a matter of concern for independent musicians [1]. Indeed, online communities of musicians have dedicated multiple discussion threads on online forums (subreddits) to discussing potential solutions to this problem[2]. Some common suggestions on these forums include using image retrieval to source artworks (from websites like Google Images or DeviantArt) or using design tips and image editing software to produce DIY album art. However, these strategies risk copyright infringement and are limited by musicians' design skills and access to relevant software.

To this end, we propose the use of state-of-the-art image generation and styling techniques in computer vision and deep learning, and we add an interface aimed to be fast, convenient, and intuitive for users. We describe a system architecture that relies on user-submitted music and lyrics as input for two models: a Generative Adversarial Network (GAN) [8] text-to-image model, which generates seed images from each line

[1] Note that we use the term album art to refer to images that accompany songs that are typically used as promotional material, and not necessarily a cover image that one would find on a CD

[2] For example, https://www.reddit.com/r/makinghiphop/comments/84nthl/

of the lyrics, and a Style Transfer model, which uses information about the dominant moods in the audio file to pull related 'emotional' images from a previously assembled database and layer them onto the seed images. Style transfer [28] is a deep-learning-based method that combines the neural representations of two images, resulting in a third image that blends the two constituent images. Through this two-step approach, the tool generates a selection of copyright-free, custom cover artworks that capture some of the semantic meanings of the lyrics and the mood of the music. From this selection, musicians can download the cover artwork that best accompanies their music. We opted for a modern responsive layout and a minimalist design to make the process of album art generation as seamless as possible.

Finally, to assess whether the proposed architecture is suitable for its intended purpose, we designed a simple user study targeting musicians in diverse genres to measure the usability of the application and the suitability of the output images. We recruited 35 professional and amateur musicians from around the world, who tested the application with their own songs and subsequently assessed (1) whether the GAN-generated images were usable as cover artwork and (2) whether the system itself was easy to use.

## 3.2 Related Work

By combining semantic information from the lyrics and acoustic information from the audio, out proposed system can synthesize album art while also providing a new way for users to visualize music. Therefore, we consider previous work on both music visualization and album art generation.

One of the approaches of music visualization is through musical moods. Laurier et al. [29] designed an application that visualizes moods probabilities in real-time, displaying the probability of each mood during a specific time frame using bar charts. Husain et al. [30] followed the same approach, designing visual textures based on the relationship between moods and different visual elements, such as color and shapes. Another approach is the visualization from lyrics. Funasawa et al. [31] created a system to identify keywords within each line, and used image retrieval to form a slideshow that displays images corresponding to words in each line of lyrics. Visualyre adopts both approaches in a multimodal fashion. The lyrics are visualized line-by-line, but rather than using keywords to retrieve existing images, it uses image generation to synthesize new images using each sentence as an input. The moods extracted from the audio are visualized via Style Transfer, using pictures related to each mood as texture images, generating copyright-free images suitable for use as album art.

Current methods to create album art rely on websites that aid users in generating different kinds of cover art for various media from a base stock image [3]. The selection of the images is quite limited and may not be to the one desired by the user. A limitation of these applications is that other users may use the same stock image, which makes the resulting album art no longer unique.

Hepburn et al. addressed the limitation above by using a DC-GAN [10] model. This model was trained on the One Million Audio Cover Images for Research (OMACIR) [4] dataset to automatically synthesize album covers [32]. They also experimented with the AC-GAN [33] model to generate album arts for a specific genre. Visualyre adopts a similar approach to synthesize album covers. However, instead of using musical genres as input, it uses a set of lyrics and an audio file from a user-provided song. A GAN model synthesizes an image from the lyrics, which is further enhanced with the moods of an audio file. Moreover, the authors only showcased the results of their model, while our work incorporates a user interface that allows users to generate album covers in an interactive and personalized way.

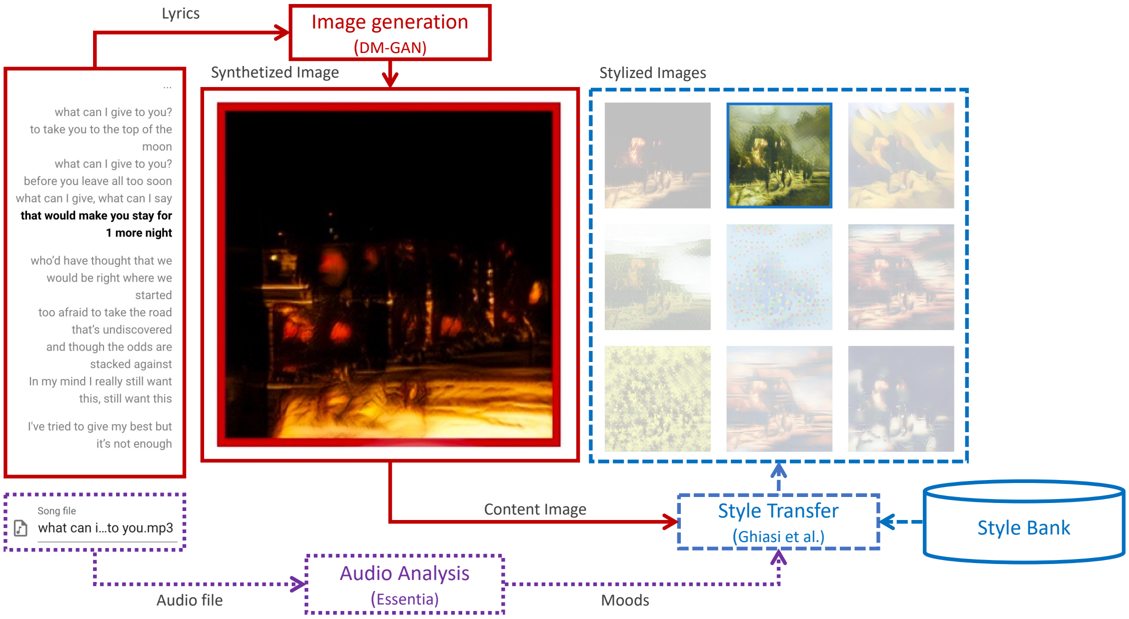

Lyrics
Image generation
(DM-GAN)
Synthetized Image
Stylized Images
...
what can I give to you?
to take you to the top of the moon
what can I give to you?
before you leave all too soon
what can I give, what can I say
that would make you stay for 1 more night
who'd have thought that we would be right where we started
too afraid to take the road that's undiscovered
and though the odds are stacked against
In my mind I really still want this, still want this
I've tried to give my best but it's not enough
Song file
what can i...to you.mp3
Content Image
Style Transfer
(Ghiasi et al.)
Style Bank
Audio Analysis
(Essentia)
Audio file
Moods

Figure 6: System Architecture: Lyrics are used to synthesize images through a text-based image generator. The audio file is analyzed to evaluate the presence of four different moods. This evaluation is used to sample which style images are used in Style Transfer.

[3] https://www.canva.com/

[4] https://archive.org/details/audio-covers

## 3.3 System Description

Given that album artwork could serve to preliminary communicate the themes and sentiments of a song (or collection of songs) to a would-be listener, the most helpful image-generation tool for independent musicians would do more than produce attractive images—it would account for both the semantic meanings of the lyrics and the emotional moods of the instrumentation. As such, Visualyre uses the lyrics (text) and music (audio) of the input song as features for the generation of representative images.

To do this, it deploys three modules that pair the information from the audio files and the lyrics: (1) a text-to-image GAN model to synthesize images from lyrics, (2) an audio analyzer to evaluate the moods of a song file based on its audio features, and (3) a Style Transfer model to adapt the image from the first module to a particular mood. In this way, users are able to fine-tune and adjust the generated images, achieving a similar effect to image filters on smartphone camera applications. The system architecture is shown in Figure 6.

### 3.3.1 Image Generation

In the first module, a text-encoder extracts features from each line in the input lyrics and uses these features to conditionate the synthesis of images. This is done by employing a text-to-image generation model. Multiple approaches have been proposed over the years to improve these models, such as using a stacked approach to refine the images [13], employing attention to synthesize fine-grained details [14], and inferring semantic information to guide the image generation [34].

Visualyre uses a DM-GAN to visualize the lyrics, generating an image for each sentence in the input lyrics. DM-GAN [2] makes use of dynamic memory to refine the content of the synthesized images, even if the quality of the previous images is not good. This generative model is trained using the MS COCO Dataset [35], which contains a total of 123,287 images (82,783 for training and 40,504 for validation) with five captions per image. The captions of the images are short descriptions in English of a particular scene; as such, the model can only generate images from English sentences. Once our system generates a selection of images, users must select one before proceeding to the next step. Figure 7 shows a selection of one of the synthesized images.

### 3.3.2 Audio Analysis

Music encompasses more than just lyrics; in many cases, and while outside the target base for our application, music may be completely instrumental and devoid of lyrics completely. Thus, to make more efficient use of the source material (music), one must also attend to the emotionality of the instrumentation [36]. To generate cover art capable of representing these musical emotions, we utilize binary classifiers to detect the relative presence of four different emotions—anger, happiness, sadness, and relaxation—based on the musical features extracted from the song's audio signal. These classifiers were introduced by Laurier et al. using Support Vector Machine models and were recently ported to pre-trained deep convolutional neural network models by the Music Technology Group [37]. Although the classifiers include three additional moods—acoustic, electronic, and party—, we did not include them during the audio analysis as they do not have a consistent visual representation. We also considered segmenting the audio analysis according to the lyrics, but the poor performance of the lyrics transcription made this approach unfeasible

For our model, we normalized the positive label (e.g., 'angry' instead of 'not angry') of these classifiers and used this information to derive an audio-based Mood Probability score, which assigns a probability to each mood based on the score of the classifiers. The higher the score, the more likely the mood will be sampled for Style Transfer (through the process described in Section 3.3.3). Likewise, if a particular score has a very small value, its sampling probability will be very low and other moods will be sampled instead. This would allow us to enhance the images generated in the first phase by giving them an artistic feel.

### 3.3.3 Style Transfer

To visually represent the moods obtained from the audio analysis described above, we use a technique called Style Transfer. Style Transfer is the process of changing the style of an image to match the style of another image while still preserving the original image's content [38]. In previous work, Style Transfer models were only capable to change the style to a specific style, as these models used a different trained model for each style image [28, 39]. However, Ghaidi et al. [9] introduced a style prediction network that can predict the style embeddings of an arbitrary style image, effectively enabling stylization using any pair of content and style images. Their network is trained using content images from the ImageNet dataset[5] and the Kaggle Painter By Numbers (PBN) dataset,[6] which consists of 79,433 paintings.

For Visualyre, we utilized Ghaidi et al. Style Transfer model by preparing a small dataset of 'emotional' images called 'Style Bank' by querying Unsplash[7] for images that reflect the four different moods assessed by our binary classifiers (Section 3.3.2): anger,

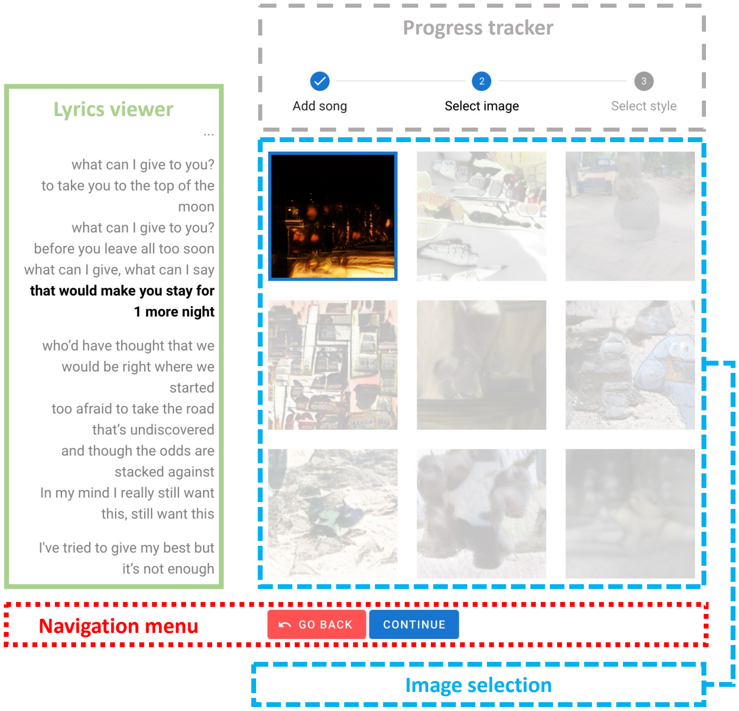
[5] https://www.image-net.org/

[6] https://www.kaggle.com/c/painter-by-numbers

[7] https://unsplash.com/

happiness, sadness, and relaxation. We used the moods and words with similar meanings (e.g., ‘fury,’ ‘joy,’ ‘calm,’ and ‘sorrow’) as search queries to obtain images relevant to a particular mood. Our intention was to use Style Transfer to apply the ‘mood’ from the images with the images generated by the GAN model (Section 3.3.1). To determine which ‘mood’ image should be combined with the GAN generated image, we relied on the Mood Probability scores derived from the audio analysis. After sampling a mood, the system selects an image representative of the chosen mood from the Style Bank, which serves as the style image. This style image is then used to generate eight different versions of the previously selected seed image. Together with the base synthesized image, we show the user a total of nine different images to choose from, which is shown in Figure 6

### 3.3.4 Graphical User Interface



...

what can I give to you?
to take you to the top of the moon
what can I give to you?
before you leave all too soon
what can I give, what can I say
**that would make you stay for 1 more night**

who'd have thought that we would be right where we started
too afraid to take the road that's undiscovered
and though the odds are stacked against
In my mind I really still want this, still want this

I've tried to give my best but it's not enough

Figure 7: Graphical User Interface of Visualyre and its components. This figure shows the selection of a synthesised image from the image generator prior to applying Style Transfer.

As per the above description, Visualyre synthesizes cover artworks in a sequential manner. To make this process as seamless and dynamic as possible for the end user, we opted for a responsive layout and a minimalist design, showing each component only when it is needed. This means that the entire process occurs on the same page, without reloading or redirecting to another URL. Instead, the UI components and their respective contents change reactively according to the user's interaction with Visualyre. Figure 7 shows the graphical user interface and its different components (which are also listed below).

- **Song input:** A form where users add lyrics and upload an audio file. The users input the lyrics by copying and pasting into a text box and upload the song by choosing a file from their device.
- **Navigation menu:** Buttons that enable travel to the next step or to the previous one. This component also indicates whether the system has finished analyzing the moods from the submitted audio file or if there occurred any errors during the analysis.
- **Progress tracker:** allows users to check their current step at any time. The name of each step provides a subtle description of the desired action during a specific step.
- **Image selection:** A grid where users can select or hover over generated or stylized images. Multiple images are displayed simultaneously, users can scroll to display further images when necessary. To advance to the style selection, a user must select one of the synthesized images. Similarly, users must select a stylized image to download an image.
- **Lyrics viewer:** An area where users can view the input lyrics while selecting synthesized images. When the user hovers over or selects an image, the system highlights the fragment of lyrics used to generate that image.

## 3.4 System Interaction

To use Visualyre, users proceed through a series of steps, which are detailed below and illustrated in the user experience flow in Figure 8. The letters indicate the current node from the flowchart.

1. An empty form (A) is displayed to the user.
2. The user inputs the lyrics in the text field and uploads an audio file (B). The user clicks the SUBMIT to start the audio analyzer.
3. The top left button in the navigation changes to READY (C), indicating that the mood detection of the audio file has finished.

4. The user clicks CONTINUE and is redirected to the image selection. In the background, the generator starts generating images from the lyrics.

5. The images are generated and shown in a grid-like fashion.

6. The user selects a synthesized image (D).

7. The user clicks CONTINUE and is redirected to the style selection.

8. Style Transfer is applied to the selected image by sampling style images based on the results of Step 5.

9. The stylized images are created and shown in a grid-like fashion.

10. The user selects a stylized image (E). The image is shown on the left with a DOWNLOAD button.

11. The user clicks DOWNLOAD, acquiring their desired stylized image.

## 3.5 Evaluation Methodology

### 3.5.1 Prescreening

We used the Prolific[8] platform to recruit an expert group of musicians to evaluate the usability of Visualyre and its suitability in generating cover art for their songs. Prolific allows us to restrict participants according to certain categories, one of them being musical instruments. Thus, prior to recruitment, we utilized this item ("Do you play a musical instrument; if so for how many years?") to pre-select an eligible sample population from Prolific who had the highest likelihood of having experience writing songs.

We then conducted a prescreening survey, where we asked participants to share their previous musical experience. Additionally, because Visualyre requires the use of an original song with English lyrics, we asked participants whether they had previously written a song with English lyrics and whether they were willing to share such a song to evaluate our application. Based on this criteria, initially recruited N = 621 participants for the prescreening survey, of which N = 95 suitable candidates were invited to participate in the user study. All prescreening survey participants were compensated with $0.2. We paid 0.2 dollars to all participants, regardless of their answers. The prescreening includes an attention check [40], which ensures that participants were reading the questions prior to answering them.

[8] https://prolific.co/

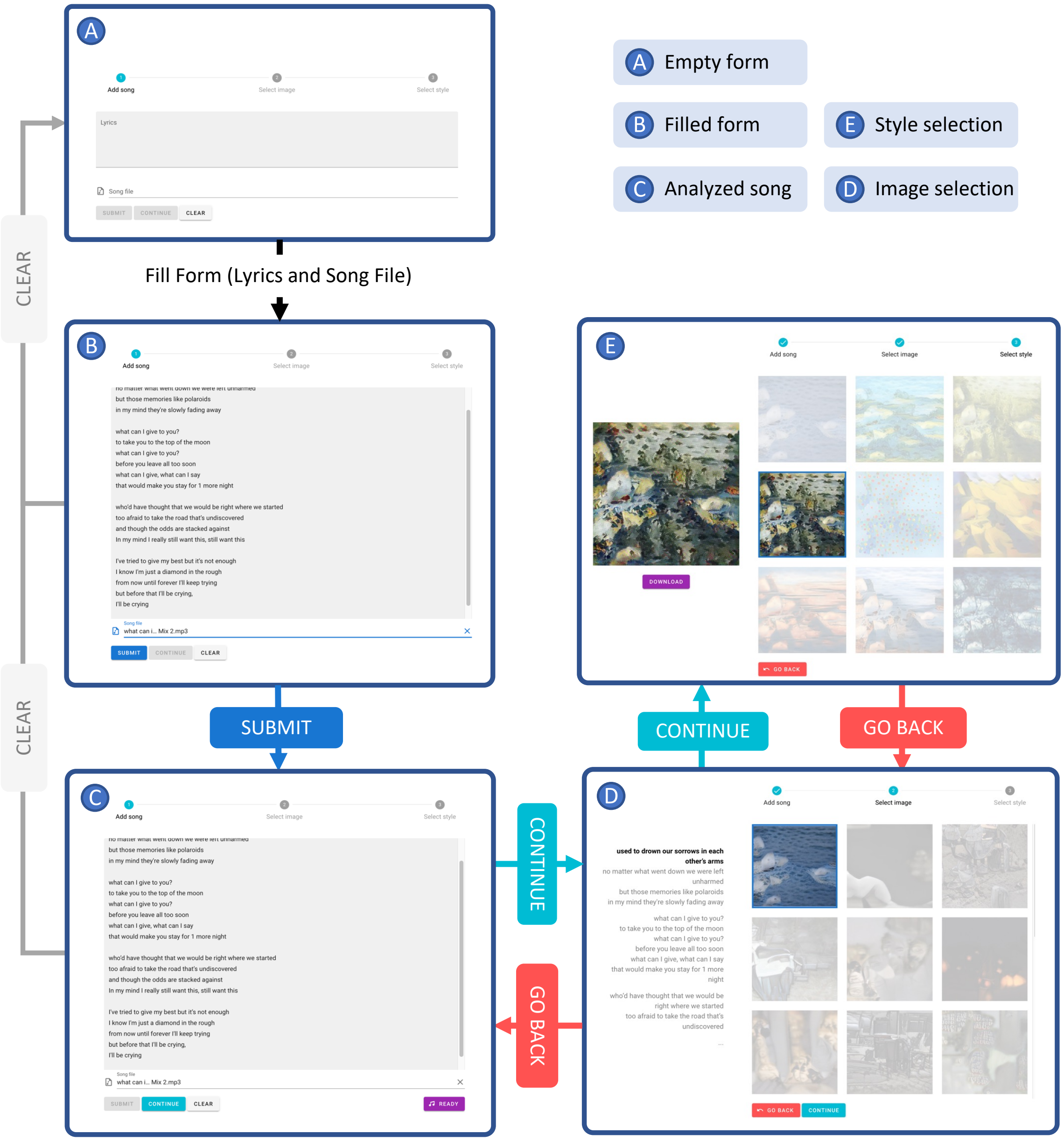


Figure 8: User experience flow. This figure shows the route that the users must follow to synthesize a stylized image from a song file and its respective lyrics. The letter indicates the current status of the application.

### 3.5.2 Telemetry

We used telemetry in our application to (1) confirm that participants interacted with the application prior to answering the Evaluation Survey (described in Section 3.5.3, below) and (2) determine how long they spent performing each step. To do this, we added a ”Create ID” button at the start of the application that generates a unique 7-digit ID, which is used to log the participants’ actions with a corresponding timestamp.

Table 1 shows all the action types and the moment when they are logged. By checking whether a specific ID has completed the *download* action, we can assess whether a participant has used the application from start to finish, as instructed.

Table 1: Types of actions logged through telemetry

| **Action type** | **Logged when** |
|---|---|
| *clear* | The user resets the form to an empty state. |
| *continue* | The user continues to the next step. |
| *download* | The user downloads an image in the final step. |
| *file_error* | An error occurs when analyzing the audio file. |
| *generate_images* | The system finishes generating images from the lyrics. |
| *generate_styles* | The system finishes stylizing the selected image. |
| *go_back* | The user returns to the previous step. |
| *select_image* | The user selects a synthesized image. |
| *select_style* | The user selects a stylized image. |
| *start* | The user starts the application, generating a unique 7-digit ID. |
| *submit_form* | The user submits the form after inputting the lyrics and the audio file. |
| *upload_lyrics* | The system receives the input lyrics and starts generating images. |
| *upload_moods* | The system finishes extracting moods from the audio file. |

### 3.5.3 Evaluation Metrics

We provided each eligible participant with an Evaluation Survey, which listed instructions for how to use the application, specified that the user must upload an original English song and download one of the stylized images, and indicated that participants were free to drop out from the evaluation at any time. There were no time constraints when using the application. Out of the 95 candidates, 35 participants (27 male, 7 female, 1 non-binary) answered all the survey questions and used the application from start to finish with an original English song. This was corroborated by the telemetry detailed above. Participants were compensated $5 for their involvement with the study.

The age of the participants ranged from 19 to 61 (Mean Age = 28.43, SD = 9.73);

Most participants reported using English (17) as a main language, followed by Spanish (5) and Polish (4). The rest of the languages were Hebrew (2), Portuguese (2), Czech (1), Dutch (1), French (1), Greek (1), and Swedish (1). Participants also labelled the genres of the songs they submitted, the most common genres were Rock (6) and Pop (6). The rest of the genres were R&B (4), Electronic (3), Metal (3), Acoustic (2), Alternative (2), House (2), Indie (2), Punk (2), Rap (2), and Piano-ballad (1). Some participants wrote more than one genre for their song. Usage time ranged from 2:41 to 44:14 (Mean = 11:59, SD = 9:32). Out of the 35 participants, only 5 of them spent more than 20 minutes using the application, while 8 of them used it for less than 5 minutes.

After using the application, participants were invited to complete the survey portion of the Evaluation Survey, where they rated the application across different metrics. Users had to rate eight different questions with a 6-point Likert scale ranging from strongly agree (6) to strongly disagree (1). These questions were designed to enable a quantitative evaluation of the app's usability and utility for generating cover art. To link these responses with the telemetry data, we asked participants to provide both their Prolific ID and the 7-digit ID generated by our application. Table 2 shows the questions from the the survey portion of the Evaluation Survey with its associated metric.

For the qualitative evaluation, users had to share their impressions of the application and their previous experiences seeking cover art for their music by answering two open-ended questions. These questions were intended to solicit further information about the app's effectiveness, usability, and necessity. The first, "What are your impressions of Visualyre?" (Q14) yielded responses from all 35 participants. The second was a follow-up question to "Prior to using this application, have you had any difficulties obtaining a cover image for your music?" (Q25); those respondents who answered "Yes" to this initial question were asked to elaborate on their previous experiences (Q26). An additional question, "Have you used a similar tool before?" (Q15) was added to assess the novelty of the application.

## 3.6 Findings

In this section, we showcase the results of our evaluation by grouping and analyzing the results of the quantitative and qualitative evaluations across different metrics. The results of the quantitative evaluation are summarized in the Table 3 below.

*How intuitive is Visualyre?*

As Visualyre requires multiple steps to accomplish music visualization (as per Section 3.4), we wanted to check whether users were able to complete this step-by-step procedure smoothly. Accordingly, we asked participants to rate whether "The application was easy

Table 2: Survey questions with their respective metrics

| Metric | Question |
|---|---|
| Assertion | Enter your prolific ID |
| Assertion | Enter your user 7-digit User ID. This is displayed at the bottom of the page. |
| Assertion | What's the genre of the song you submitted on the application? |
| Effectiveness | The downloaded image is suitable as cover art |
| Effectiveness | How likely would you recommend this application to other artists? |
| Necessity | Prior to using this application, have you had any difficulties on obtaining a cover image for your music? |
| Suitability | The application displayed at least an image that matches the theme of the lyrics |
| Suitability | The downloaded image matches the intent or intention of the corresponding lyrics |
| Suitability | The application displayed a style that matches the tone of the song |
| Suitability | The downloaded image matches the intent or intention of the song |
| Usability | The application was easy to use on every step |
| Usability | The application can quickly generate a final image from start to finish |
| Qualitative | What are your impressions of Visualyre? |

to use from beginning to end" (Q4) using the 6-point Likert scale described above. The mean score for this question was 5.09 (SD =0.92), suggesting that most users agreed that the application was intuitive and easy to use throughout the entire procedure.

Of the 28 participants who commented on the app's usability on Q14, 22 expressed positive sentiments. Eleven indicated that the app was easy to use, and one elaborated, "It's pretty easy to use. smooth transitions between pages and it didn't take long to upload and analyze the song" (p19). In terms of user experience, 14 classified the app as "cool," "interesting," or "creative," and four lauded the app's concept/features as "nice," "good," or even "great." Two noted that it was "fun" or otherwise enjoyable to use, with one writing, "I really enjoyed that a lyric from the song could be interpreted in different ways by choosing a different style for the artwork" (p17). Finally, three remarked that they liked the app or were happy to have used it, independent of their intention to use it again.

Table 3: Results of the quantitative analysis

| Question | Mean | SD |
|---|---|---|
| (Q4) The application was easy to use on every step | 5.10 | 0.96 |
| (Q5) The application displayed at least an image that matches the theme of the lyrics | 4.17 | 1.42 |
| (Q6) The application displayed at least style that matches the tonality of the song | 4.67 | 1.09 |
| (Q7) The downloaded image is suitable as cover art | 4.63 | 1.03 |
| (Q8) The downloaded image matches the intent or intention of the song | 4.17 | 1.21 |
| (Q9) The downloaded image matches the intent or intention of the corresponding lyrics | 4.07 | 1.23 |
| (Q12) How likely would you recommend this application to other artists? | 4.20 | 1.30 |

Although Visualyre was praised for its intuitiveness, some users expressed concerns. Three indicated that the app was incomplete in its current form, either for functional (n = 1) or aesthetic (n = 2) reasons. For example, one such respondent noted, “It doesn’t work at the moment, but I can see it might end up being something very interesting” (p6). Another, who had experienced technical difficulties, expressed the need for functional improvements, while three others recommended specific improvements or features, including the ability to re-access previous information, access a broader range of images or upload original work as a starter image (p3, p25, p27, respectively).

*How effective are the methods of text-to-image generation and Style Transfer for visualizing lyrics?*

Visualyre combines the methods of text-to-image generation and Style Transfer to synthesize the resulting images. To evaluate the first method, we asked participants to rate whether “The application displayed at least one image that matches the theme of the lyrics” (Q5). The mean score for this question was 4.26 (SD = 1.46). To evaluate the second method, we asked participants to rate whether “The application displayed at least one style that matches the tonality of the song” (Q6). The mean score for this question was 4.66 (SD = 1.08). Participants considered Style Transfer to be a more effective visualization method than text-to-image generation which may be due to the limitation of the MS COCO dataset used in training the current text-to-image model and the difficulty involved in representing abstract concepts through this dataset. Nevertheless, both mean scores were well within the positive (agree) end of the spectrum, which suggests that Visualyre was at least somewhat effective as visualizing lyrics.

*Are the images generated by Visualyre suitable as cover art?*

All participants used the application until they downloaded at least one of the stylized images. To determine whether these images could reasonably be used as cover art, we asked users to rate whether "The downloaded image is suitable as cover art" (Q7). The mean score for this question was 4.60 (SD = 1.03). We also added two follow-up questions to determine why users considered the downloaded image suitable as cover art: "The downloaded image matches the intent or intention of the song" (Q8) and "The downloaded image matches the intent or intention of the corresponding lyrics" (Q9). The first question yielded an mean score of 4.20 (SD = 1.16); the second question yielded an mean score of 4.14 (SD = 1.19). Combined with the preference for Style Transfer over text-to-image generation indicated in response to Q5 and Q6 (see Section 3.6), these results show that users are somewhat willing to use images generated by Visualyre as cover art despite the limitations of text-to-image generation.

In response to Q14, respondents elaborated on their positive and negative perceptions of the cover art. Six of the 16 participants who commented offered positive appraisals. Four said that the images were "interesting" or "unique" while two said they were "impressive." Two also described the images as "beautiful" or "nice," one based on the text-to-image generation ("makes beautiful covers for every line" (p13)) and the other based on the Style Transfer ("when combined with the different styles, they can give really nice examples" (p17)). Five respondents described them as abstract, conceptual, or random/non-representative, and in four of these cases the respondent regarded this as a negative attribute. For instance, one remarked, "Bit confused as to what some of the images were" (p1), while another clarified, "[the image] didn't add to the meaning of the song – the images were quite conceptual images rather than illustrative images" (p22). Five respondents also used more affective negative-leaning language, describing the images as "strange," "creepy," "disturbing," "trippy," or "nonsensical." One specified, "Some of the ones regarding people were scary" (p33). Only three respondents found the images to be unattractive, sub-par, or uninteresting, one of whom wrote, "The images provided were not that interesting . . . The images feel esasily identifiable as being AI generated" (p31). Finally, one participant commented not on the content of the image, but on the low file quality upon download. Figure 9 showcases some of the stylized images that our participants downloaded with their respective base images and target mood. To preserve anonymity, we only show two words from the sentence that was used to synthesize the base image.

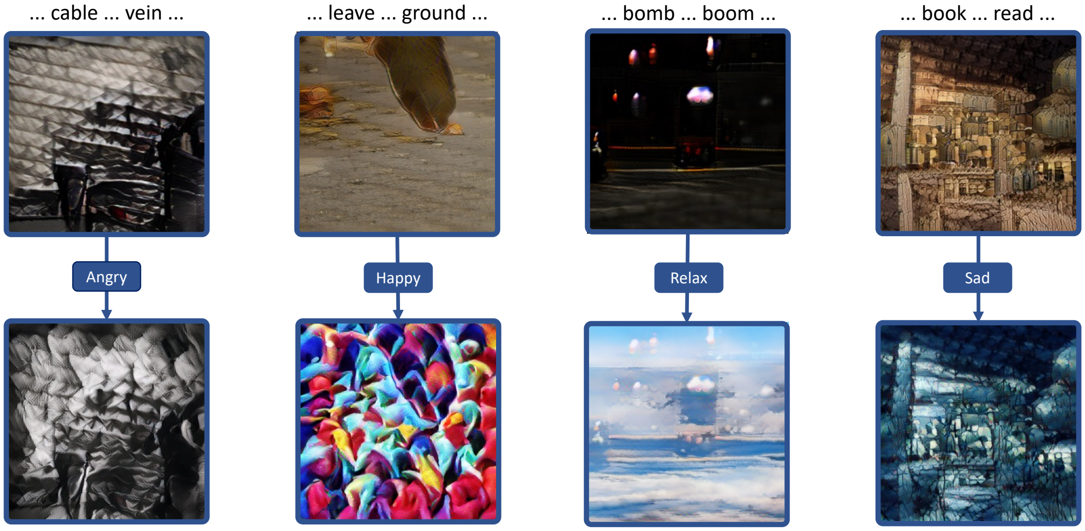


Figure 9: Synthesized images downloaded by our users. The base images are located on the top, and the stylized images are located on the bottom.

*How effective is Visualyre for musicians?*

Of the 18 participants who commented on the effectiveness of the app, 14 attested to its actual or potential efficacy. Six confirmed that they found the app generally useful, and three indicated that they would use the app for their own projects. Speaking from his own experience, one respondent summarized, “As an artist I personally know how important it to get the right art work for a song or project. Most artists struggle when it comes to creating there desired art work so Visualyre can change that by show casing defferent art for artists to just pick” (p24). Six participants regarded the app as promising, and three suggested that it might work better for artists in specific musical genres, for example, “for indie and experimental music [rather] than for traditional pop” (p14). However, three respondents expressed confusion about the purpose of the app, and one stated that they did not find the app useful in its current form, writing “I don’t know what to think about it. [. . . ] but for now I don’t believe that it can be really useful” (p28).

These findings complement the results of “How likely are you to recommend this application to other artists?” (Q12) from the quantitative part of the survey, wherein 71.4% of respondents indicated that they were somewhat likely to very likely to recommend this application to others. The average score for this question was 4.29 with an SD of 1.30. The score ranged from very likely (6) to very unlikely (1).

*Is Visualyre novel?* Though none of the respondents commented directly about the novelty of the app, we may deduce from the prevalence of "cool/interesting" comments (n = 14; i.e., "converting audio files and interpreting lyrics in a pictorial way is something really cool!" (p26)), the renderings of the app as a "concept" or "idea" (n = 3; i.e., "I think it is a great concept" (p29)), the tendency to describe the app as "promising," or as having "potential" (n = 6), or even the comments of confusion (n = 3), that users had not encountered an app like this before (there were 20 unique responses of this kind). This is further supported by the responses to another question, which asked, "Have you used a similar tool before?" All but one of the respondents answered "No," not including another who compared the app to Google's DeepMind. Thus, it seems safe to say that the app was novel for a majority of the respondents.

*Is there a need for an application like Visualyre?* Of the eight participants who responded to the second qualitative question, seven confirmed that their previous experiences obtaining/creating cover art for their music were difficult; one misinterpreted the prompt, commenting instead on their technical difficulty when utilizing Visualyre. The main reasons given for previous difficulties included sourcing issues (n = 4), including inability to find an appropriate artist or to obtain image rights, and time (n = 3). One participant explained, "Well, it takes me a lot of time to find the right cover art because I'm such a perfectionist. Sometimes I do find the right image but when I contact the owner, they won't let me use it, which is understandable" (p15). Contrary to our expectations, only two respondents cited financial obstacles, and only one cited skill-related difficulties. Three resorted to making their own artwork, though one conceded that this was "fun" despite the time commitment (p16), and another acknowledged that they have the requisite skills: "I end up doing my own artwork since I'm an illustrator" (p15). Moreover, one respondent noted that Visualyre would have been especially useful in the early phases of his career (p24), suggesting a more circumscribed area of "necessity." Indeed, it is worth noting that 27 respondents (77.1%) did not report previous issues with obtaining cover artwork. As such, it may be more appropriate to refer to the app's convenience, as it saves artists time more so than money. This may explain why, when asked "Would you pay to use a similar service?" (Q17), 42.9% of respondents answered "Yes," despite concerns with the quality of the generated images. Taken together with the predominantly positive feedback about the app's effectiveness for musicians (see Section 3.6), these comments seem to confirm that there is an audience for this kind of tool.

## 3.7 Discussion

Thus far, our data indicate that Visualyre appears to be an effective tool for musicians to generate cover images and album art for their proprietary songs. Responses to "What are your impressions of Visualyre?" (Q14) revealed that users found Visualyre to be highly usable, largely novel, and reasonably effective, though it currently does not yield wholly suitable results. Responses to "Prior to using this application, have you had any difficulties in obtaining a cover image for your music?" (Q25) confirmed the need for this tool amongst a small segment of the target population (22.9%), though for different reasons than the ones we anticipated. Most of the participants were willing to recommend this app to another musician, and 42.9% would pay for cover art generation as a service.

Overall, more than 60% of the qualitative responses leaned positive, while less than 30% leaned negative, indicating general support for the app. Of the four participants who explicitly reported on the app's performance against their expectations, two were pleased to find that it was better than expected. One stated that the app performed to plan "for someone who want a quick, abstract and unique album cover" (p8). Finally, one bemoaned that the app was different than expected because the images were nonrepresentational: "I was expecting a more finished product with actual images instead of random lines and colors" (p2). We expect that these issues could be solved by using more robust text-to-image generation models, given that participants acknowledged less concordance between lyrics and images than between tone and images in the quantitative survey.

Despite only a minority of users with prior difficulties with procuring album art, the largely positive response towards Visualyre suggests that this may be a useful tool for independent musicians to consider for online platforms and promotional material, as both the images generated and interface received satisfactory feedback. For the musician, this presents a low-cost alternative to designing album art, so the effort, time, and resources that are saved can then be diverted back to core music-making activities.

## 3.8 Conclusion

In this paper, we proposed a novel application of text-to-image computational techniques in the form of Visualyre, a lyric visualization tool that combines two techniques (text-based image generation via DM-GAN and arbitrary Style Transfer) to sequentially synthesize images based on lyric segments and the overall mood of uploaded songs. Combined with the in-built possibility for user selection and input, we think that this method allows for the generation of images that can capture some aspect of a song's semantic meaning or mood.

Moving forward, we consider some areas where Visualyre can be improved. Firstly, our GAN model is currently generating images for every line in the lyrics. Some of these lines may be too short or contain only abstract concepts, which are very difficult for the image generation model to synthesize. As such, one improvement could be to expand the training dataset used to include images trained on artistic or abstract concepts. Secondly, the range of supported moods in the audio analysis was limited and the reference images in the Style Bank were determined subjectively. Future iterations could include a more comprehensive mood detection model for audio analysis and crowdsourced annotations for a more consistent estimation of reference images' emotionality. Thirdly, Visualyre only supports lyrics written in English. Thus, another improvement would involve adding multilingual support for the application, perhaps by training GAN models with multilingual datasets or by adding machine translation capabilities within Visualyre's architecture. Finally, we note that the downloadable images have low resolution. We plan to use Super Resolution models [41] to enable high-resolution download in future iterations. Additionally, we also plan to use cloud computing for increased scalability, enabling simultaneous access to musicians around the world. This would also provide an opportunity for real-world feedback from actual users of the application, rather than paid participants used in this study. With the release of Diffusion Models [19] and CLIP [16], we think that image generation using these models may offer an alternative image-generation algorithm to the GAN used in Visualyre, though we note that the speed of CLIP-guided text-to-image synthesis is considerably slower, so retaining our current GAN model may yet be advantageous [42].

With Visualyre, one of our goals is to bring a modest contribution to the independent music community by generating artwork to match artists' musical and artistic intentions. In the future, we will explore other possible applications, such as allowing users to browse music through synthesized images.

# 4 Exploring the Use of AI Text-to-Image Generation to Downregulate Negative Emotions in an Expressive Writing Application

## 4.1 Background

Many studies have described the therapeutic benefits of expressive writing for improving resilience and managing negative emotions [43–45]. Pennebaker and Beall [46], for instance, found that writing about a traumatic event for 15 minutes a day over four days led to improved physical health six months later. Such expressive writing enables individuals to cognitively process their traumatic experiences [47], elaborate on associated negative emotions [48, 49], and thereby habituate to these negative experiences [50]. Furthermore, writing therapies are easy to deploy in unsupervised settings, and may be a good option for the estimated 42% of individuals who primarily turn to the internet for guidance on mental health issues [51]. However, research has also warned about the limited and sometimes deleterious effects of expressive writing. In a meta-analysis of 39 randomised controlled trials, Reinhold et al. [52] found that expressive writing exerted only limited effects in reducing depressive symptoms, and sometimes exacerbated results in unpredictable ways. In unsupervised settings, expressive writing can occasionally increase psychological distress [53, 54], even to the point of causing patients to discontinue their treatment [55,56]. These conflicting findings indicate that, despite the efficacy and convenience of expressive writing for emotion regulation, care should be taken when deploying this approach in self-directed practices.

Writing can also be a form of distraction for the individual. When used appropriately, distraction tasks can alleviate negative emotions during stressful situations. In two separate experimental studies, Trask and Sigmon [57] and Nolen-Hoeksema and Morrow [58] found that participants with negative emotions (induced negative emotion, in the first case; depression in the second) who completed a distraction task reported lower levels of that negative emotion post-task than those who completed a rumination task. Distraction tasks have even been effective for short-term coping [59], and they pose little to no risk of worsening baseline emotions in depressed individuals [60]. These attributes make distraction tasks ideal for online, unsupervised settings. In this research, we specifically aim to induce positive distraction, which shifts attention away from negative stimuli [61]. To do this, we pair creative writing with representational or gentle abstract art, which has been shown to facilitate positive distraction for stress and anxiety regulation [62]. Specifically, we explore the use of text-to-image generation models to generate such images in real-time based on users' narratives.

We designed and piloted an online writing tool called StoryWriter, which pairs expressive writing and deep-learning-based text-to-image functionality, to enable ongoing positive distraction. In the application, users are tasked to write creative introductions to a fictional narrative, which in turn creates machine-generated artworks that can direct them away from negative emotions. In this paper, we describe the design, implementation, and outcomes of two studies we performed to gauge the preliminary efficacy of our application. The first – an experimental study conducted with 388 users to quantify the emotional effect of the application – is recounted in Section 4.4. The second – a qualitative study with 54 remote users was performed to provide depth and context for our other findings and is recounted in Section 4.5. After reporting the results of these studies, we discuss the benefits and limitations of our application for negative emotion downregulation.

## 4.2 Related Work

In this section, we cover previous research related to (1) the use, benefits, and disadvantages of technology-mediated writing activities, and (2) the previous application of digital technology for emotion regulation and positive distraction, and the potential use of artificial intelligence for similar purposes.

### 4.2.1 Technology-Mediated Writing Activities

In contemporary society, writing activities are increasingly mediated through myriad technologies and carried out online. While classic expressive writing exercises used in therapeutic care are generally paper-based and mostly conducted on-site, burgeoning studies indicate that computer-based writing exercises may be equally effective while providing a more cost-effective, accessible, and anonymous alternative [63–65]. Online writing activities have yielded positive user feedback and promising results in fields such as Prolonged Grief Disorder (PGD) [49], Post-Traumatic Stress Disorders (PTSD) [66,67], and depression [68] and other mood disorders [69,70]. Thus, internet-based technologies have been applied in the field of healthcare to promote better self-monitoring [71, 72]. However, researchers have also warned about the potential negative effects of technology-mediated writing tools, highlighting how unsupervised writing exercises that require users to write about distressing past events could stimulate negative emotions and reduce the exercise's overall effectiveness [53, 56].

While technology-mediated writing activities seem to provide some positive effects, they are often implemented under psycho-therapeutic protocols and in a controlled environment. We argue that, in real-life situations, individuals frequently utilise online resources to directly or indirectly access mental health advice. Many people cannot easily access moderated writing exercises due to financial, systemic, or logistical constraints. As such, there is a need for technologies that can facilitate safe, online, emotionally regulating writing experiences.

### 4.2.2 Digital and Artificial Intelligence Technologies for Emotion Regulation

Whereas earlier emotion regulation practices tended to rely on participants manually self-monitoring, recognising, and adjusting their emotions in response to challenging circumstances, recent advances in digital technologies have led researchers to examine how these tools might facilitate this process [73, 74]. Ubiquitous technology, in particular, has shown considerable promise on this front. For instance, one study demonstrated that commercially available smart watches can be used to (1) automatically detect emotional outburst patterns for individuals with Autism Spectrum Disorders and (2) provide them with self-regulation strategies that can reduce unregulated anger episodes [75]. Wearable devices have also been used to help calm users who are anticipating socially stressful situations (such as a public speech) [76] or undergoing exams [77].

On a different but related front, researchers have examined how tools such as a digitally mediated nature soundscape [78] and an interactive installment featuring projected Augmented Reality and smart floor displays [79] can be used to positively distract users from stressful or anxiety-inducing situations. Immersive technologies such as virtual reality and ambient forms of digital art have been particularly effective for this purpose [62, 80]. Yet, in most of these cases, the digital content within these systems needed to be manually crafted beforehand, making it difficult to tailor the content to the specific needs and preferences of each individual.

This is where Artificial Intelligence (AI) can play a decisive role. Previous AI research on emotion regulation focused on developing algorithms to accurately detect human emotions through physiological cues like facial expressions [81, 82]. This has allowed computerised systems to perceive users' emotional states and automatically deploy various strategies to help regulate them. [83, 84]. Recent developments in high-performing deep learning models, particularly in the fields of image recognition and natural language processing, have made it possible to deploy AI more broadly to support emotion regulation beyond simply detecting emotions. Examples include chat bots being developed for text-based communication platforms to enable emotional management amongst distributed teams [85] and deep-learning-based models being trained and used to help service employees regulate customer emotions through real-time emotional feedback [86].

Further advances in artificial intelligence have led to the development of Generative Adversarial Networks (GANs), which have made it possible for AI to create novel content such as stories, images, and music [5–7]. Using such models, practitioners can generate personalised content that matches the specific characteristics of each individual, which could then be used almost immediately in emotion regulation procedures. Despite this potential, few GAN models have been applied to the issue of emotion regulation. One notable exception is an AI mirror system developed by Rajcic and McCormack [87], which uses OpenAI's GPT-2 model to generate poetry based on users' facial expressions and thereby provoke emotional self-reflection. To begin filling this gap, we explore the use of GAN models as a means for positive distraction. We deploy our prototype within an online writing-related exercise to ascertain whether the GAN mechanism can mitigate the issues acknowledged in Section (4.2.1).

## 4.3 Design and Development of StoryWriter

Our design process was guided by the following objectives: (1) users should experience improved emotions after using StoryWriter for short-term coping, (2) the mechanism must pose little to no risk of exacerbating pre-existing negative emotions, making it accessible to as many people as possible, and (3) the application's infrastructure and image generation technology should be easy to replicate, modify and scale upwards, making it straightforward for other developers to adapt and utilize our work.

To make the interface as easy to use as possible, we designed StoryWriter as a Single Page Application (SPA). We developed it using the *Vue* framework (vue.js)[9] for the front-end, and Python's Flask framework[10] for the back-end in order to forward the user's text as an input to a text-to-image generation model. Figure 10 shows the architecture of the application. Its components are described below.

### 4.3.1 Text-to-image Generation Model

Different GAN architectures offer different features that could lend themselves to text-to-image generation. For instance, StackGAN delegates image generation to multiple GAN layers, thereby increasing the resolution of each subsequent layer [12], while AttnGAN introduces an attention mechanism to match word features with subregions in the output image [6], thereby capturing fine-grained details in the generated images. We wanted a model that could solve one of the key weaknesses of a multilayered approach: that is, the way the quality of the first layer constrains the quality of the last

[9] `https://vuejs.org/`

[10] `https://flask.palletsprojects.com/en/2.0.x/`

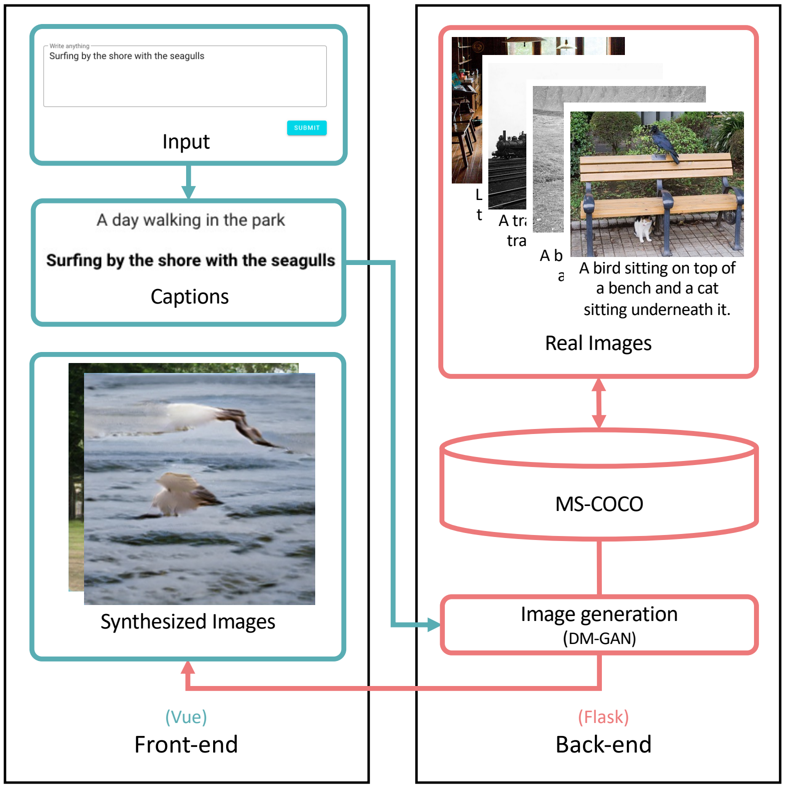

Write anything
Surfing by the shore with the seagulls
SUBMIT
Input
A day walking in the park
Surfing by the shore with the seagulls
Captions
Synthesized Images
A bird sitting on top of a bench and a cat sitting underneath it.
Real Images
MS-COCO
Image generation
(DM-GAN)
(Vue)
Front-end
(Flask)
Back-end

Figure 10: System Architecture of *StoryWriter*

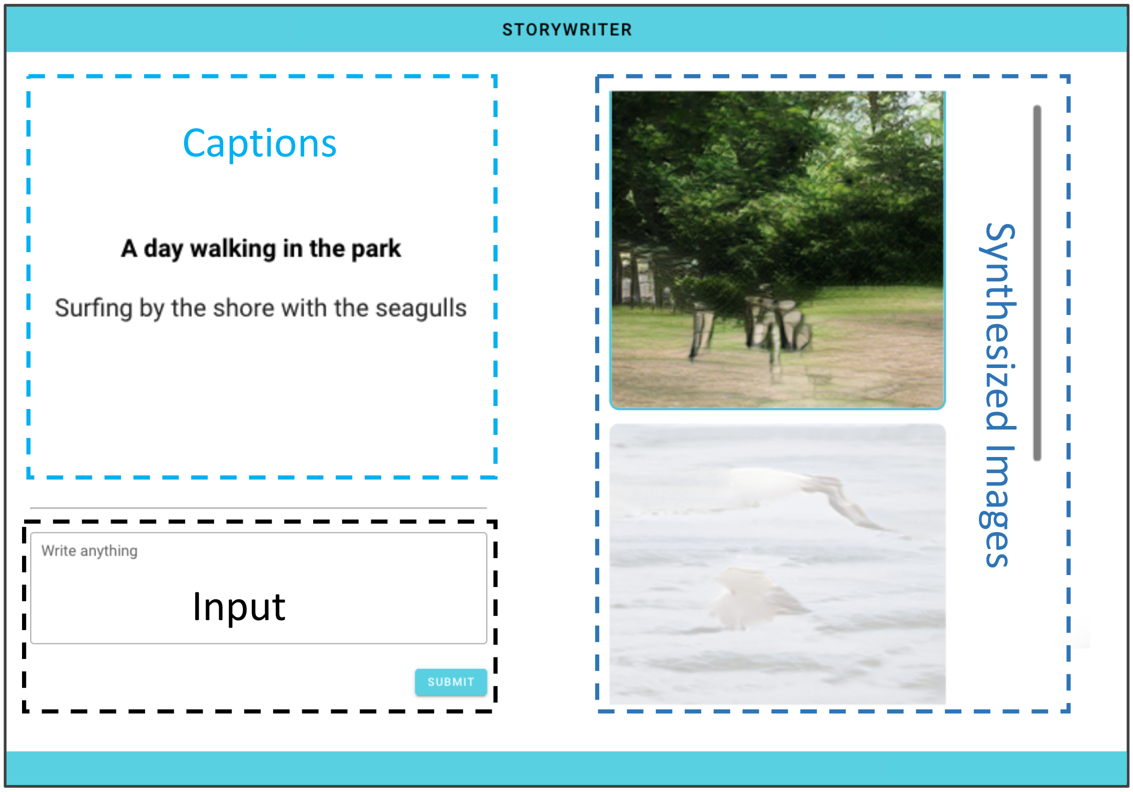
layer [2]. Thus, we decided to use Dynamic Memory (DM)-GAN in our application, as it incorporates a memory mechanism that can refine the image contents generated in the early layers of the models, yielding higher quality images [2]. This model was trained on the MS-COCO dataset [35], a dataset comprising 123,287 images with five annotated descriptions per image. This large dataset contains a total of 886k object instances across 80 categories.

### 4.3.2 User Interface

The StoryWriter interface, shown in Figure 11, is divided into three main sections: Input, Synthesised Images, and Captions. We describe each of these sections below.

#### Input

The input area refers to the text box where users can write and submit their narratives. Users can write anything within this text area, but they are directed to submit content one sentence at a time to facilitate image generation. Once the user clicks the submit button, the DM-GAN model will synthesise an image using the submitted text as an



**A day walking in the park**

Surfing by the shore with the seagulls

Figure 11: User Interface of StoryWriter

input. The application will then display the newly synthesised image in the image section, highlighting the corresponding text in the caption section.

**Synthesised Images**

This area displays the images generated by the DM-GAN model. Users can scroll up and select any of the previously synthesised images. Clicking or hovering over any of the images will highlight the respective user-inputted captions (i.e., the input text used to synthesise that image in the model).

**Captions**

This area displays all the previous texts submitted by the user. The caption of the currently selected image is highlighted by default. As with the image area, users can scroll up and see any of the previously submitted texts.

### 4.3.3 Task Tracking

Given the risks involved in autobiographical expressive writing and the fact that our application did not include counseling or human-based interventions, we decided not to adopt the conventional expressive writing paradigm used in offline settings, as it might prove too risky in the application's anonymous, unsupervised, online setting. Instead, we opted to use a story-based expressive writing task that allows users to submit fictional narratives. Participants were instructed to 'write the introduction of a short story' and to 'describe its setting and place while being as imaginative as possible' (see [88]). They were also instructed to write for a set period of time (three minutes for Study 1, five minutes for Study 2).

To track the user's interaction with the application, we added an interface mechanism to record users' Telemetric data via timestamps and unique codes. We utilised SQLAlchemy to incorporate a database at the back-end of the StoryWriter application. Each user is assigned a unique 7-digit identification (ID) code, generated at their first input keystroke. The user's first keystroke also activates a countdown timer, which, once elapsed, reveals a unique completion code that participants can input when the system is deployed as part of a survey platform (e.g. Prolific) . Every time the user clicks on the 'submit' button to generate an image, we store the input text with a corresponding timestamp in our database. The small region in the top right of the interface displays the countdown timer and (once elapsed) the ID code used to identify the participants.

## 4.4 Study 1: Quantitative Experiment

We conducted an experiment with emotion- and application-related control groups to determine StoryWriter's efficacy in downregulating negative emotions. We hypothesised that users induced to feel negative emotions such as anger, sadness, anxiety, or stress would exhibit a greater reduction of negative emotion after using StoryWriter than users assigned to the control condition. An overview of the experimental procedure can be seen in Figure 12.

### 4.4.1 Participants

We collected data from a total of 421 participants from Prolific,[11] an online crowdsourcing website for psychological research. After excluding participants who failed attention checks or provided invalid responses (mismatched IDs or blank submissions), we included data from 388 participants in our study (men = 131, women = 252, other = 4, rather not say = 1; mean age = 30.1, SD = 8.9). Of these, 295 participants were randomly assigned to the negative emotion conditions (anxiety = 89, sadness = 99, stress = 107) and 93 participants were assigned to the neutral condition. Moreover,

[11] `https://prolific.co`

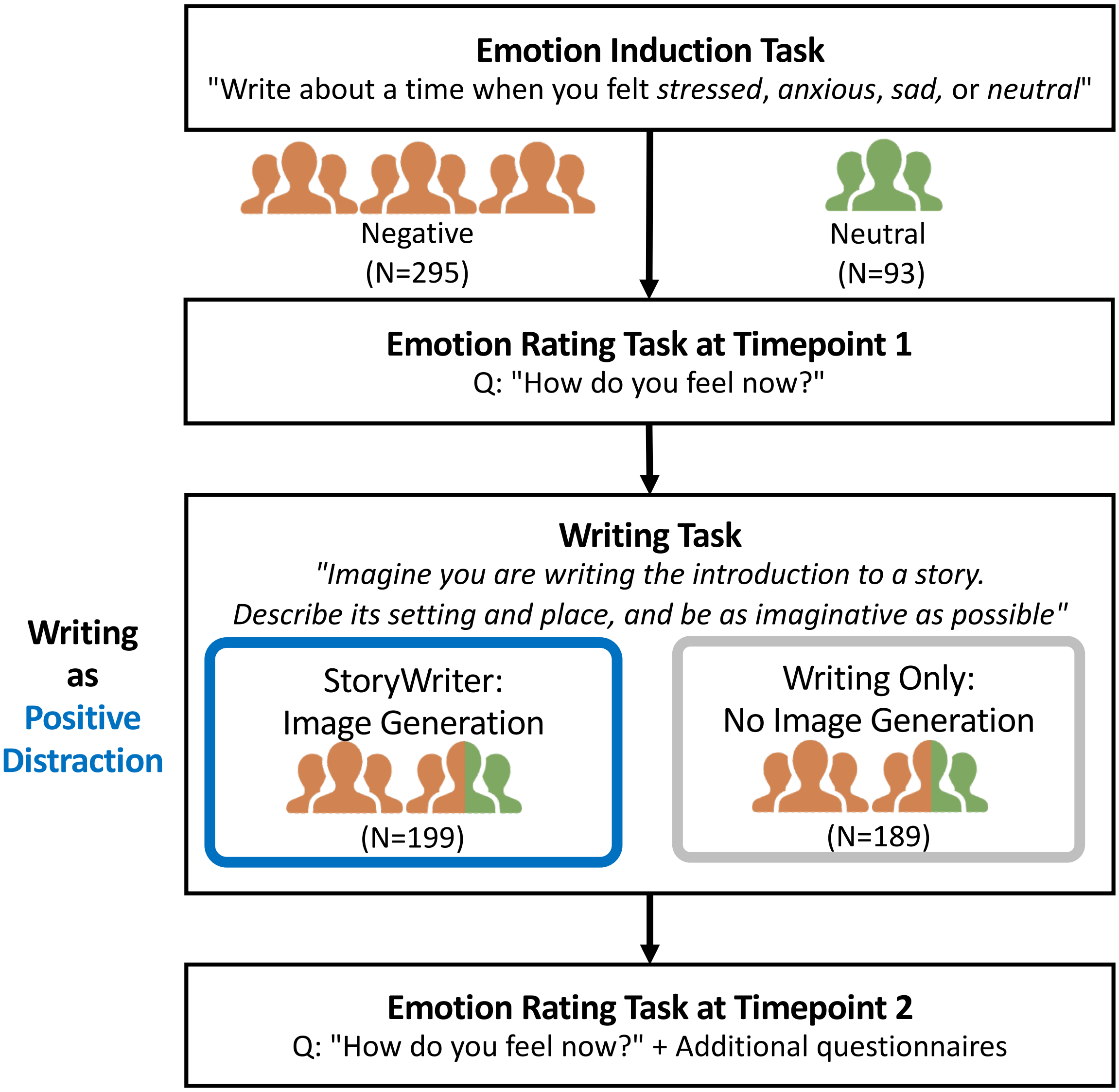

Emotion Induction Task
"Write about a time when you felt *stressed*, *anxious*, *sad,* or *neutral*"
Negative
(N=295)
Neutral
(N=93)
Emotion Rating Task at Timepoint 1
Q: "How do you feel now?"
Writing
as
Positive
Distraction
Writing Task
"Imagine you are writing the introduction to a story.
Describe its setting and place, and be as imaginative as possible"
StoryWriter:
Image Generation
(N=199)
Writing Only:
No Image Generation
(N=189)
Emotion Rating Task at Timepoint 2
Q: "How do you feel now?" + Additional questionnaires

Figure 12: The flow of the experiment. Participants were first induced to feel a negative (vs. neutral) emotion through an emotion induction task before using the StoryWriter application.

199 participants were assigned to the StoryWriter (writing with image generation) condition, while 189 participants were assigned to the control (writing only) condition. Participants were compensated with GBP £2.05 for their involvement in the study. All participants held either U.S. nationality or residence, and all but two (Spanish) reported using English as a first language. This study received approval from the Institutional Review Board of the Nara Institute of Science and Technology (2021-I-2), and participants provided informed consent before participating in the study, and were allowed to withdraw from the study at any point in time. Anonymised participant data is available online at our Open Science Framework (OSF) repository (anonymous review link:

https://osf.io/6cpjv/?view_only=4e355c167b99447b9a37add11dacb66f).

### 4.4.2 Procedure

As we were examining the efficacy of the application for users with pre-existing negative emotions, we started the study with an emotion induction task. After providing consent to participate, participants were induced to feel either negative emotions or a neutral emotion through an autobiographical recall task (see [89]) where participants recalled (and wrote about) a time they felt anxious, sad, stressed, or – for the neutral condition – performed an everyday routine task. Following this, participants rated the extent to which they felt the following emotions on a seven-point Likert scale, from 1 (not at all) to 7 (extremely): happy, awe, angry, touched/moved, grateful, excited, calm, sad, anxious, stressed, generally positive, and generally negative. We refer to this as Timepoint 1 (T1). Participants were then redirected to either the StoryWriter application or the control application (which included all the same features as the StoryWriter application except for image generation and display) and were asked to complete the task described in Section 4.3.3. They were also instructed to use the application for a minimum of three minutes. Thereafter, participants repeated the emotion rating task. We refer to this as Timepoint 2 (T2). Participants then evaluated the applications and completed measures of well-being (Satisfaction with Life Scale [SWLS] [90]), psychopathological tendencies (Depression, Anxiety, and Stress Scale [DASS] [91, 92]), and creativity (Self-Perceived Creativity scale [SPC] [93]). They also provided demographic information (age, gender, languages spoken, nationality). For usability and engagement, participants rated the 'fun' and 'ease of use' of the application.

### 4.4.3 Measurements and Analyses

To evaluate the changes in emotion before and after using the applications, we created a difference score (D) by subtracting T1 ratings for each emotion category from the T2 scores, focusing especially on the angry, sad, anxious, stressed, and generally negative items. Subjective evaluations of the application focused on emotion improvement ('using this writing application helped me get into a better mood') and usability ('the writing application was easy/fun to use').

We used the Telemetry infrastructure described in Section 4.3.3 to measure the character count of participants' narratives and the time spent between the first keystroke and the last 'submitted' text on the application. These figures were used to evaluate the extent to which participants actively engaged with the application. Participants then copied the generated completion ID codes from the application into the survey platform, which allowed us to combine both sources of data for each user.

All analyses were conducted using 2x2 ANCOVAs via the Jamovi software [94] with negative/neutral emotion induction (Emotion) and StoryWriter/control (Writing) conditions as the independent variable (IV). We also included SWLS, DASS (Depression, Anxiety, and Stress factors), and SPC as covariates. We used an alpha level of $p = .05$ for significance testing. All $p$-values reported are uncorrected. More detailed results (including non-significant findings) are available on our OSF repository (anonymous review link: `https://osf.io/6cpjv/?view_only=4e355c167b99447b9a37add11dacb66f`).

### 4.4.4 Results and Discussion

**Emotion Induction**

Users induced to feel negative emotions reported corresponding increases in negative emotions at T1. Users induced with negative emotions reported higher anger (F(1,192) = 18.72, $p < .001$), sadness (F(1,182) = 33.34, $p < .001$), anxiety (F(1,142) = 11.69, $p < .001$), stress (F(1,150) = 15.53, $p < .001$), and general negativity (F(1,173) = 15.28, $p < .001$), as well as lower calmness (F(1,152) = 19.54, $p < .001$), than those in the neutral emotion condition. This suggests that the emotion induction task was largely effective in its intended effect, and users in the negative emotion condition began their interactions with StoryWriter from a negative emotional state.

**Application Efficacy**

Of the primary variables considered in our experimental study, only anger, sadness, emotion improvement, and fun yielded statistically significant overall models when designated as the dependent variable. This suggests that participants' anger, sadness, emotion improvement, and fun were meaningfully impacted by the other variables in our study. However, none of the variables had significant main effects in both the Emotion (negative vs. neutral) and Writing (StoryWriter vs. control) conditions. The only main effects observed were with the SPC covariate, and only amongst the items for emotion improvement, fun, and ease of use. Namely, users with higher self-rated creativity were more likely to rate the app as effective in emotion improvement and as fun or easy to use (see Table 4).

Figure 13 shows the results of the StoryWriter and control conditions for our hypothesised effect (regulating negative emotions). Figure 14 shows the results for emotion improvement, usability and engagement, and Telemetric data. Figure 15 shows some examples of submitted narratives with their respective output images.

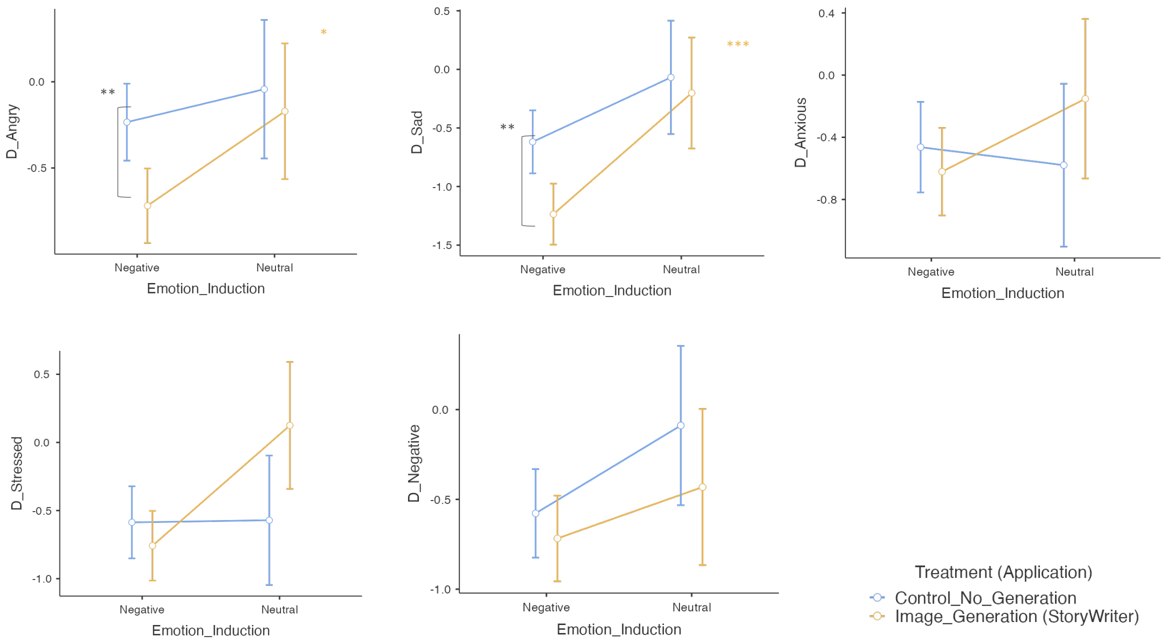


Figure 13: The results of the StoryWriter/control application for regulating negative emotion. ‘*’ indicates $p < .05$, ‘**’ indicates $p < .01$, and ‘***’ indicates $p < .001$

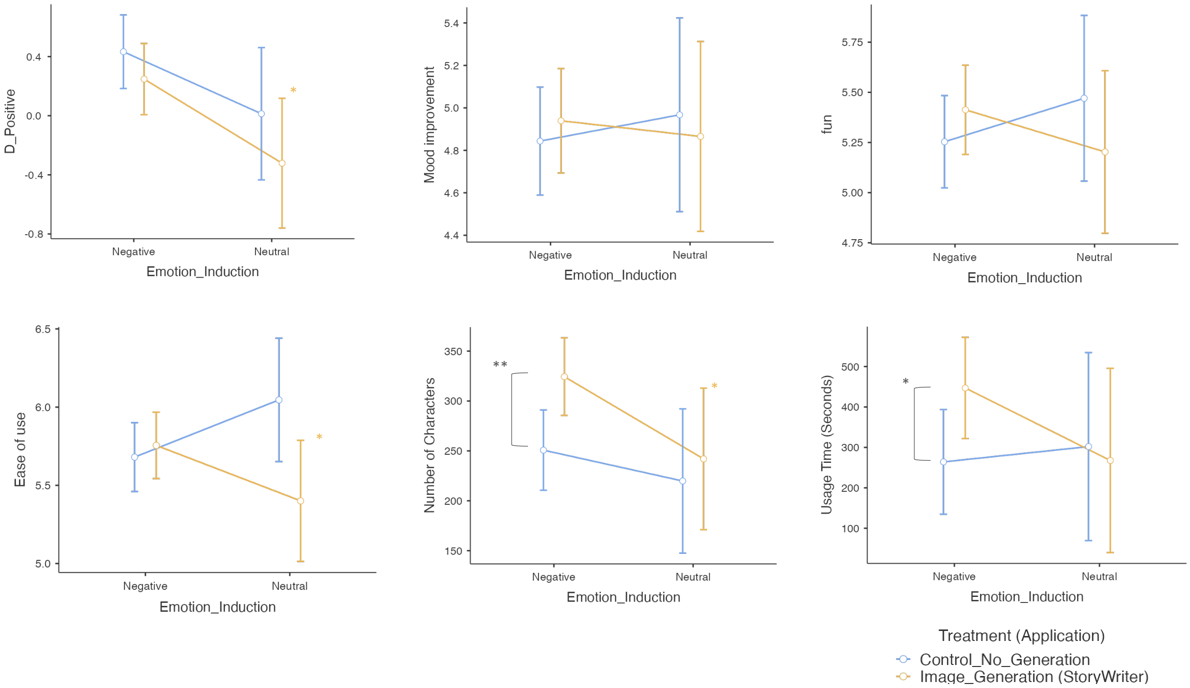
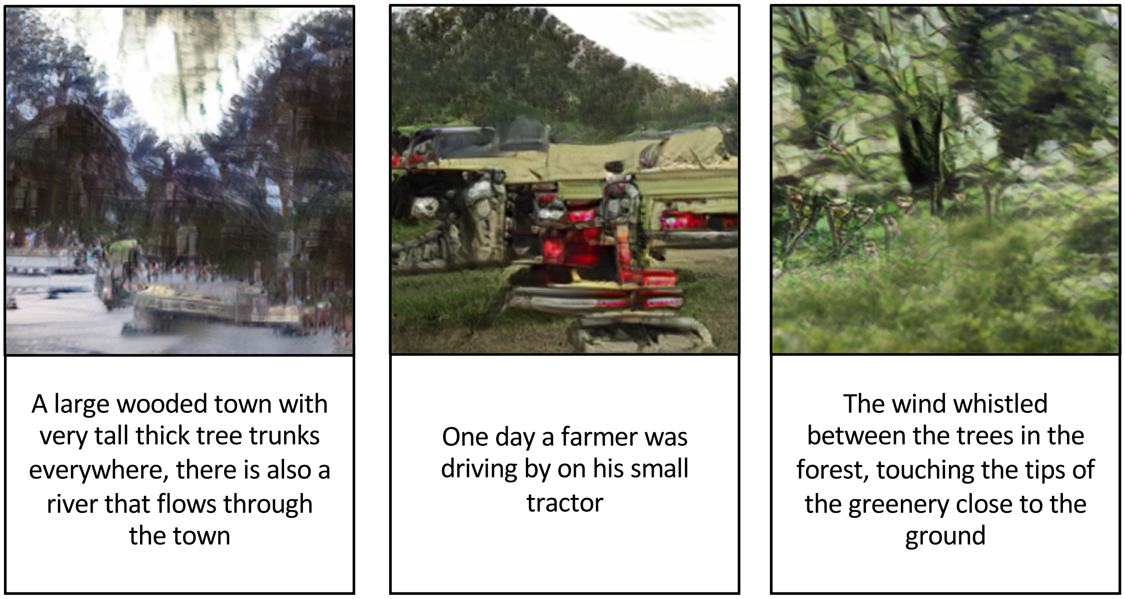


Figure 14: The results of the StoryWriter/control applications for emotion improvement, usability and engagement, and Telemetry. ‘*’ indicates $p < .05$, and ‘**’ indicates $p < .01$

| **Variables** | **Model** | | **Main Effects (SPC)** | | **Interaction Effects** | |
|---|---|---|---|---|---|---|
| | **F** | ***p*** | **F** | ***p*** | **F** | ***p*** |
| **Emotion** | | | | | | |
| Anger | F(23:364) = 1.97 | .005** | F(1:364) = 0.004 | 0.95 | F(1:364) = 0.47 | .59 |
| Sadness | F(23:364) = 1.78 | .016* | F(1:364) = 0.36 | 0.55 | F(1:364) = 1.32 | .25 |
| Anxiety | F(23:364) = 1.01 | .448 | F(1:364) = 0.92 | 0.338 | F(1:364) = 2.56 | .110 |
| Stress | F(23:364) = 0.49 | .070 | F(1:364) = 2.59 | 0.11 | F(1:364) = 0.25 | .616 |
| Negative | F(23:364) = 1.24 | .208 | F(1:364) = 1.23 | 0.27 | F(1:364) = 0.42 | .515 |
| Positive | F(23:364) = 1.34 | .139 | F(1:364) = 0.34 | 0.56 | F(1:364) = 0.42 | .515 |
| **Efficacy** | | | | | | |
| Emotion Impr. | F(23:363) = 1.99 | .005** | F(1:363) = 19.79 | <.001* | F(1:363) = 0.29 | .593 |
| **Usability** | | | | | | |
| Ease of Use | F(23:363) = 1.41 | .101 | F(1:363) = 4.24 | .040* | F(1:363) = 0.76 | 0.39 |
| Fun | F(23:364) = 1.99 | .005* | F(1:364) = 15.97 | < .001*** | F(1:364) = 1.44 | 0.23 |
| **Engagement** | | | | | | |
| Character Count | F(23:364) = 1.49 | .070 | F(1:364) = 3.15 | 0.077 | F(1:364) = 0.91 | .339 |
| Time Spent | F(23:364) = 0.41 | .993 | F(1:364) = 0.06 | 0.80 | F(1:364) = 0.39 | .530 |

Table 4: Results from respective Analysis of Covariate (ANCOVA) models. We report the main effects of creativity (SPC) and the interaction effect between induced emotion (negative versus neutral) and task (StoryWriter versus control). Note: * indicates $p < .05$, ** indicates $p < .01$, and *** indicates $p < .001$. A detailed list of results (including effect sizes) is available on our OSF repository.



A large wooded town with very tall thick tree trunks everywhere, there is also a river that flows through the town

One day a farmer was driving by on his small tractor

The wind whistled between the trees in the forest, touching the tips of the greenery close to the ground

Figure 15: Examples of written narratives and the respective images synthesised by the text-to-image DM-GAN model

| Variables | StoryWriter | | | | | | | | Negative Induction | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Negative | | Neutral | | | | | | StoryWriter | | Control | | | | | |
| | **M** | **SE** | **M** | **SE** | **t** | **df** | **p** | **d** | **M** | **SE** | **M** | **SE** | **t** | **df** | **p** | **d** |
| **Emotion** | | | | | | | | | | | | | | | | |
| Anger | -0.72 | 0.11 | -0.17 | 0.20 | -2.40 | 364 | .017* | -0.40 | -0.72 | 0.11 | -0.23 | 0.11 | 3.07 | 364 | .002** | 0.36 |
| Sadness | -1.24 | 0.13 | -0.20 | 0.24 | -3.76 | 364 | <.001*** | -0.63 | -1.24 | 0.13 | -0.62 | 0.14 | 3.24 | 364 | .001** | 0.38 |
| Anxiety | -0.61 | 0.14 | -0.15 | 0.26 | -1.58 | 364 | .116 | – | -0.61 | 0.14 | -0.46 | 0.15 | 0.76 | 364 | .445 | – |
| Stress | -0.76 | 0.13 | -0.13 | 0.24 | -3.27 | 364 | .001** | -0.55 | -0.76 | 0.13 | -0.59 | 0.13 | 0.92 | 364 | .359 | – |
| Neg. Affect | -0.72 | 0.12 | -0.43 | 0.22 | -1.14 | 364 | .257 | – | -0.72 | 0.12 | -0.58 | 0.13 | 0.80 | 364 | .423 | – |
| Pos. Affect | 0.25 | 0.12 | -0.32 | 0.22 | 2.24 | 364 | .026* | 0.38 | 0.25 | 0.12 | 0.43 | 0.13 | 1.05 | 364 | .294 | – |
| **Efficacy** | | | | | | | | | | | | | | | | |
| Emotion Impr. | 4.94 | 0.13 | 4.87 | 0.223 | 0.28 | 364 | .776 | – | 4.94 | 0.13 | 4.84 | 0.13 | -0.53 | 364 | .595 | – |
| **Usability** | | | | | | | | | | | | | | | | |
| Ease of Use | 5.76 | 0.11 | 5.40 | 0.20 | 1.58 | 363 | 0.115 | – | 5.76 | 0.11 | 5.68 | 0.11 | -0.48 | 363 | 0.631 | – |
| Fun | 5.41 | 0.11 | 5.20 | 0.21 | 0.89 | 364 | 0.372 | – | 5.41 | 0.11 | 5.25 | 0.12 | -0.98 | 364 | 0.328 | – |
| **Engagement** | | | | | | | | | | | | | | | | |
| Cha. Count | 324 | 19.8 | 242 | 36.0 | -2.04 | 364 | .046* | 0.34 | 324 | 19.8 | 251 | 20.5 | -2.56 | 364 | .010* | -0.30 |
| Time Spent | 447 | 63.6 | 268 | 115.9 | 1.36 | 364 | .176 | – | 447 | 63.6 | 264 | 65.8 | -2.00 | 364 | .046* | -0.23 |

Table 5: Table of planned post-hoc pairwise comparisons. Note: * indicates $p < .05$, ** indicates $p < .01$, and *** indicates $p < .001$. Effect sizes (Cohen's d) were reported only for significant effects.

Results of planned post-hoc comparisons are shown in Table 2. Significant differences were observed between variables across conditions. Anger, sadness, and character count yielded differences between both sets of conditions: users assigned to StoryWriter exhibited significantly reduced anger and sadness and increased character counts when induced with negative rather than neutral emotions, while users assigned to the negative emotion condition exhibited similar outcomes when using StoryWriter rather than the control application. Some variables, such as stress, general positive affect, ease of use, and time spent, exhibited significant differences in only one set of study conditions. For instance, users assigned to StoryWriter exhibited significantly reduced stress and significantly increased general positive affect when induced with negative rather than neutral emotions, but there were no significant differences for users assigned to the negative emotion condition when they used StoryWriter compared to the control application. On the other hand, users assigned to the negative emotion condition spent significantly more time writing when they used StoryWriter rather than the control application, but there were no significant differences for StoryWriter users assigned to the negative vs. neutral emotion condition. Participants in the control condition generally found the application easier to use than those in the StoryWriter condition (F(1,363) = 5.50, $p$ = .020), which is unsurprising given the relative difficulty of writing for image generation purposes. Finally, variables like anxiety, general negative affect, emotion improvement, and fun yielded no significant differences between either set of study conditions. Moreover, none of the variables had significant interaction effects between the Emotion and Writing conditions.

In short, while there is some evidence that (1) StoryWriter allowed users to positively distract themselves from negative emotions and (2) self-reported creativity had an impact on users' engagement with the application, we do not have sufficient evidence to determine the overall efficacy of StoryWriter in achieving its intended effect. Moreover, we examined individuals with induced negative emotional states, which differs from our initial objective of assessing positive distraction through creative writing for depressed individuals. Given the complexity of these unanswered questions, we decided to run a qualitative user study focusing more on the mechanisms involved in positive distraction amongst depressive StoryWriter users.

## 4.5 Study 2: Qualitative User Study

In Study 2, we aimed to examine how users in our target user-base (users with depression) perceived the experience of using StoryWriter. Though we collected stories and images from the experimental study, we were eager to solicit users' appraisals of the images and obtain more open-ended feedback about their emotions after using the application. Due to operational constraints and the ongoing impact of COVID-19, we chose to administer this qualitative component remotely as a long-form questionnaire to build a foundational understanding of user impressions that could inform future user studies involving text-to-image therapeutic writing applications.

### 4.5.1 Participants

We first used Prolific to recruit two sets of participants to a prescreening survey, which included the DASS, SPC, and SWLS inventories from Study 1, alongside demographic questions and questions about participants' creative pastimes (both writing and other artforms), mental health diagnoses, and COVID and holiday-related experiences (the study was completed between December 2021 and January 2022). This would provide context for their current emotional baseline and writing approach. The first participant set targeted a general population above the age of 20 who are located in the U.S. (following Study 1's inclusion criteria), while the second set included such users with past mental health conditions (as a filter applied through Prolific). We used this stratified sampling strategy to ensure that results reflect diverse perspectives – especially given the prospective therapeutic use of our application – while also including enough neutral-range participants to contextualise findings from Study 1.

From the 314 individuals who engaged with our prescreening survey (233 from the mental health-specific group, 81 from the general population), we received 302 responses, 271 of which were complete and usable. After excluding one entry that didn't meet the inclusion criteria, we ranked the remaining entries based on their response quality (length and depth), DASS scores, SPC scores, and diversity, prioritising greater engagement with the survey and more stratified DASS scores (highest and lowest), and taking care to include some participants with the least represented creativity and demographic information in the sample (for instance, low SPC, high income, etc.). Ultimately, we invited 100 of these individuals to participate in the main qualitative survey – 50 each from the high and low DASS groups (glossed as the 'DASS' and 'Neutral' groups). Of these, 65 responded, yielding 62 total responses, 54 of which were complete and usable (31 DASS, 23 Neutral). Thus, our analysis includes 54 participants (30 women, 19 men, 5 non-binary; mean age = 33.7, SD = 11.7) who are either U.S. citizens or residents and who, with one exception (Chinese), speak English as their dominant language. Participants in the prescreening survey were compensated with GBP £0.84, while participants in the main survey were compensated with GBP £3.00. This study received approval from the Institutional Review Board of the Nara Institute of Science and Technology (2021-I-2), and participants provided informed consent before participating in the study, and were allowed to withdraw from the study at any point in time.

### 4.5.2 Procedure

All participants in the main survey engaged with the StoryWriter application (the control application from Study 1 was not used in Study 2). We devised two versions of the same survey: one with an induction task that invited respondents to reflect on a recent sadness-, anxiety-, or stress-inducing experience (following Study 1), and one without this task. This was to examine differences between baseline (typical) and escalated (atypical) negative emotions. To account for extraneous factors and ascertain baseline emotions, each survey began with an open-ended question inviting participants to talk about their day and current feelings, as well as a multi-select question where participants could record common external factors contributing to their current mindset (e.g., tiredness or physical discomfort, preoccupation with life events, etc.). We did not include structured emotion checks because the objective was not to use this information for empirical analysis, and because we did not want to cause participants to draw preemptive conclusions about the purpose of our study (which could affect the way they answered later questions).

We directed DASS and Neutral participants into each of the two surveys (glossed here as 'ES', for the emotion-induction survey, and 'NS', for the naturalistic (no-induction) survey), taking particular care to allocate enough DASS users to the ES survey to provide insight into prospective application use cases; 19 DASS and 10 Neutral participants completed the ES survey, while 12 DASS and 13 Neutral participants completed the NS survey. After providing consent to participate in the study, participants completed the brief reflection, the priming task (where applicable), and then immediately proceeded to sample the StoryWriter tool with the same task used in the experimental study (Section 4.3.3). For this study, we adjusted the timer from three to five minutes in order to encourage slightly longer stories, and we adjusted the instruction text with updated study information, but preserved all other features.

Following the StoryWriter segment, participants were invited to reflect on their writing process, including the decisions that went into their particular story and the overall ease or difficulty of writing. Thereafter, participants were asked to elaborate on their current emotional state and to rank the factors influencing their emotions (those specified at the beginning of the survey alongside StoryWriter components: interface, images, and writing process). Upon completing this portion, participants proceeded to answer a set of user experience questions ('what did you like and dislike about the application?' 'what did you think of the application's layout?'), questions about the images' suitability and effect on the writing process, and questions about their satisfaction with their story and appraisal of the StoryWriter application.

### 4.5.3 Analyses

We used NVivo to process and code the survey responses and the accompanying stories and images. One researcher undertook this portion of the project, deploying an analysis approach that is best described as Template Analysis: an iterative but systematic mode of thematic analysis involving the initial development and subsequent adjustment of a coding 'template' [95]. This means that codes were developed both deductively (from pre-defined themes in Study 1) and inductively (based on new patterns observed in the data; p. 203). *A priori* themes included positive vs. negative. vs. neutral distraction, emotion change and valence, and specific emotion groups (anger, sadness, anxiety, stress), and they were progressively tailored to shed light on the following questions: (1) what do distraction and engagement look like in practice? (2) if there is distraction, what precisely is causing it? (3) what makes distraction positive (vs. neutral or negative)? Furthermore, as per Template Analysis, codes were arranged both hierarchically (within broader themes) and laterally (based on 'integrative' resonances across clusters; p. 204). Our hierarchical arrangements tended to specify domains of data, or topic groups informed by our survey questions (our final template is organised by survey phases, like Pre-Study, Study, Stories & Images, and by application features: images, interface, writing), whereas our lateral arrangements specified emotions that appeared at various stages of the study. We determined this scope to be sufficient for describing individuals' views and experiences over the course of their interaction with the tool, and for the purpose of contextualising the findings of Study 1.

We generally favored semantic (rather than latent) meanings, or codes that described participants' wording/express orientations, though we did interpret latencies in some behaviors – for example, negative distraction in instances where participants deliberately avoided looking at the StoryWriter images). These latent meanings are not intended as absolute standards for how the material should be interpreted, but as one of many possible interpretations that can enrich our understanding of the observed phenomena (e.g., distraction, engagement, emotion downregulation, etc.). Indeed, we took care during the analysis to identify alternate ways in which the target phenomena were being described – for instance, distraction occurring relative to the writing process (as a negative, unwelcome detour) rather than to negative emotions (as a positive intervention, as defined in Study 1). Thus, we would say that we performed our analysis from a 'contextual constructivist' position [96], which 'assumes that there are always multiple interpretations to be made of any phenomenon, and that these depend upon the position of the researcher and the specific social context of the research' [95] (p. 205). This position is not typically amenable to quantitative reliability measures like inter-rater reliability, given that 'all accounts, whether those of participants or of researchers, are understood to be imbued with subjectivity and therefore not *prima facie* invalidated by conflicting with alternative perspectives [96] (pp. 9–10). Rather, we pursued consistency and comprehensiveness by (1) designing our survey to examine given concepts from multiple vantage points (for instance, through open-ended questions as well as scales and ranking questions), providing a degree of triangulation that will become evident in our Results and Discussion; (2) recruiting new participants until we were satisfied with the coverage of topics; and (3) reviewing our codes and transcripts several times, as per typical Template Analysis procedure, and redeploying/standardising the template until it was able to accommodate all parts of the participants' responses to the survey questions.

Emotional states were coded at four points: baseline (the first emotion check-in question), induction (prompt that asked participants to reflect upon and recount a recent negative event), story (based on the emotions conveyed by characters or projected onto the story environment), and post-study (emotional check-in question). We began by extracting the affective language used by the participants (e.g., *irritated*), then grouped these together into related sentiment groups (e.g., *irritated*, *annoyed*, *bothered*), which became part of our larger, *a priori* emotion themes (namely, anger-adjacent, anxiety/stress-adjacent, and sadness-adjacent). We distinguished stress from anxiety based on the degree to which the sentiment could be interpreted as situational (instigated by circumstances) vs. dispositional (endemic, habitual). Changes in emotional states were determined primarily based on changes in the number of items coded to each emotion code, and corroborated by examining emotion change in terms of the presence of new affective language categories or absence of previous ones.

### 4.5.4 Results and Discussion

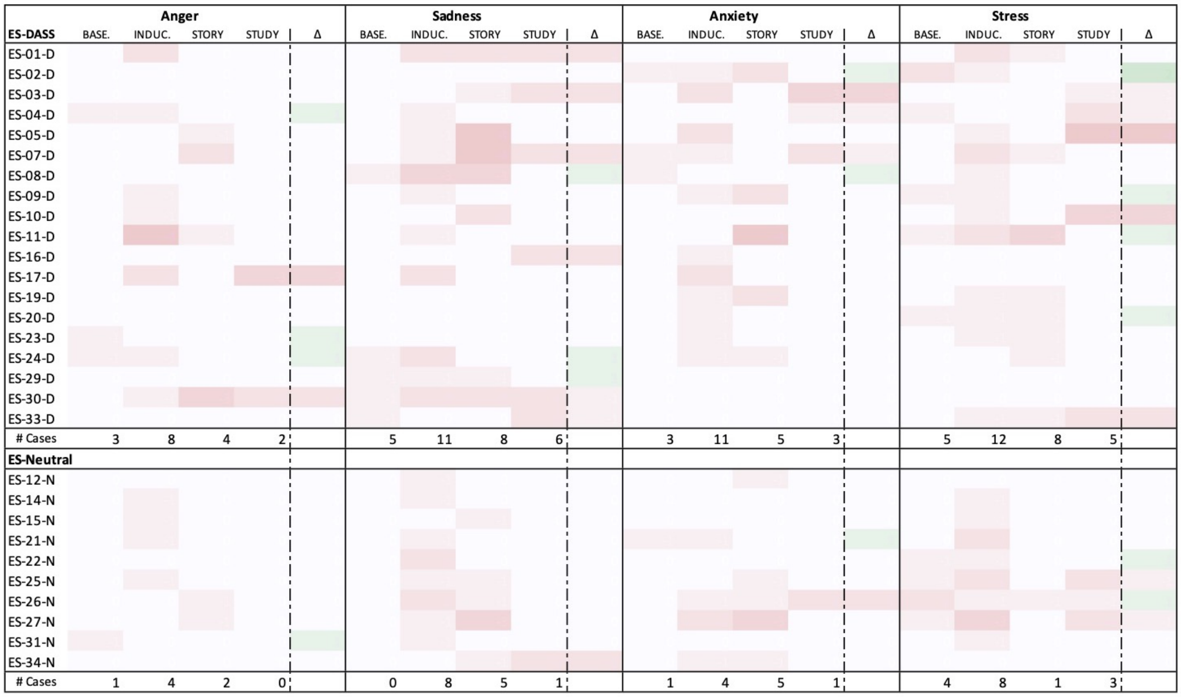
Our overarching assumption was that GAN-generated images could help to moderate users' negative emotions during online writing activities by providing positive distraction. Additionally, we hypothesised that participants induced with negative emotions would achieve better emotional outcomes than neutral participants. Below we discuss these results in light of the qualitative findings.

#### Negative Emotion Downregulation

According to the results of Study 1, participants with negative emotions who used StoryWriter exhibited small but significant reductions in anger and sadness, and some reduction in stress, particularly when compared to users who either did not use the

StoryWriter application or users who were not induced with negative emotions prior to using the application.

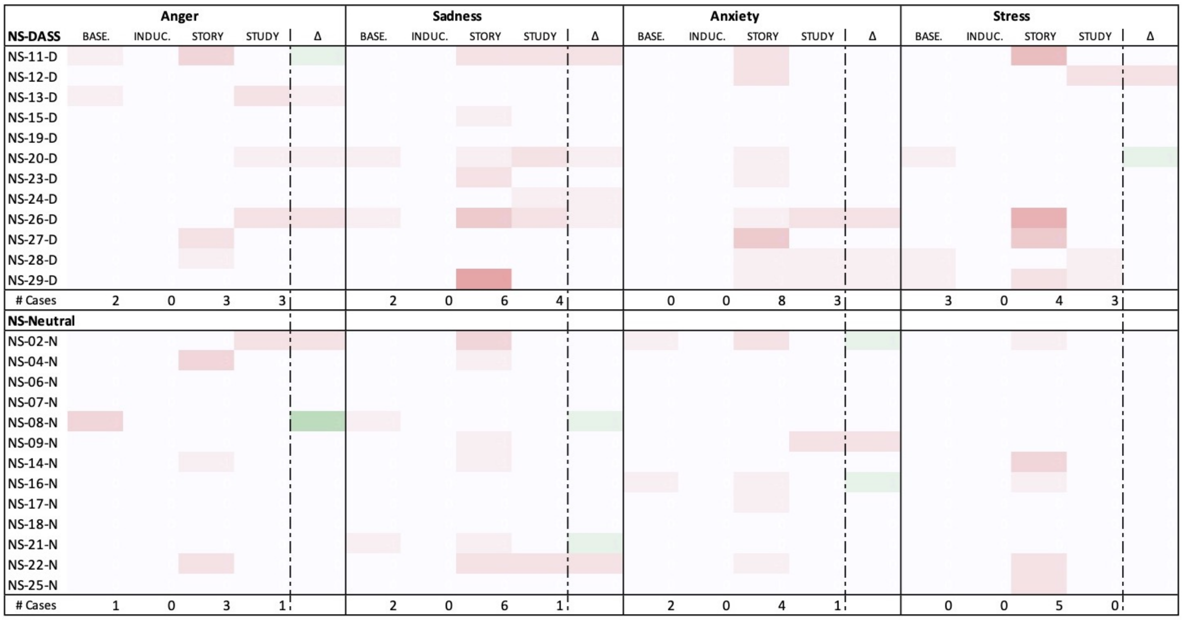
Similarly, though Study 2's qualitative results provided only limited evidence for negative emotion downregulation due to the small sample size, it did appear that more negative baseline emotions corresponded somewhat with better emotion outcomes post-study (see Figures 16–18). Induced emotional escalation, however, did not correspond much with emotion outcomes.



| | Anger | | | | | Sadness | | | | | Anxiety | | | | | Stress | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ES-DASS | BASE. | INDUC. | STORY | STUDY | Δ | BASE. | INDUC. | STORY | STUDY | Δ | BASE. | INDUC. | STORY | STUDY | Δ | BASE. | INDUC. | STORY | STUDY | Δ |
| ES-01-D | | | | | | | | | | | | | | | | | | | | |
| ES-02-D | | | | | | | | | | | | | | | | | | | | |
| ES-03-D | | | | | | | | | | | | | | | | | | | | |
| ES-04-D | | | | | | | | | | | | | | | | | | | | |
| ES-05-D | | | | | | | | | | | | | | | | | | | | |
| ES-07-D | | | | | | | | | | | | | | | | | | | | |
| ES-08-D | | | | | | | | | | | | | | | | | | | | |
| ES-09-D | | | | | | | | | | | | | | | | | | | | |
| ES-10-D | | | | | | | | | | | | | | | | | | | | |
| ES-11-D | | | | | | | | | | | | | | | | | | | | |
| ES-16-D | | | | | | | | | | | | | | | | | | | | |
| ES-17-D | | | | | | | | | | | | | | | | | | | | |
| ES-19-D | | | | | | | | | | | | | | | | | | | | |
| ES-20-D | | | | | | | | | | | | | | | | | | | | |
| ES-23-D | | | | | | | | | | | | | | | | | | | | |
| ES-24-D | | | | | | | | | | | | | | | | | | | | |
| ES-29-D | | | | | | | | | | | | | | | | | | | | |
| ES-30-D | | | | | | | | | | | | | | | | | | | | |
| ES-33-D | | | | | | | | | | | | | | | | | | | | |
| # Cases | 3 | 8 | 4 | 2 | | 5 | 11 | 8 | 6 | | 3 | 11 | 5 | 3 | | 5 | 12 | 8 | 5 | |
| ES-Neutral | | | | | | | | | | | | | | | | | | | | |
| ES-12-N | | | | | | | | | | | | | | | | | | | | |
| ES-14-N | | | | | | | | | | | | | | | | | | | | |
| ES-15-N | | | | | | | | | | | | | | | | | | | | |
| ES-21-N | | | | | | | | | | | | | | | | | | | | |
| ES-22-N | | | | | | | | | | | | | | | | | | | | |
| ES-25-N | | | | | | | | | | | | | | | | | | | | |
| ES-26-N | | | | | | | | | | | | | | | | | | | | |
| ES-27-N | | | | | | | | | | | | | | | | | | | | |
| ES-31-N | | | | | | | | | | | | | | | | | | | | |
| ES-34-N | | | | | | | | | | | | | | | | | | | | |
| # Cases | 1 | 4 | 2 | 0 | | 0 | 8 | 5 | 1 | | 1 | 4 | 5 | 1 | | 4 | 8 | 1 | 3 | |

Figure 16: Negative Emotional Trajectory of Participants in the ES Survey. Leftmost column lists the participants; center columns indicate instances of the given emotion at each time point (greater saturation represents more instances); rightmost column indicates change in the given emotion between Baseline and Post-Study time points (more green represents lower presence of given negative emotion, more red represents greater presence of given negative emotion). # Cases = number of participants who reported given emotion at each time point.

Over the course of thematic analysis, we assembled positive emotions into two groups (or themes), which we glossed as 'action-oriented' and 'contemplative'. Action-oriented positive emotions included adjectives like capable, competent, motivated, excited, curious, stimulated, and so on – all words that seemed to suggest a subsequent action ('capable of...', 'motivated to', 'curious to [see]...' etc.). Contemplative positive emotions included adjectives like comfortable, content, grateful, happy, hopeful, calm, etc. – words that seemed to suggest a kind of appreciative stasis. The distinction partly resembles the situational vs. dispositional divide used to distinguish stress from anxiety in user responses. Interestingly, there was a noticeable increase in action-oriented posi-

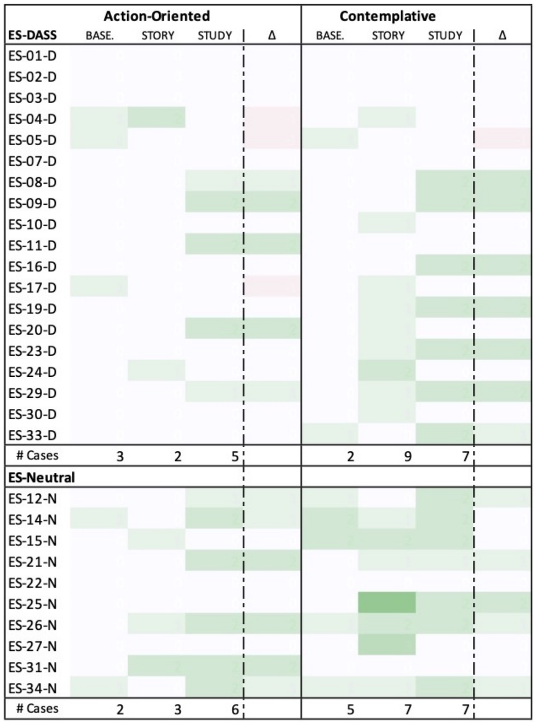


| | Anger | | | | | Sadness | | | | | Anxiety | | | | | Stress | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NS-DASS | BASE. | INDUC. | STORY | STUDY | Δ | BASE. | INDUC. | STORY | STUDY | Δ | BASE. | INDUC. | STORY | STUDY | Δ | BASE. | INDUC. | STORY | STUDY | Δ |
| NS-11-D | | | | | | | | | | | | | | | | | | | | |
| NS-12-D | | | | | | | | | | | | | | | | | | | | |
| NS-13-D | | | | | | | | | | | | | | | | | | | | |
| NS-15-D | | | | | | | | | | | | | | | | | | | | |
| NS-19-D | | | | | | | | | | | | | | | | | | | | |
| NS-20-D | | | | | | | | | | | | | | | | | | | | |
| NS-23-D | | | | | | | | | | | | | | | | | | | | |
| NS-24-D | | | | | | | | | | | | | | | | | | | | |
| NS-26-D | | | | | | | | | | | | | | | | | | | | |
| NS-27-D | | | | | | | | | | | | | | | | | | | | |
| NS-28-D | | | | | | | | | | | | | | | | | | | | |
| NS-29-D | | | | | | | | | | | | | | | | | | | | |
| # Cases | 2 | 0 | 3 | 3 | | 2 | 0 | 6 | 4 | | 0 | 0 | 8 | 3 | | 3 | 0 | 4 | 3 | |
| NS-Neutral | | | | | | | | | | | | | | | | | | | | |
| NS-02-N | | | | | | | | | | | | | | | | | | | | |
| NS-04-N | | | | | | | | | | | | | | | | | | | | |
| NS-06-N | | | | | | | | | | | | | | | | | | | | |
| NS-07-N | | | | | | | | | | | | | | | | | | | | |
| NS-08-N | | | | | | | | | | | | | | | | | | | | |
| NS-09-N | | | | | | | | | | | | | | | | | | | | |
| NS-14-N | | | | | | | | | | | | | | | | | | | | |
| NS-16-N | | | | | | | | | | | | | | | | | | | | |
| NS-17-N | | | | | | | | | | | | | | | | | | | | |
| NS-18-N | | | | | | | | | | | | | | | | | | | | |
| NS-21-N | | | | | | | | | | | | | | | | | | | | |
| NS-22-N | | | | | | | | | | | | | | | | | | | | |
| NS-25-N | | | | | | | | | | | | | | | | | | | | |
| # Cases | 1 | 0 | 3 | 1 | | 2 | 0 | 6 | 1 | | 2 | 0 | 4 | 1 | | 0 | 0 | 5 | 0 | |

Figure 17: Negative Emotional Trajectory of Participants in the NS Survey. Leftmost column lists the participants; center columns indicate instances of the given emotion at each time point (greater saturation represents more instances); rightmost column indicates change in the given emotion between Baseline and Post-Study time points (more green represents lower presence of given negative emotion, more red represents greater presence of given negative emotion). # Cases = number of participants who reported given emotion at each time point.

tive words over the course of the study, and in the number of participants who used such words (12 at baseline, 21 post-study). Contemplative positive emotions appeared more consistently, at least in the overarching group (there was a sharper increase amongst DASS participants in the ES survey), with 24 participants reporting contemplative emotions at baseline compared to 27 post-study.

These finding complement the results of Study 1, which showed that negative preexisting emotions yielded generally more positive emotion outcomes, but with differential effects for different emotions. The stronger results for anger and sadness in Study 1 (as compared to anxiety and stress) may be due to the approach (vs. avoidant) nature of these negative emotions [97]. Approach orientation (or action tendencies) describe the motivational system that drives individuals towards a certain desired goal or target. This is opposed to avoidant orientation, which describes the motivational system that drives users to withdraw from a target [98]. Sadness and anger – and, arguably, the specified action-oriented positive emotions – are associated with approach orientations: anger (and motivation, curiosity, competency, etc.) directly motivates individuals to approach that goal, and sadness arises from the presence of an unattainable goal [99]. While anger and sadness represent opposing ends of the approach–motivational spectrum, they nevertheless are mutually defined by their 'approach' orientation towards a



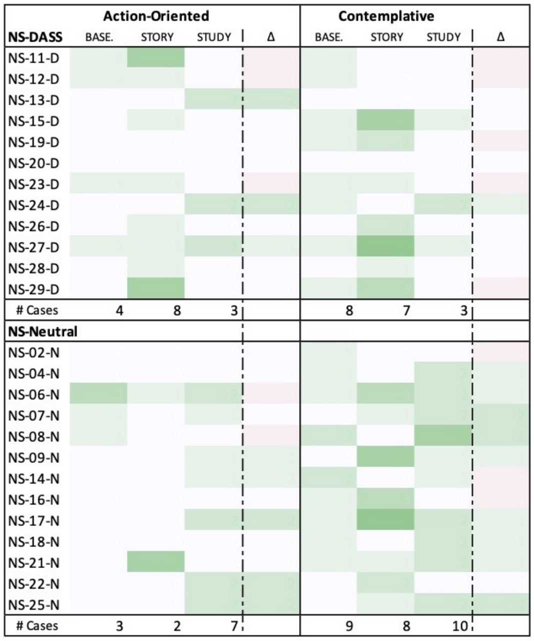


Figure 18: Positive Emotional Trajectory of Participants. Leftmost column lists the participants; center columns indicate instances of the given emotion at each time point (greater saturation represents more instances); rightmost column indicates change in the given emotion between Baseline and Post-Study time points (more green represents greater presence of given positive emotion, more red represents lower presence of given positive emotion). # Cases = number of participants who reported given emotion at each time point.

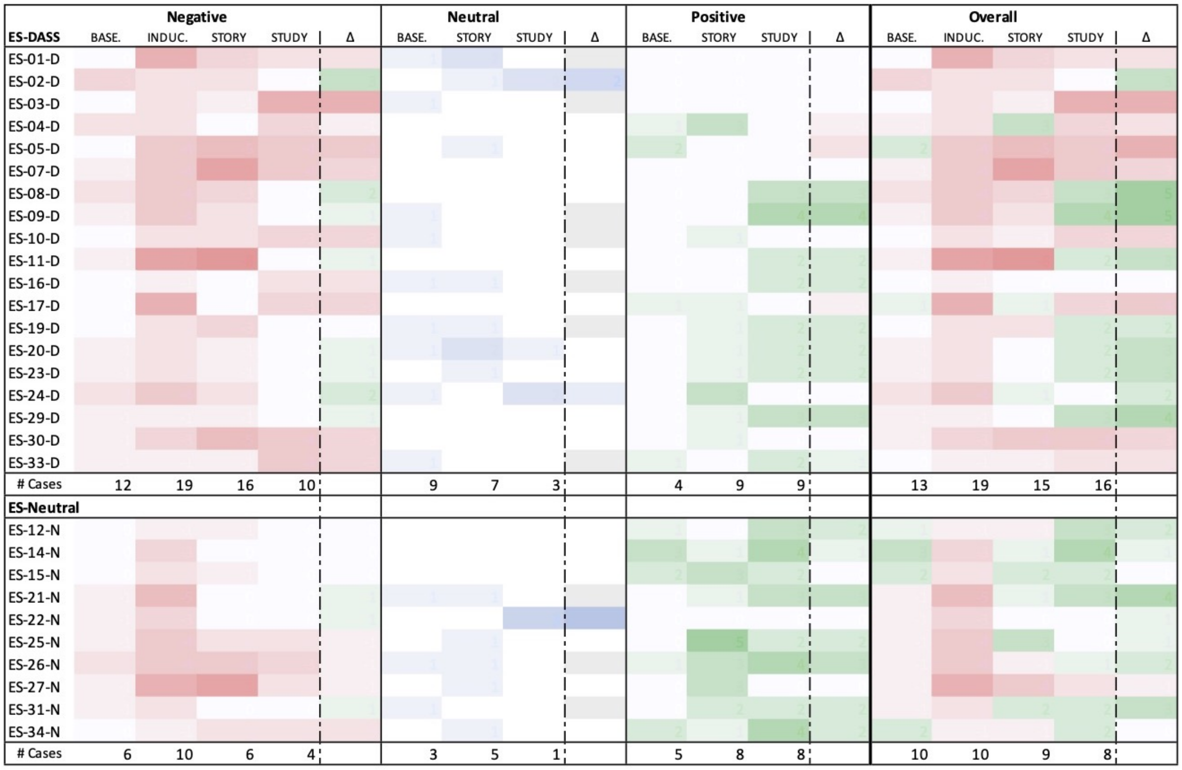


Figure 19: Total Emotional Trajectory of Participants in the ES Survey. Leftmost column lists the participants; center columns indicate the sum of given emotions at each time point (positive emotions as positive values, negative emotions as negative values, neutral emotions as zero value); rightmost column indicates change in total emotional valence between Baseline and Post-Study time points (more green represents more positive emotion change, more red represents more negative emotion change). # Cases = number of participants who reported a given emotion at each time point.

stimulus. Thus, anger and sadness may have been more effectively mitigated in Study 1 – and action-oriented positive emotions may have been more consistently bolstered in Study 2 – due to our external stimulus (StoryWriter), which distracts users by redirecting them towards a goal-oriented writing task. This congruence between emotion and task could drive an orientation-matching effect [99], thereby reducing the intensity of anger and sadness. By contrast, anxiety is avoidance-oriented, stress shows differing patterns, and contemplative emotions were defined by their relative stasis, so the goal-oriented distraction task employed here may not be as effective in downregulating anxiety and stress or upregulating contemplative positive emotions.

One question that remains involves the role of affect in the stories, given that our tool is designed to downregulate negative emotions that come up in the process of writing. Hypothetically, participants who used more negative words in their story should have gleaned more benefit from the image distraction. However, there was not much correspondence between negative story affect and post-study emotion outcomes,

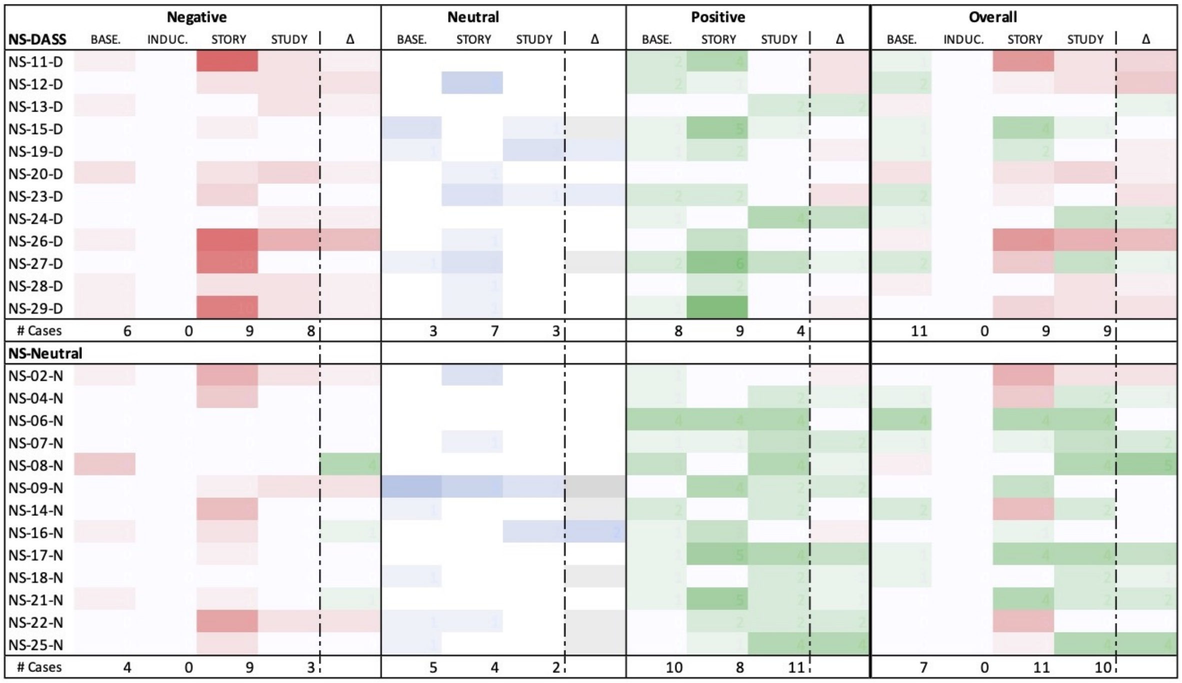


Figure 20: Total Emotional Trajectory of Participants in the NS Survey. Leftmost column lists the participants; center columns indicate the sum of given emotions at each time point (positive emotions as positive values, negative emotions as negative values, neutral emotions as zero value); rightmost column indicates change in total emotional valence between Baseline and Post-Study time points (greener represents more positive emotion change, redder represents more negative emotion change). # Cases = number of participants who reported a given emotion at each time point.

apart from outlier cases with very negative story affect, which seemed to correspond somewhat with worse emotion outcomes.

### Engagement Enhancement

Based on the Telemetry results of Study 1, users induced to feel negative emotions appeared to engage more with the writing task, writing longer passages and spending more time on the application. Previous research indicated that greater engagement yields better response quality [100,101], hence our findings likely suggested that StoryWriter could support therapeutic outcomes, as meaningful engagement and thoughtful outputs are crucial to the efficacy of writing exercises.

One potential concern is that engagement is not a singular and homogeneous construct – there may be different registers of engagement that cannot be detected through time spent and character counts alone. For instance, four participants in Study 2 described their engagement with the application as a form of entertainment – like a game, a novelty, or something fun to explore with friends – but not as a writing tool that they would apply to serious purposes. Moreover, six of the 15 respondents who indicated that they would not use the application for their own writing were quite prolific in their writing outputs, writing above the average character count of 551.39 and, in three cases, above 1,000.

More importantly, it is not immediately clear from time and character count figures *what* participants are engaging with when they use the application. Indeed, 35 of the 54 participants (64.82%) exhibited some kind of lack of engagement with the StoryWriter images. In some cases, this may have been due to a technical problem, like device compatibility issues (n = 5), in others to users' poor appraisal of the images themselves (n = 11), either as blurry and indecipherable or as unrelated to the story at hand. In most cases (n = 30, 55.56%), participants indicated that the images had no effect on their writing process. In contrast, 14 (25.93%)) participants reportedly engaged with some other aspect of the interface – specifically, the way the text formatting and simplicity of the tool helped them to focus on the writing process itself. One commented, 'It helped me take the writing process one sentence at a time, which isn't my usual style but was good for therapy purposes' (ES-31-N, woman, 26 years old). Another stated, 'I occasionally write short stories and this would be a great application to use. I write my short stories usually in MS Word and that application can be very distracting. It offers [way] to many choices and most of them I don't need' (NS-20-D, man, 35 years old).

In the qualitative study, we used two tactics/constructs to approximate the extent and object of users' engagement with the application: ease of writing and effect ranking. After participants had sampled the tool, we asked them, 'How easy or difficult was it to continue writing? Were there any points where you wanted to stop writing or using the application?' Though 50.00% (n = 27) of the participants found it easy to write, there was an even split between those who were motivated to keep writing (n = 12, 22.22%) and those who wished to stop writing (n = 12, 22.22%), either because of the submission pacing (i.e., the need to press enter/submit after each written sentence; n = 4), the images themselves (n = 3), or some other factor (n = 5). One stated, 'The images were very distracting. When the first image appeared, I really wanted to stop using the application all together. It [got] better the further I went, though' (ES-20-D, woman, 22 years old).

After reflecting on the writing process, participants reviewed their current emotions and attributed them to either study-related or previously-registered factors (e.g., tiredness, physical discomfort, preoccupation with life events, etc.). Of the 54 total participants, 22 (40.74%) attributed their emotions to the writing process first; 19 (35.19%) ranked non-study-related factors first; five (9.26%) ranked the StoryWriter interface first; and four (7.41%) ranked the StoryWriter images first (four indicated that none of the factors influenced their current emotion). Thus, the writing process was the only StoryWriter component deemed more effective (or engaging), on average, than other circumstances outside the app itself. In this light, any engagement outcomes noted in the experimental study could derive from the process of imaginative writing itself, which would explain the lack of significant differences in many emotion reduction outcomes between the StoryWriter and control groups for Study 1. The effect rankings and user commentary – which often framed the interface and images as relative/contingent to the user stories – indicate that the StoryWriter interface and images may only support negative emotion downregulation insofar as they support or mediate this writing process, rather than yielding distinct downregulating effects in and of themselves.

**Positive Distraction**

In Study 1, engagement was used as a proxy for distraction, with greater engagement indicating that users were more distracted by the mechanism. Additionally, positive distraction was defined as distraction that yielded positive emotional outcomes. Given these premises, the Telemetry results from Study 1 supported the conclusion that StoryWriter achieves some degree of negative emotion downregulation (for anger and sadness, and primarily amongst negatively-induced participants) through positive distraction. This would add AI-based interventions to the list of viable positive distraction tactics available for online settings (alongside digital art installments [79, 80] or active imagination tasks in supervised therapeutic activities [57, 58]).

In Study 2, we ascertained distraction through participants' responses to questions such as 'Why do you feel this way?' (following the emotion check-in) and 'How did the images affect your writing process, and what relationship did they have to your story?' We could also observe the extent to which study elements (like the interface, images, or writing process) superseded circumstantial factors in the aforementioned effect ranking. Moreover, we evaluated whether any given distraction was positive, neutral, or negative based on users' description of how certain features helped (positive) or hurt (negative) their writing process.

Of the 54 total participants, 27 (50.00%) explicitly stated that some study-related element affected their emotions. For instance, one participant who wrote a story based on recent events said, 'I think I feel this way because I'd been disappointed about the weather, and writing about it helped me process that (and 'strike back' at the snow)' (NS-08-N, woman, 42 years old). Another explained their more negative sentiments, saying, 'That's what the images made me feel like. The sadness is from what I was writing and the anxiousness is from the images and the recounting I had to do beforehand' (ES-03-D, woman, 20 years old). Fourteen of the 27 (or 51.85%) indicated that they had been positively affected, 13 due to the writing process (ex., 'I think the writing process was therapeutic and helped take my mind off my current troubles'; ES-31-N, woman, 26 years old), and one due to the app's overall distraction ('I have at minimum ... been effectively distracted from what I was doing and how I was feeling previously'; ES-11-D, woman, 41 years old). Four of the 27 (14.82%) were ambivalently affected, two due to the writing task or writing subject (ex., 'I feel [a bit sad but also ... hopeful] because I was thinking of issues that are very complex but with science there is hope for these issues being more understood and improved upon every day'; ES-34-N, woman, 39 years old), one due to the overall app experience ('I don't know why I feel [a bit disassociated], other than it's the first time I've had to myself all day and this exercise was a little unusual'; NS-16-N, woman, 45 years old), and one due to the survey (annoyed because 'The story portion was fun and imaginative but now we're on survey question which don't really require imagination'; NS-13-D, woman, 50 years old). Nine participants (33.33%) reported that they were negatively affected, four because of the images (ex., disturbed because 'It seemed to stack a lot of body parts or like parts of human bodies with animal bodies together and certain things just looked like chaotic'; NS-12-D, woman, 52 years old), two because of the writing process (ex., 'I always feel [depressed] when I start thinking. And to write a story, I need to think'; ES-07-D, woman, 22 years old), one because of the interface (lost, frustrated because 'I wish I could go back and change sentences to add more detail I'd forgotten about'; ES-17-D, man, 23 years old), one because the survey reset ('It's taken a toll on my focus and spirits to have to spend so much time redoing my work here'; NS-02-N, man, 31 years old), and one because of the priming task ('I feel like this because yesterday's events were very unpleasant and sad. It also reminded me that my friends didn't even ask if I had a good day yesterday'; ES-01-D, woman, 37 years old).

Overall, participants who explicitly reported more positive emotion changes tended to attribute this to the writing process, rather than to the application or images; those who explicitly attributed their emotions to the images tended to describe negative emotion change. This is somewhat reinforced by the effect ranking, where those who listed the writing process amongst the top three factors for their current emotion tended to report positive emotions (19 positive vs. 10 negative, 18 other), while those who listed the interface or image within the top three tended to report more neutral or ambivalent emotions (interface: 20 neutral vs. 11 positive, 5 negative; images: 24 neutral vs. 11 positive, 5 negative). Indeed, only one of the participants who ranked the StoryWriter images above extraneous factors and the writing process itself (n = 5) exhibited positive emotional change (one exhibited neutral emotion both before and after, and three reported worse emotions after the exercise because they found the images unsettling). The effect ranking also indicated that study elements superseded circumstantial factors for 23 participants (42.59%), and that four of the 35 individuals who had previously reported a non-study-related emotion influence ceased to consider it an influence after using the tool.

It is possible that the images/interface had a subconscious effect on participants, even if they did not report image/interface effects. For instance, many participants did acknowledge that the images successfully redirected their attention away (n = 28, 51.85%). Twelve of the 28 individuals (or 42.86%) thought that the images facilitated their writing. For example, participants used them to generate ideas (ex. 'I mostly used them when I was stuck. I would pause and then kind of let my thoughts wander and then look over at the images and see if I could generate anything based on them'; NS-11-D, man, 34 years old), visualise their stories ('They seemed to generate images of what my characters looked like, and I took that inspiration to describe them'; ES-25-N, non-binary, 24 years old), or adjust their tone ('I think they prompted me to be a bit darker and more fantastical than the simple romantic morning I initially planned to write about'; NS-27-D, man, 22 years old). Eight participants (28.57%) said that the images rendered no effect on their writing or emotions, while another eight reported that the images undermined the writing process to the extent that they avoided looking at/engaging with the images at all. For example, one participant stated, 'I enjoyed the thinking, planning and writing, but didn't like the pictures that came up next to each sentence. They threw my thoughts off balance and felt a bit creepy' (NS-06-N, woman, 48 years old). Another specified that their writing process was 'hindered ... with the pictures that made no sense' (NS-13-D, woman, 50 years old).

Thus, while it is clear that StoryWriter offers a good deal of distraction, it is unclear to what extent we can label the distraction as positive. This is especially evident because, amongst the 12 who reported positive image effects for writing, only four reported improved emotions thereafter (five reported worsened emotions). Only one of these 12 participants explicitly linked their emotion changes to the images themselves, and only five ranked images above extraneous factors in terms of their effect on their emotions.

### Users' Evaluations of the StoryWriter Features

In Study 1, participants evaluated the efficacy, usability, and experience of using StoryWriter, ultimately determining that it was slightly more fun, harder to use, and somewhat more efficacious for emotion improvement than the control. In Study 2, we provided participants with opportunities to divulge their impressions of the images, interface, and writing results in a more open-ended way, alongside more structured rankings for ease of use and image suitability.

*Interface.* Of the 44 participants who provided descriptions of the interface, 40

(90.91%) used positive words to describe the layout, usability, or application overall. In terms of layout, participants described the app as *functional* (n = 6), *visually appealing* (n = 3), and especially *simple*, *clean*, or *minimal* (n = 18). On the usability side, participants described the app as *fun*, *entertaining*, or *funny* (n = 5), *helpful*, *useful*, or *convenient* (n = 3), *intriguing* (n = 2), and especially *easy to use* or *easy to read* (n = 17). This was more or less corroborated by the numeric ease of use score participants assigned through a sliding scale, with zero indicating extreme ease and 100 extreme difficulty (the average for all participants was 25.98, SD = 22.30). Positive descriptions of the app as a whole included adjectives like *nice* (n = 7), *novel* (n = 1) and *quick* (n = 1). Meanwhile, 10 of the 44 specified participants (22.73%) used neutral words to describe the app, such as *fine* or *okay*, while four (9.10%) used negative descriptors such as *awkward*, *confusing*, or *plain*. Despite the predominantly positive appraisal, 41 (75.93%) of the 54 total participants frequently acknowledged specific issues with the interface. Foremost among these was bad pacing, or the way that submitting sentences one at a time slowed down the writing process (n = 27). Other issues included the lack of some word processing or customisation features (n = 16), including the ability to tailor the AI itself; layout grievances, such as the location of the instructions, the size of the text box, or the relative positions of the text input and image output (n = 12); glitch and compatibility issues, namely with mobile phones (n = 11); and punctuation and formatting issues, such as the omission of full stops (n = 9).

*Images.* Of the 31 participants who deployed adjectives to describe the StoryWriter images, 22 (70.97%) included negative words like *scary*, *unsettling*, or *creepy* (n = 15), *confusing*, *chaotic*, or *incoherent* (n = 6), *stressful* (n = 1), and *boring* (n = 1). Twelve (38.71%) included neutral descriptors like *abstract* (n = 6), *curious*, *odd*, or *peculiar* (n = 6), *okay* (n = 2), or *psychedelic* (n = 1). Thirteen (41.94%) included positive descriptors like *interesting* or *imaginative* (n = 12), *funny* (n = 2), and *relaxing* (n = 1). Of the 54 total participants, 38 individuals (70.37%) reported specific issues with the images. These participants took particular issue with the irrelevance of the images for their stories (n = 28), and with the quality of the images as a whole (n = 24). Some even commented on the limited scope of the images (n = 6), in that there was little variation between images and no opportunity to select between image options. One suggested that images be excluded from the application altogether. Nonetheless, 46.30% of all participants (n = 25) liked the concept of automatic image generation in theory, describing it as *interesting*, *fun*, or *novel*, and some did glean actual enjoyment from the application and/or confirm that some of the images matched the tone of their story (n = 7, 12.96%). Overall, the average suitability rating (indicated on a sliding scale) suggested that only 26.63% of images were suitable for the writing task.

*Writing.* After completing most of the survey, participants were asked to disclose their satisfaction with their story as a way of evaluating the writing process. Roughly half (n = 28, 51.85%) of all participants reported that they were satisfied with their story, with some (n = 5, 9.26%) indicating that there were aspects of their story that they would like to explore further. Fifteen individuals (27.78%) were not satisfied with their story, and 11 (20.37%) felt somewhat neutral about their story. In terms of issues or suggestions, 16 participants (29.63%) imagined improvements to the writing prompt and instructions, or to the writing tool itself so as to enable more intentional, guided writing. One participant explained, 'I would like if there was a new prompt to each submission. I keep seeing the same prompt (write the introduction of a short story) at the instructions section but I've already wrote a body and an ending' (ES-07-D, woman, 22 years old). Another suggested, 'Have a box at the side that threw out descriptive adjectives/metaphors etc. that might relate to what I was writing. Maybe even some suggested whole phrases' (NS-06-N, woman, 48 years old).

*Overall Evaluation.* Participants' overall evaluation of the images was rather decisive: of the 19 participants who provided such an evaluation, 11 (57.89%) indicated their dislike and 9 (47.37%) indicated their neutrality (ex., 'I neither liked nor disliked any of them'; NS-24-D, woman, 43 years old). While the participants took time to describe individual images that they liked more than others, only two participants mentioned that they liked (or, in one case, 'didn't mind'; NS-08-N, woman, 42 years old) the images as a whole. One of these individuals explained, 'Although I liked the photos, the real photos were going on inside my head' (NS-17-N, woman, 63 years old), indicating that the images were, to some extent, superficial. Responses to the application were more positive: of the 47 individuals who provided such an evaluation, 17 (36.17%) found the application to be sufficient (nothing more to add) or promising, while 11 (23.40%) indicated that they liked the application or its layout, with five (10.64%) indicating that they would use it again. Respondents particularly prized the motivating and streamlining aspects of the tool, which helped them to focus their thoughts on specific details in the story. Fifteen of the 47 individuals (31.92%) regarded the app as insufficient for story writing, indicating that they would not use the tool again (at least barring substantial changes), though only three (6.38%) explicitly stated their dislike for the application. Thirteen participants (27.66%) did not see much difference between the application and other writing tools, while 11 (23.40%) did not feel any particular way about the application or their experience using it. Some users envisioned interesting use cases for the application; one in particular stated that the app could help individuals with aphantasia – an inability to form mental images – to make their creative processes more visual. Although many participants commented on the distracting quality of the application or its features, and a few reported actual or anticipated therapeutic benefits, none envisioned a use case for the application that involved emotion regulation.

## 4.6 Discussion and Conclusion

In this paper, we reported the development and user evaluation of StoryWriter, an AI-enhanced writing application that uses Dynamic Memory-Generative Adversarial Networks to generate real-time images from users' written stories. These images are intended to serve as positive distractions that mitigate the impact of negative emotions in online writing activities. Based on our experimental study (Study 1) with 388 users and two control conditions (neutral emotion vs. negative emotion and no image generation application vs. StoryWriter application), we found that:

1. Users who were assigned to StoryWriter and/or induced with a negative emotion exhibited lower post-study anger and sadness than those in the control conditions, but not lower anxiety or stress

2. Users were somewhat successful in downregulating negative emotions through the application, but not in upregulating positive emotions (indicating certain boundary effects)

3. Users induced with a negative emotion wrote significantly more characters and spent more time writing than those in the control condition

While these results appear to support the usage of GAN-based image generation as a tool for emotion regulation, Study 2 – our qualitative study carried out with 54 users – yielded somewhat different insights:

1. Though users with pre-existing negative emotions did seem to be at least somewhat distracted from these emotions when using the application, the distraction was not consistently positive, and there was not much of an effect for induced escalated emotions.

2. Though users in Study 1 did not exhibit substantial positive emotion upregulation, some users in Study 2 did exhibit some improvements in certain positive emotions – namely, those with a bias towards action.

3. Furthermore, increased usage of the application did not necessarily mean that users were positively engaged with all (or even target) features of the StoryWriter application. Indeed, users appeared to attribute emotion regulatory benefits to the writing task, but not to the image generation feature.

4. Independent of other safety indicators, many participants found the images to be grotesque and unsettling, even to the extent that some participants avoided looking at them.

Taking these findings together, and coupled with the lack of positive appraisals of the application's images, we find that even if StoryWriter was successful in distracting participants (an outcome that applied to only about half of the participants in Study 2), the emotion regulatory mechanisms it facilitates appear unstable and driven by the fictional writing (therapy) rather than the concurrent image generation. Furthermore, the negative appraisals of the images indicate the model's failure to safely strengthen emotional outcomes. Yet it also demonstrates the potential of the technology, in that the generated images leave a strong impression on the writer. Instead of deploying StoryWriter as a writing therapy tool, future studies may consider how such real-time visual feedback can be used by writers – especially those who may have difficulty with imagination – to help visualize their scenes and scenarios in real-time.

Though our initial goal was to develop an application that generates images from users' text input in order to facilitate emotion regulation, we find that StoryWriter has only a limited effect in alleviating negative emotions during expressive writing exercises, though it could enhance engagement in ways that support therapeutic outcomes. Our study also demonstrates how quantitative and qualitative approaches can be combined for a thorough analysis of user evaluations and feedback to new technologies.

Given our theoretical approach, does the limited effect of StoryWriter imply that positive distraction was ineffective in facilitating emotion regulation? Qualitative evidence from Study 2 suggests that for a majority of users who experienced an upregulation in their affective state, these were attributed to the writing process, instead of the generated image. One possibility could be that, for some individuals, the writing task (imaginative fictional writing) could be sufficiently distracting from their existing negative emotional state, thus facilitating positive emotional change. This would be a departure from standard autobiographical writing therapies (see Background), in providing a relatively safer environment for online writing therapies through fictional narratives. Accordingly, our image generation should facilitate, rather than hinder or draw attention away from the writing process, which we discuss in more detail in the following section.

### 4.6.1 Limitations

There were several limitations present in our studies which should be noted. Firstly, StoryWriter was designed for English-speaking users, relying on a GAN image generator trained on English language annotations. Our study also recruited (predominantly)

American users with mostly native English language proficiency. As such, we have yet to evaluate the application in cultures and languages outside of the English-speaking U.S. Because StoryWriter's textual input is fictional and anonymous, not demanding the disclosure of personal, traumatic incidents (as in conventional expressive-writing-based therapies), we expect that the tool will be particularly accessible to individuals from East-Asian countries, who are often uncomfortable with self-disclosure [88, 102]. More research is recommended to appraise the tool's efficacy for emotion regulation outside of Anglo-American contexts.

One large limitation would be that StoryWriter currently utilises the DM-GAN model trained on photographs. This resulted in the generated images giving (in some cases) a '*grotesque*' form of reality, which negatively impact some users and drew attention away from the writing task. Given that some users were also disgruntled by the images' irrelevance for their story, one solution would be to redesign the user interface such that the generated images are given less prominence than the writing task. By placing less importance on the generated images, we hope that it could play a supporting role in guiding users to continue with their development of the story, instead of discouraging users to write due to the incongruence of the image. Also, we propose that training an image general model using newer image generation technologies (such as Diffusion Models) that are trained on considerably larger datasets would reduce the abstractness of the generated images while producing more accurate representations of the text. Alternatively, this could also be achieved through image retrieval technologies, where images are retrieved from a database of pictures (rather than generated from scratch). This would allow for greater control over the type of images displayed, and image modification technologies (such as style transfer) could also be applied.

Given prevailing circumstances and logistical constraints, we completed our qualitative evaluation remotely through a long-form, open-ended survey. However, in-person user tests may have yielded richer, more detailed information, while psychiatric supervision would have allowed the researchers to explore participants' negative emotions in greater depth (in ways that cannot be safely replicated in an online, remote survey). Future investigations may also benefit from a longitudinal, user-diary-based approach, allowing participants to use the application over a number of weeks and to record their impressions in a free-form way alongside brief, periodic interviews. Additionally, future work should explore other factors that can enhance or disrupt writing therapy, such as the impact of different image styles, the presence of music during the writing process, or the location where the writing takes place.

# 5 Conclusion

In this chapter, we summarize the results of the previous studies, discuss the limitations and propose future research directions.

## 5.1 Summary

In Chapter 3, we introduce Visualyre, a tool designed to help musicians create album art by combining Generative Adversarial Networks (GANs) and Style Transfer to create unique visuals from the lyrics and mood of the song. The system takes lyrics and audio files as input, generates images from the lyrics, analyzes the mood of the audio, and then applies style transfer to the images based on the detected mood.

We evaluated the effectiveness of the application with a user study involving 35 amateur and independent musicians. Results showed that the application was intuitive and easy to use and that the generated images were suitable as cover art. Participants found the application novel and expressed a willingness to use and recommend it to others. Some limitations were noted, such as the limited range of supported moods and the low resolution of the downloadable images.

In Chapter 4, we introduce StoryWriter, a writing tool that generates images in real-time based on users' narratives. This tool explores the potential of using AI-powered text-to-image generation to help regulate negative emotions during expressive writing.

We conducted two studies to evaluate the efficacy of StoryWriter. The first study was a quantitative experiment with 388 participants designed to measure the emotional impact of the application, where participants were induced to feel either negative or neutral emotions and then used either StoryWriter or a control group application (without image generation). The findings revealed that participants who used StoryWriter and were induced to feel negative emotions experienced a greater reduction in anger and sadness compared to those in the control group. However, there was no significant difference in anxiety or stress reduction. The second one was a qualitative user study with 54 participants to examine the user experience of StoryWriter. The study found that while the application did distract participants from their negative emotions, the distraction was not always positive. Some participants found the generated images unsettling or irrelevant to their stories. Additionally, the emotional benefits were primarily attributed to the writing process itself rather than the image generation feature.

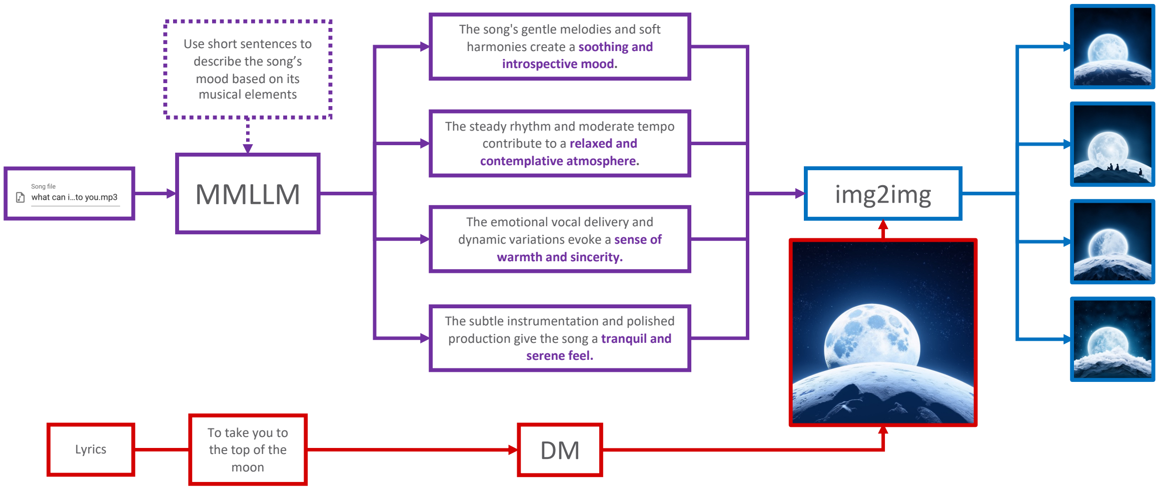
## 5.2 Limitations and Future Work

This dissertation has several limitations. We explain each of them below and propose future work to address them.

The text-to-image generation model used in both studies uses the MS-COCO dataset for training. This dataset contains around 120k images with five captions per image, considerably smaller than the most recent datasets used to train the latest text-based image generation models. In Chapter 3, this limitation was evident in the feedback from musicians who found that the generated images did not always match the themes of their lyrics. In Chapter 4, participants found the generated images to be “grotesque and unsettling”, “creepy”, or “chaotic”. These adverse reactions highlight the need for more refined image generation models trained with large datasets that allow the synthesis of images that can better cater to users’ diverse and nuanced needs in creative and therapeutic applications. The latest advancements in image generation, such as diffusion models, allow a style to be included as part of the prompt. Incorporating style prompts could enhance the therapeutic experience by allowing users to generate images that better align with their personal preferences and the specific emotions they are trying to process. This could lead to a stronger connection with the generated images, provide consistency in the visuals of their stories, and potentially improve the effectiveness of the writing therapy.

The mood detection models used in Chapter 3 were limited to identifying only four basic emotions: anger, happiness, sadness, and relaxation. This limited range may not fully capture the nuances and complexities of emotions conveyed in music, potentially leading to a mismatch between the generated artwork and the intended emotional expression of the artist. Additionally, the style transfer strategy has limitations compared to img2img in diffusion models; it relies on an initial dataset and offers less control over style selection and content preservation, limiting its versatility and potential applications. Future research could explore the use of Multi-Modal Large Language Models (MMLLMs) to extract and the application of img2img to visually represent a wider range of emotions and themes from the audio in a zero-shot setting. Figure 21 shows a proposal for this architecture; it uses a Multi-Modal Large Language Model to extract textual representations of the moods from the audio file, which are then applied to a base image using img2img. The base image can be generated from the lyrics using a Diffusion Model (DM) or provided by the user.

The study presented in Chapter 4 only examined the immediate effects of using text-to-image generation tools on emotional regulation. The long-term impact of these tools, particularly their potential to influence emotional regulation strategies over extended periods, remains unclear. Further research could address this limitation by conducting



Use short sentences to describe the song's mood based on its musical elements

The song's gentle melodies and soft harmonies create a **soothing and introspective mood.**

The steady rhythm and moderate tempo contribute to a **relaxed and contemplative atmosphere.**

The emotional vocal delivery and dynamic variations evoke a **sense of warmth and sincerity.**

The subtle instrumentation and polished production give the song a **tranquil and serene feel.**

To take you to the top of the moon

Figure 21: Proposed Architecture that makes use of Diffusion Models (DMs) and Multi-Modal Large Language Models (MMLLMs)

longitudinal studies that track users' emotional responses and regulation patterns over weeks or months of using these tools, providing valuable insights into the sustainability of the observed effects and the potential for long-term benefits in emotional well-being.

Finally, both studies were conducted exclusively in English, limiting the generalizability of the findings to other languages and cultural contexts. Further research should include participants from different cultures to investigate how users from diverse linguistic and cultural backgrounds interact with and respond to AI-generated content in creative and therapeutic applications.

# Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Eiji Aramaki, Associate Professor Shoko Wakamiya, Assistant Professor Shuntaro Yada in Social Computing Laboratory and Wan Jou (Lavender) She (currently in Kyoto Institute and Technology), Kongmeng Liew (currently in University of Canterbury, New Zealand) and Lis Pereira (currently in Center for Information and Neural Networks) for all the guidance throughout my time in NAIST as doctoral student.

The staff from the Japan Embassy in Venezuela and the Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT) for granting me the opportunity to study abroad in Japan at Nara Institute of Science and Technology. Atsuko and Yuuya, whom I had the opportunity to meet thanks to the Host Family Program offered by Osaka University.

My parents and my family for giving me all their unconditional love and support, even though we are far, far away, and I still feel that they are by my side. Finally, I want to thank all my friends and labmates for the good memories and the fun events we have shared together during my doctoral course.

# References

[1] C. Little, M. Elliot, R. Allmendinger, and S. S. Samani, “Generative adversarial networks for synthetic data generation: a comparative study,” *arXiv preprint arXiv:2112.01925*, 2021.

[2] M. Zhu, P. Pan, W. Chen, and Y. Yang, “Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5802–5810, 2019.

[3] Kainat, “Image generation: Diffusion model,” 2023.

[4] I. Aristimuño, “An introduction to diffusion models and stable diffusion,” Nov 2023.

[5] K. Wang and X. Wan, “Sentigan: Generating sentimental texts via mixture adversarial networks,” in *Proceedings of International Joint Conference on Computational Intelligence*, pp. 4446–4452, 2018.

[6] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, “Attngan: Fine-grained text to image generation with attentional generative adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1316–1324, 2018.

[7] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, “Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’14, (Cambridge, MA, USA), p. 2672–2680, MIT Press, 2014.

[9] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, and J. Shlens, “Exploring the structure of a real-time, arbitrary neural artistic stylization network,” *arXiv preprint arXiv:1705.06830*, 2017.

[10] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico,*

*May 2-4, 2016, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2016.

[11] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

[12] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 5907–5915, 2017.

[13] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan++: Realistic image synthesis with stacked generative adversarial networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1947–1962, 2018.

[14] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1316–1324, 2018.

[15] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12873–12883, 2021.

[16] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning* (M. Meila and T. Zhang, eds.), vol. 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, PMLR, 18–24 Jul 2021.

[17] K. Crowson, S. Biderman, D. Kornis, D. Stander, E. Hallahan, L. Castricato, and E. Raff, "Vqgan-clip: Open domain image generation and editing with natural language guidance," in *Proceedings of the European Conference on Computer Vision*, pp. 88–105, Springer, 2022.

[18] A. Borji, "Pros and cons of gan evaluation measures," *Computer vision and image understanding*, vol. 179, pp. 41–65, 2019.

[19] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[20] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741*, 2021.

[21] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *Proceedings of the International Conference on Machine Learning*, pp. 8821–8831, Pmlr, 2021.

[22] 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.

[23] J. Betker, G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee, Y. Guo, *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.

[24] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

[25] S. James, *The Evolution of Album Art Through Decades of Music.* PhD thesis, Radford University, 2014.

[26] R. J. Belton, "The narrative potential of album covers," *Studies in Visual Arts and Communication: an international journal*, pp. 1–7, 2015.

[27] S. Mühlbach and P. Arora, "Behind the music: How labor changed for musicians through the subscription economy," *First Monday*, vol. 25, Mar. 2020.

[28] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *CoRR*, vol. abs/1508.06576, 2015.

[29] C. Laurier and P. Herrera, "Mood cloud: A real-time music mood visualization tool," in *Proceedings of the Computer Music Modeling and Retrieval*, 2008.

[30] A. Husain, M. F. Shiratuddin, and W. W. Kok, "Establishing a framework for visualizing music mood using visual texture," in *Proceedings of the International Conference on Computing and Informatics*, 2015.

[31] S. Funasawa, H. Ishizaki, K. Hoashi, Y. Takishima, and J. Katto, “Automated music slideshow generation using web images based on lyrics,” in *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 63–68, 2010.

[32] A. Hepburn, R. McConville, and R. Santos-Rodrıguez, “Album cover generation from genre tags,” in *Proceedings of the International Workshop on Machine Learning and Music*, 2017.

[33] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier gans,” in *Proceedings of the International Conference on Machine Learning*, pp. 2642–2651, PMLR, 2017.

[34] S. Hong, D. Yang, J. Choi, and H. Lee, “Inferring semantic layout for hierarchical text-to-image synthesis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7986–7994, 2018.

[35] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Proceedings of the European Conference on Computer Vision*, pp. 740–755, Springer, 2014.

[36] C. Laurier, O. Meyers, J. Serra, M. Blech, and P. Herrera, “Music mood annotator design and integration,” in *Proceedings of the International Workshop on Content-Based Multimedia Indexing*, pp. 156–161, IEEE, 2009.

[37] P. Alonso-Jiménez, D. Bogdanov, J. Pons, and X. Serra, “Tensorflow audio models in essentia,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 266–270, IEEE, 2020.

[38] X. Xie, F. Tian, and H. S. Seah, “Feature guided texture synthesis (fgts) for artistic style transfer,” in *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, pp. 44–49, 2007.

[39] C. Li and M. Wand, “Precomputed real-time texture synthesis with markovian generative adversarial networks,” in *Proceedings of the European Conference on Computer Vision*, pp. 702–716, Springer, 2016.

[40] D. M. Oppenheimer, T. Meyvis, and N. Davidenko, “Instructional manipulation checks: Detecting satisficing to increase statistical power,” *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 867–872, 2009.

[41] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[42] M. K. Slifka and J. L. Whitton, “Clinical implications of dysregulated cytokine production,” *J. Mol. Med.*, vol. 78, pp. 74–80, 2000.

[43] J. D. Sexton, J. W. Pennebaker, C. G. Holzmueller, A. W. Wu, S. M. Berenholtz, S. M. Swoboda, P. J. Pronovost, and J. B. Sexton, “Care for the caregiver: benefits of expressive writing for nurses in the united states,” *Progress in Palliative Care*, vol. 17, no. 6, pp. 307–312, 2009.

[44] A. Tonarelli, C. Cosentino, D. Artioli, S. Borciani, E. Camurri, B. Colombo, A. D’Errico, L. Lelli, L. Lodini, and G. Artioli, “Expressive writing. a tool to help health workers. research project on the benefits of expressive writing,” *Acta Bio Medica: Atenei Parmensis*, vol. 88, no. Suppl 5, p. 13, 2017.

[45] S. F. Allen, M. A. Wetherell, and M. A. Smith, “Online writing about positive life experiences reduces depression and perceived stress reactivity in socially inhibited individuals,” *Psychiatry research*, vol. 284, p. 112697, 2020.

[46] J. W. Pennebaker and S. K. Beall, “Confronting a traumatic event: toward an understanding of inhibition and disease.,” *Journal of abnormal psychology*, vol. 95, no. 3, p. 274, 1986.

[47] J. W. Pennebaker, J. K. Kiecolt-Glaser, and R. Glaser, “Disclosure of traumas and immune function: health implications for psychotherapy.,” *Journal of consulting and clinical psychology*, vol. 56, no. 2, p. 239, 1988.

[48] J. Pizarro, “The efficacy of art and writing therapy: Increasing positive mental health outcomes and participant retention after exposure to traumatic experience,” *Art Therapy*, vol. 21, no. 1, pp. 5–12, 2004.

[49] W.-J. She, P. Siriaraya, C. S. Ang, and H. G. Prigerson, “Living memory home: Understanding continuing bond in the digital age through backstage grieving,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021.

[50] E. B. Foa, E. Hembree, and B. Rothbaum, *Prolonged Exposure Therapy for PTSDEmotional Processing of Traumatic Experiences, Therapist Guide: Emotional Processing of Traumatic Experiences, Therapist Guide.* New York, NY: Oxford University Press, 01 2015.

[51] J. Powell and A. Clarke, "The www of the world wide web: who, what, and why?," *Journal of Medical Internet Research*, vol. 4, no. 1, p. e4, 2002.

[52] M. Reinhold, P.-C. Bürkner, and H. Holling, "Effects of expressive writing on depressive symptoms—a meta-analysis," *Clinical Psychology: Science and Practice*, vol. 25, no. 1, p. e12224, 2018.

[53] V. Paquin, J. Bick, R. Lipschutz, G. Elgbeili, D. P. Laplante, B. Biekman, A. Brunet, S. King, and D. Olson, "Unexpected effects of expressive writing on post-disaster distress in the hurricane harvey study: a randomized controlled trial in perinatal women," *Psychological medicine*, pp. 1–9, 2021.

[54] K. M. Krpan, E. Kross, M. G. Berman, P. J. Deldin, M. K. Askren, and J. Jonides, "An everyday activity as a treatment for depression: the benefits of expressive writing for people diagnosed with major depressive disorder," *Journal of affective disorders*, vol. 150, no. 3, pp. 1148–1151, 2013.

[55] J. M. Smyth, "Written emotional expression: effect sizes, outcome types, and moderating variables.," *Journal of consulting and clinical psychology*, vol. 66, no. 1, p. 174, 1998.

[56] J. W. Pennebaker, *Opening up: The healing power of confiding in others.* William Morrow & Company, 1990.

[57] P. C. Trask and S. T. Sigmon, "Ruminating and distracting: The effects of sequential tasks on depressed mood," *Cognitive Therapy and Research*, vol. 23, no. 3, pp. 231–246, 1999.

[58] S. Nolen-Hoeksema and J. Morrow, "Effects of rumination and distraction on naturally occurring depressed mood," *Cognition & emotion*, vol. 7, no. 6, pp. 561–570, 1993.

[59] C. E. Waugh, E. Z. Shing, and R. M. Furr, "Not all disengagement coping strategies are created equal: positive distraction, but not avoidance, can be an adaptive coping strategy for chronic life stressors," *Anxiety, Stress, & Coping*, vol. 33, no. 5, pp. 511–529, 2020.

[60] J. Joormann, M. Siemer, and I. H. Gotlib, "Mood regulation in depression: Differential effects of distraction and recall of happy memories on sad mood," *Journal of abnormal psychology*, vol. 116, no. 3, p. 484, 2007.

[61] M. M. Shepley, "The role of positive distraction in neonatal intensive care unit settings," *Journal of Perinatology*, vol. 26, no. 3, pp. S34–S37, 2006.

[62] S. Jiang, "Positive distractions and play in the public spaces of pediatric healthcare environments: A literature review," *HERD: Health Environments Research & Design Journal*, vol. 13, no. 3, pp. 171–197, 2020.

[63] A. Barak, L. Hen, M. Boniel-Nissim, and N. Shapira, "A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions," *Journal of Technology in Human services*, vol. 26, no. 2-4, pp. 109–160, 2008.

[64] T. Hanley and D. Reynolds, "Counselling psychology and the internet: A review of the quantitative research into online outcomes and alliances within text-based therapy," *Counselling Psychology Review*, vol. 24, no. 2, pp. 4–13, 2009.

[65] E. Kaltenthaler, G. Parry, and C. Beverley, "Computerized cognitive behaviour therapy: A systematic review," *Behavioural and Cognitive Psychotherapy*, vol. 32, no. 1, pp. 31–55, 2004.

[66] A. Lange, J. Q. v. d. Ven, B. Schrieken, B. Bredeweg, and P. Emmelkamp, "Internet-mediated, protocol-driven treatment of psychological dysfunction," *Journal of Telemedicine and Telecare*, vol. 6, no. 1, pp. 15–21, 2000.

[67] N. A. Sayer, S. Noorbaloochi, P. A. Frazier, J. W. Pennebaker, R. J. Orazem, P. P. Schnurr, M. Murdoch, K. F. Carlson, A. Gravely, and B. T. Litz, "Randomized controlled trial of online expressive writing to address readjustment difficulties among us afghanistan and iraq war veterans," *Journal of Traumatic Stress*, vol. 28, no. 5, pp. 381–390, 2015.

[68] S. W. Lee, I. Kim, J. Yoo, S. Park, B. Jeong, and M. Cha, "Insights from an expressive writing intervention on facebook to help alleviate depressive symptoms," *Computers in Human Behavior*, vol. 62, pp. 613–619, 2016.

[69] K. A. Baikie, L. Geerligs, and K. Wilhelm, "Expressive writing and positive writing for participants with mood disorders: An online randomized controlled trial," *Journal of affective disorders*, vol. 136, no. 3, pp. 310–319, 2012.

[70] M. Hirai, S. T. Skidmore, G. A. Clum, and S. Dolma, "An investigation of the efficacy of online expressive writing for trauma-related psychological distress in hispanic individuals," *Behavior therapy*, vol. 43, no. 4, pp. 812–824, 2012.

[71] W. J. She, L. Burke, R. A. Neimyer, K. Roberts, W. Lichtenthal, J. Hu, and M. Rauterberg, “Toward the development of a monitoring and feedback system for predicting poor adjustment to grief,” in *Proceedings of the Conference on Design and Semantics of Form and Movement-Sense and Sensitivity, DeSForM 2017*, IntechOpen, 2017.

[72] E. Morton, ““it’s not just a bunch of scores”–using the quality of life tool to self-monitor wellness in bipolar disorder,” 2020.

[73] A. H. Bettis, T. A. Burke, J. Nesi, and R. T. Liu, “Digital technologies for emotion-regulation assessment and intervention: A conceptual review,” *Clinical Psychological Science*, p. 21677026211011982, 2021.

[74] G. Wadley, W. Smith, P. Koval, and J. J. Gross, “Digital emotion regulation,” *Current Directions in Psychological Science*, vol. 29, no. 4, pp. 412–418, 2020.

[75] J. C. Torrado, J. Gomez, and G. Montoro, “Emotional self-regulation of individuals with autism spectrum disorders: Smartwatches for monitoring and interaction,” *Sensors*, vol. 17, no. 6, p. 1359, 2017.

[76] R. T. Azevedo, N. Bennett, A. Bilicki, J. Hooper, F. Markopoulou, and M. Tsakiris, “The calming effect of a new wearable device during the anticipation of public speech,” *Scientific reports*, vol. 7, no. 1, pp. 1–7, 2017.

[77] J. Costa, F. Guimbretière, M. F. Jung, and T. Choudhury, “Boostmeup: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch,” in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 1–23, ACM New York, NY, USA, 2019.

[78] J. W. Newbold, J. Luton, A. L. Cox, and S. J. J. Gould, “Using nature-based soundscapes to support task performance and mood,” in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’17, (New York, NY, USA), p. 2802–2809, Association for Computing Machinery, 2017.

[79] B. Lim, Y. Rogers, and N. Sebire, “Designing to distract: Can interactive technologies reduce visitor anxiety in a children’s hospital setting?,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 26, no. 2, pp. 1–19, 2019.

[80] M. A. Dabrowska, *The Role of Positive Distraction in the Patient’s Experience in Healthcare Setting: A Literature Review of the Impacts of Representation of Na-*

*ture, Sound, Visual Art, and Light*. PhD thesis, Georgia Institute of Technology, 2020.

[81] K. Ćosić, S. Popović, M. Horvat, D. Kukolja, B. Dropuljić, B. Kovač, and M. Jakovljević, "Computer-aided psychotherapy based on multimodal elicitation, estimation and regulation of emotion," *Psychiatria Danubina*, vol. 25, no. 3, pp. 340–346, 2013.

[82] J. F. Cohn, "Foundations of human computing: facial expression and emotion," in *Proceedings of the 8th international conference on Multimodal interfaces*, pp. 233–238, 2006.

[83] A. Fernández-Caballero, A. Martínez-Rodrigo, J. M. Pastor, J. C. Castillo, E. Lozano-Monasor, M. T. López, R. Zangróniz, J. M. Latorre, and A. Fernández-Sotos, "Smart environment architecture for emotion detection and regulation," *Journal of biomedical informatics*, vol. 64, pp. 55–73, 2016.

[84] A. Fernández-Caballero, J. M. Latorre, J. M. Pastor, and A. Fernández-Sotos, "Improvement of the elderly quality of life and care through smart emotion regulation," in *Proceedings of the International Workshop on Ambient Assisted Living*, pp. 348–355, Springer, 2014.

[85] I. Benke, M. T. Knierim, and A. Maedche, "Chatbot-based emotion management for distributed teams: A participatory design study," in *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, pp. 1–30, ACM New York, NY, USA, 2020.

[86] A. P. Henkel, S. Bromuri, D. Iren, and V. Urovi, "Half human, half machine–augmenting service employees with ai for interpersonal emotion regulation," *Journal of Service Management*, 2020.

[87] N. Rajcic and J. McCormack, "Mirror ritual: An affective interface for emotional self-reflection," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[88] M. Manabe, K. Liew, S. Yada, S. Wakamiya, E. Aramaki, *et al.*, "Estimation of psychological distress in japanese youth through narrative writing: Text-based stylometric and sentiment analyses," *JMIR Formative Research*, vol. 5, no. 8, p. e29500, 2021.

[89] E. Siedlecka and T. F. Denson, "Experimental methods for inducing basic emotions: A qualitative review," *Emotion Review*, vol. 11, no. 1, pp. 87–97, 2019.

[90] E. Diener, R. A. Emmons, R. J. Larsen, and S. Griffin, "The satisfaction with life scale," *Journal of personality assessment*, vol. 49, no. 1, pp. 71–75, 1985.

[91] P. Lovibond and S. Lovibond, "The structure of negative emotional states: Comparison of the depression anxiety stress scales (dass) with the beck depression and anxiety inventories," *Behaviour Research and Therapy*, vol. 33, no. 3, pp. 335–343, 1995.

[92] L. Parkitny and J. McAuley, "The depression anxiety stress scale (dass)," *J Physiother*, vol. 56, no. 3, p. 204, 2010.

[93] S. Kreitler and H. Casakin, "Self-perceived creativity: The perspective of design," *European Journal of Psychological Assessment*, vol. 25, no. 3, pp. 194–203, 2009.

[94] The jamovi project, "jamovi," 2021.

[95] J. Brooks, S. McCluskey, E. Turley, and N. King, "The utility of template analysis in qualitative psychology research," *Qualitative Research in Psychology*, vol. 12, pp. 202–222, 2015.

[96] A. Madill, A. Jordan, and C. Shirley, "Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies," *British Journal of Psychology*, vol. 91, pp. 1–20, 2000.

[97] C. Carver and E. Harmon-Jones, "Anger is an approach-related affect: evidence and implications," *Psychological bulletin*, vol. 135 2, pp. 183–204, 2009.

[98] C. Carver, "Approach, avoidance, and the self-regulation of affect and action," *Motivation and Emotion*, vol. 30, pp. 105–110, June 2006.

[99] A. A. Labroo and D. D. Rucker, "The orientation-matching hypothesis: An emotion-specificity approach to affect regulation," *Journal of Marketing Research*, vol. 47, no. 5, pp. 955–966, 2010.

[100] I. Elkin, L. Falconnier, Y. Smith, K. E. Canada, E. Henderson, E. R. Brown, and B. M. Mckay, "Therapist responsiveness and patient engagement in therapy," *Psychotherapy Research*, vol. 24, no. 1, pp. 52–66, 2014.

[101] D. Sabo Mordechay, Z. Eviatar, and B. Nir, "Emotional engagement in expressive writing: Clinical and discursive perspectives," *Narrative Inquiry*, 2021.

[102] A. Asai and D. C. Barnlund, "Boundaries of the unconscious, private, and public self in japanese and americans: a cross-cultural comparison," *International Journal of Intercultural Relations*, vol. 22, no. 4, pp. 431–452, 1998.

# List of Publications

## Journals

1. Gamar Azuaje, Kongmeng Liew, Rebecca Buening, Wan Jou She, Panote Siriaraya, Shoko Wakamiya, Eiji Aramaki. “Exploring the use of AI Text-to-Image Generation to Downregulate Negative Emotions in an Expressive Writing Application”, Royal Society Open Science, 9, 220238, Apr. 2023

2. Gamar Azuaje, Kongmeng Liew, Elena Epure, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki. “Visualyre: multimodal album art generation for independent musicians”, Personal and Ubiquitous Computing, May 2023

## International Conferences

1. M. E. Aburto-Gutierrez, G. Azuaje, V. Mishra, S. Osmani and K. Ikeda. “JaSenpai: Towards an Adaptive and Social Interactive E-Learning Platform for Japanese Language Learning”, 2022 International Conference on Advanced Learning Technologies (ICALT), Bucharest, Romania, 2022. (pp. 236-238).