Journal of the Brazilian Computer Society, 2025, 31:1, doi: 10.5753/jbcs.2023.4166

From Pampas to Pixels: Fine-Tuning Diffusion Models for Gaúcho Heritage

William Alberto Cruz-Castañeda [**Alana AI Research** | ***williamalberto.cruz@gmail.com***]
Marcellus Amadeus [**Alana AI Research** | ***7marcellus@gmail.com***]
André Felipe Zanella [**Alana AI Research** | ***aft.zanella@gmail.com***]
Felipe Rodrigues Perche Mahlow [**Alana AI Research** | ***felipe.mahlow@gmail.com***]

Alana AI, Av. Rebouças, 1585 - Pinheiros, São Paulo - SP, 05401-200, Brazil.

Received: 26 February 2024 • **Accepted:** 06 February 2025 • **Published:** 14 April 2025

Abstract Generative Artificial Intelligence has become pervasive in society, witnessing significant advancements in various domains. Particularly in the domain of Text-to-Image (TTI) models, Latent Diffusion Models (LDMs) showcase remarkable capabilities in generating visual content based on textual prompts. This paper addresses the potential of LDMs in representing local cultural concepts, historical figures, and endangered species. In this study, we use the cultural heritage of Rio Grande do Sul (RS), Brazil, as an illustrative case. Our objective is to contribute to the broader understanding of how generative models can help to capture and preserve the regional culture and historical identity. The article outlines the methodology, including subject selection, dataset creation, and fine-tuning process. The results showcase the picture generation alongside the challenges and feasibility of each concept. In conclusion, this work shows the power of these models to represent and preserve unique aspects of diverse regions and communities.

Keywords: Diffusion Models, Text-to-image Generation, Generative Artificial Intelligence, Computer Vision

1 Introduction

In recent years, the domain of computer science and artificial intelligence (AI) has witnessed significant advancements in the discipline of Generative AI (GenAI). GenAI focuses on creating models capable of generating text, audio, and images based on patterns learned from training datasets. Generative models like GPT-4 [OpenAI, 2023] have demonstrated remarkable abilities to produce coherent and informative text, enabling applications in chatbots, machine translation, and high-quality content creation. In the audio field, notable generative models such as WaveGAN [Donahue *et al.*, 2019] have gained prominence. Furthermore, recent advances in text-to-speech (TTS) models have contributed to the evolving landscape of audio synthesis [Kaur and Singh, 2023; Lobato *et al.*, 2023]. Within the field of image generation, groundbreaking models such as DALL·E 3 [Betker *et al.*, 2023; Ramesh *et al.*, 2021], MidJourney[1] and Stable Diffusion [Rombach *et al.*, 2022] are capable of generating images from textual descriptions. Applications of these models are in graphic design [Hughes *et al.*, 2021], augmented reality [Liu *et al.*, 2020], and more, underscoring the significant impact GenAI has on creativity and innovation across diverse fields.

Text-to-image (TTI) models refer to a research and development area aimed at creating methods and algorithms capable of converting written text into visual images. This approach seeks to establish a bridge between natural language and visual representations, exploring ways to transform textual descriptions into coherent and meaningful visual content. An advancement in this domain lies in the emergence and adoption of Latent Diffusion Models (LDMs) [Rombach *et al.*, 2022]. These models have become essential in image creation, especially those of artistic and cultural relevance. Rooted in the foundational framework of Diffusion Probabilistic Models (DPMs) [Ho *et al.*, 2020], LDMs have wrought a transformative impact on the panorama of image generation, facilitating the generation of high-resolution images while achieving exceptional diversity and photorealism. LDMs have been increasingly utilized in various domains, such as in the medical field for generating brain images [Pinaya *et al.*, 2022], reconstructing images from brain activity [Takagi and Nishimoto, 2023], and chest X-rays [Packhäuser *et al.*, 2023; Weber *et al.*, 2023]. They have also been used for video generation [Zhou *et al.*, 2023; He *et al.*, 2023], and more broadly, are being proposed as strategies in architecture [Yıldırım, 2022; Ploennigs and Berger, 2023], news illustration [Liu *et al.*, 2022], and in nursing education [Reed *et al.*, 2023].

This work focuses on three aspects. First, assess TTI models as tools for representing regional concepts. Second, address the representation of historical personages and events with scarce visual depictions. Third, exploring the feasibility of depicting animals and plants at risk. To achieve this, we exemplify the application of TTI in the cultural heritage of Rio Grande do Sul (RS), Brazil. This regional culture, known for its historical, traditional, and visually unique attributes, embodies a rich repository of cultural symbols, costumes, breathtaking landscapes, and pivotal historical figures that significantly shape the cultural identity of southern Brazil. Our work seeks to demonstrate the feasibility of employing diffusion models to generate images that authentically represent the cultural and historical value of Rio Grande

[1] https://www.midjourney.com

do Sul. We present the image generation process alongside the results and a discussion of challenges regarding each concept.

The work is organized as follows: Section 2 describes the methodology to conduct the study, including the definition of training subjects, datasets, and training process; Section 3 summarizes the details of the experimental process carried out in the case study to validate the proposed methodology. In Results Section 4 presents the practical application of these techniques in three key areas: Cultural Aspects, Fauna and Flora, and the Farroupilha Revolution, describing the importance of each concept and the challenges faced by each one. Finally, in Section 5 conclusions and perspectives for future improvements are presented.

2 Method

Recent advances in TTI models allow fine-tuning of pre-trained models to inject a specific subject into the output domain of the model, i.e., the model can generate diverse instances of the same subject. This section describes subject-driven models using existing pre-trained models and open-source libraries.

2.1 Training topics

The subject selection is a pivotal step as it determines the model trend toward the type of images representing that subject. Subject-driven generation allows development in different contexts while maintaining the distinctive features of the concept. There are no restrictions on the choice of concepts. They can range from simple objects and living beings (specific individuals or animals) to complex concepts such as war battle paintings. However, they may require additional effort in subsequent steps, including image selection to compose a dataset and fine-tuning experimentation.

2.2 LDMs - Stable Diffusion

Latent Diffusion Models (LDMs) are a class of generative models that perform diffusion-based sampling in a compressed latent space rather than directly on high-dimensional pixel data. This approach significantly reduces the computational cost while preserving high-quality image generation. In LDMs, an encoder $\mathcal{E}$ is first used to project the input image x into a lower-dimensional latent space, denoted as $z = \mathcal{E}(x)$. Then, during training, a sequence of denoising autoencoders $\epsilon_\theta(x_t, t)$ learns to iteratively remove noise from a latent variable x_t, which is a noisy version of z. The goal is to predict a denoised version of the latent representation, and finally, the decoder D reconstructs the image from the denoised latent variable Rombach *et al.* [2022]; Podell *et al.* [2023].

The training process is governed by the following objective:

$$L_{LDM} := \mathbb{E}_{\mathcal{E}(x),y,\epsilon\sim N(0,1),t}\left[\|\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y))\|_2^2\right] \quad (1)$$

where $\epsilon_\theta(z_t, t, \tau_\theta(y))$ represents the denoising autoencoder, conditioned on time step t and the text representation $\tau_\theta(y)$. The process begins with a noisy latent variable z_t, obtained from the encoder $\mathcal{E}$, and gradually reduces noise to recover the underlying latent image representation z.

In our work, we employ Stable Diffusion versions 1.4[2] and 1.5[3]Rombach *et al.* [2022], which use a Variational AutoEncoder (VAE) structure for both the encoder and decoder, sharing the same principles outlined above. The models are conditioned on text through a cross-attention mechanism, where text inputs are projected into intermediate representations $\tau_\theta(y)$ to guide image generation. By working in this latent space, LDMs significantly improve the efficiency and scalability of the diffusion process, allowing us to train on higher-resolution images while maintaining reasonable computational demands.

2.3 Dataset creation

Within the contexts defined, the next step involves selecting relevant images to compose a dataset for each specific scenario. In our analysis, the amount of samples needed to cover a dataset was associated with the complexity of the concept. The first concern is to assess the image availability related to the specific subject. If there is a shortage of images, the dataset will restricted to a few available images. A web crawling algorithm can be adopted to automate the task across available image databases. Data curatorship after collection and annotation is also desirable, as it involves discarding out-of-context images and also poor resolution and low DPI.

2.4 Fine-tuning process

Fine-tuning allows the model to adapt to specific nuances and characteristics associated with individual concepts. The Diffusers [von Platen *et al.*, 2022] library offers fine-tuning scripts for training pre-trained diffusion models, enabling subject-driven generation. In addition to standard TTI training, specific techniques designed for this purpose, such as DreamBooth [Ruiz *et al.*, 2023], are available within the framework.

DreamBooth, introduced in 2022, represents one of the most widely adopted methods in TTI synthesis. This approach addresses the growing demand for personalized image generation by fine-tuning pre-trained diffusion models. By associating unique identifiers with specific subjects, DreamBooth enables the creation of photorealistic images set in various contexts, all guided by user-defined text prompts. Notably, it expands the model's language-vision dictionary to encompass user-specified subjects and leverages autogenous, class-specific prior preservation mechanisms to support the generation of diverse instances within the same class. The applications of DreamBooth extend to subject recontextualization, property modification, and creative art generation. Despite being very recent, DreamBooth has already been used to generate automotive images [Sutedy and Qomariyah, 2022], to improve the performance of CNNs [Zhang, 2023] and 3d generation [Raj *et al.*, 2023]. The

[2] https://huggingface.co/CompVis/stable-diffusion-v1-4

[3] https://huggingface.co/runwayml/stable-diffusion-v1-5

DreamBooth script is available at Diffusers[4] and can be employed for training with any diffusion-based TTI model, such as Stable Diffusion 1.5 or Stable Diffusion XL.

Fine-tuning TTI models are costly and demand powerful GPUs. To mitigate this, several strategies exist to optimize GPU RAM utilization, available at Diffusers. These techniques enable training with GPUs with less memory. Some of these optimization strategies are enumerated below:

- LoRA (Low-Rank Adaptation): LoRA [Hu *et al.*, 2021] can effectively reduce the memory requirements of deep learning models by approximating the weight matrices with low-rank factors, enabling the training of larger models with limited GPU memory.
- Gradient Accumulation: Implement gradient accumulation to split the backpropagation process into smaller batches, which can help reduce GPU memory requirements. This approach allows you to accumulate gradients over multiple forward passes before updating model parameters.
- Reduce Batch Size: If feasible, consider reducing the batch size during training. Smaller batch sizes require less GPU memory but may result in longer training times. It's important to find a balance between batch size and training time based on specific requirements.
- 8-bit Adam: using a lower bit optimizer, or *fp16* for training can significantly reduce memory requirements.

3 Experiments

In this section, we describe our case study of creating subject-driven TTI models to represent the regionalism, culture, and historical value of the state of RS, Brazil.

3.1 Subjects and Datasets

In this work, we study and evaluate the generative capabilities of Stable Diffusion related to concepts from the following categories: regional biome, historical events and personalities, cultural costumes, and traditional attire. As for the biome, the selected concepts include the “Araucária”, a distinctive tree native to the southern region, as well as the “Gato-do-mato-pequeno” and “Sapinho-admirável-de-barriga-vermelha”, which are characteristic animals of the area. The Farroupilha Revolution is a significant historical event for RS that influenced Brazilian history. We chose to represent historical figures of this conflict: Giuseppe Garibaldi, Anita Garibaldi, and Bento Gonçalves, as well as portrayals of battles, the proclamation of the Rio Grandense Republic, a milestone in the revolution, and the importance of the female figure during the conflict. Regarding customs, we depict the “Chimarrão”, a typical beverage of the southern region, and, finally, the traditional attire of RS people.

With the concepts defined, images were collected to create a dataset. The quantity and diversity of these images depend on each of the concepts. For example, there is a large number of images regarding the biomes, the “Araucaria” trees, but a limited number of images regarding the Farroupilha Revolution and its related concepts, like the person of Anita Garibaldi.

A three-stage curatorship process was undertaken to generate the datasets intended for training. In the initial stage, a web crawling algorithm was employed using the concepts as keyworks to collect images associated with the specified concepts. The second stage involved a visual curatorship to ensuring in an initial set of images with high-quality. We established exclusion criteria to compose the training data, excluding i) images with low correlation with the given concept and ii) images with low DPI and resolution. The first criterion was necessary to assess the suitability of the samples for each concept; for example, the web crawling algorithm collected some images of buildings named “Araucaria”, rather than the tree itself. Table 1 presents the final number of images in each dataset.

Concept	**Images**
Cuia e Bomba de Chimarrão	28
Indumentária Gaúcha	84
Araucárias	97
Pampas	46
Gato-do-mato-pequeno	41
Sapinho-admirável-de-barriga-vermelha	31
Anita Garibaldi	14
Bento Gonçalves	12
Giuseppe Garibaldi	22
Batalhas Farroupilha	49

Table 1. Number of training images used for fine-tuning each concept.

3.2 Training Details

A summary of the methods/parameters and hardware used can be found in the table 2. In the process of training our models, we used the DreamBooth training script, available at Diffusers. It is relevant to highlight that we fine-tuned each dataset separately, generating a specialized model that learned a concept. For each dataset, we assigned a distinct token, creating a unique training instance prompt, such as “A painting of **[V]**”.

The models were trained on the default resolution of 512×512 of SD v-1.4 and v-1.5, with a batch size of 1, using the AdamW optimizer. We experimented with different learning rate values ranging from 3×10^{-6} to 5×10^{-7}. We also employed a “constant” learning rate scheduler, ensuring a consistent rate throughout training. The number of training steps to achieve good results varies according to the dataset. In our experiments, we performed extensive testing within a range of 1.000 to 10.000 training steps. Fewer steps did not effectively capture the desired concepts, even though a substantial increase in the step count resulted in a model that became overly biased for specific concepts, limiting its ability to absorb additional prompts effectively. During inference, we used the PNDMScheduler[5] with 50 steps, 7.5 classifier-free guidance, at resolution 512×512. All the fine-tuning experiments were conducted on instances using a single NVIDIA

[4] https://github.com/huggingface/diffusers/tree/main/examples/dreambooth

[5] https://huggingface.co/docs/diffusers/api/schedulers/pndm

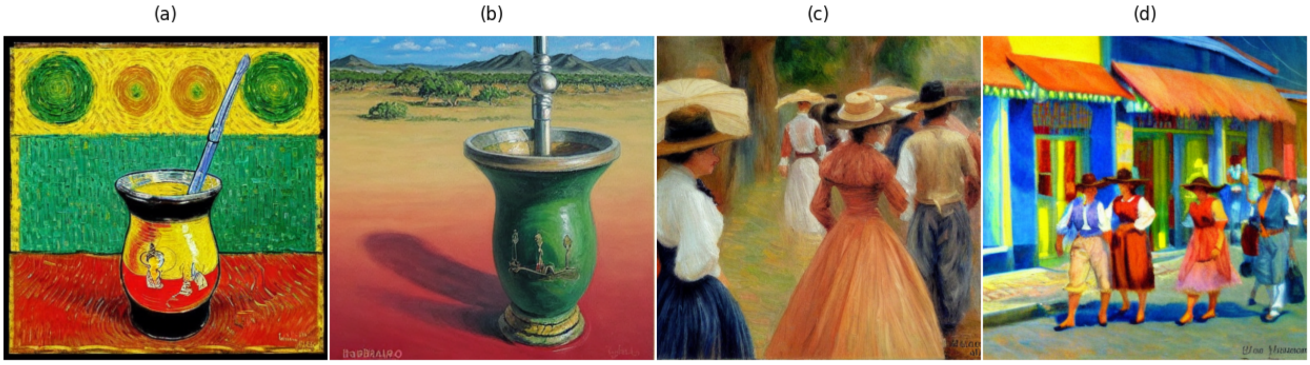

Figure 1. Images generated by Dreambooth to represent the tea and clothing culture of Rio Grande do Sul, Brazil. The prompts adopted were (a) A painting of *cuiaBombaChimarrao*, Van Gogh Style. (b) A painting of *cuiaBombaChimarrao*. (c) A painting of *indumentariaGaucha*, impressionism style. (d) A painting of *indumentariaGaucha*, impressionism style.

Configuration	**Value/Description**
Pre-trained Models	Stable Diffusion v1.4, v1.5
Training Resolution	512 × 512
Batch Size	1
Optimizer	AdamW
Learning Rate	3×10^{-6} to 5×10^{-7}
Learning Rate Scheduler	Constant
Training Steps	1,000 to 10,000
Token/Prompt Template	"A painting of **[V]**"*
Inference Scheduler	PNDMScheduler
Inference Steps	50
Classifier-Free Guidance	7.5
Inference Resolution	512 × 512
Training Hardware (VRAM)	NVIDIA A100 GPU (80 GB)
Inference Hardware (VRAM)	NVIDIA V100 GPU (16 GB)

Table 2. Summary of Parameters Used for Fine-Tuning and Image Generation.*Refers to the token used for each concept, where **[V]** is the name of the concept, as presented in Table (1).

A100 GPU with 80GB of memory. Inference was carried out on a V100 GPU instance with 16 GB of memory.

To incorporate stylistic elements into the images, the prompts were augmented with phrases such as "Van Gogh Style" or "Impressionism style". The prompts utilized are provided in the figure captions accompanying the results.

4 Results

This section discusses the results obtained from the DreamBooth fine-tuning experiments, dividing them into three main results. First, focus on cultural aspects that explore the representation of the "Cuia e Bomba de Chimarrão", essential elements of the mate tea culture, and the "Indumentária Gaúcha" (traditional gaucho costume). Additionally, we explore the representation of fauna and flora, with emphasis on the "Pampas" landscape, "Araucárias" trees, and endangered endemic animals, such as the *Melanophryniscus admirabilis* ("Sapinho-admirável-de-barriga-vermelha") and the *Leopardus tigrinus* ("Gato-do-mato-pequeno"). The narrative of the Farroupilha Revolution delves deeper into historical figures. We also explore some significant battles during this period. Our studies through these visual narratives showcases DreamBooth's ability to portray elements of Brazil's cultural, ecological, and historical events.

4.1 Cultural Aspects

In the scope of cultural aspects, we have successfully derived subject-driven models capable of representing images of some of the customs of Rio Grande do Sul. As training concepts, we choose to explore the traditional drink of the southern region of Brazil, "chimarrão" and the traditional attire of Gaúchos. The chimarrão is a mate-based drink introduced by the Guarani indigenous people, making it a deep-rooted typical drink. The "cuia" and "bomba de chimarrão", usually known as "mate straw" [Bernardes, 2021], are used for the preparation and consumption of the drink. Figure 1 (a) portrays a "Cuia de Chimarrão" in the style of Van Gogh, with the colors of the Rio Grande do Sul flag in the background, while Figure 1 (b) portrays a painting without a specified artistic style, demonstrating Dreambooth's versatility in creating diverse visual representations. Regarding the latter concept, the traditional attire, or "indumentária gaúcha" is characterized by its unique elements, which include "bombacha" (Refers to the loose-fitting, baggy trousers or pants, often made of durable fabric), "gaita" (sash or belt), a shirt, traditional leather boots, a neckerchief or bandana, wide-brimmed hat (often known as the "chimarrão" hat), "poncho" (a shoulder-worn fabric piece), and a waistband or sash. This clothing encompasses the traditional man's attire. On the other hand, the "prenda" symbolizes the representation of the rural woman, including long skirts, ornate blouses, scarves, and boots. These components collectively make up the traditional attire of gaúchos, serving both functional and cultural purposes. Figures 1 (c) and 1 (d) depict some of these elements of gaúcho attire in the impressionist style.

Regarding the "chimarrão" a substantial number of images containing a "cuia" and "bomba de chimarrão" were found to build the database. However, a significant challenge arose due to several featured logos, resulting in distorted logos in the output domain of the model. Consequently, the curatorship process had to filter out items without logos for the database. Another issue pertained to the close-up shots of many photos depicting the concept, with the object occupying a substantial portion of the image. This proximity caused the model to lack a realistic sense of the mate gourd's size. When asked the model to depict a person drinking mate, the mate gourd would appear larger than its actual size, or the model failed to understand that the person would drink through a straw, treating the mate gourd as if it were a cup. Furthermore, due to the similar shape, the model occasionally added flowers inside the mate gourd, as if it were a plant pot. After extensive experimentation with different configurations, we generated highly satisfactory images related to the concept.

In the case of "indumentária gaúcha" (traditional attire), there was also an abundance of images available on the internet. However, a significant challenge emerged as the clothing from various periods often became intertwined, blending with more contemporary adapted pieces. Additionally, certain concepts like "bombacha" and "gaita" often had to be included in separate images (not worn by individuals but displayed in showcases, for example) to enable the model to capture their specific details. Moreover, the prevalence of images featuring multiple individuals wearing the traditional attire, often from events where dozens of people wear these garments, seemed to bias the model towards generating images of groups of people in the attire rather than focusing on one or two individuals. Despite these challenges, the model captures the essence of the attire remarkably.

4.2 Fauna and Flora

In our exploration of fauna and flora, we focused on the "Pampas", extensive natural plains covering 2% of Brazilian territory. For the regional flora, we chose to represent "Araucárias", a group of evergreen trees native to South America and part of the Atlantic Forest, which has experienced severe deforestation, with only $1-4\%$ of its original area remaining [Mantovani *et al.*, 2004]. Figures 2 (a) depict "Araucárias" in the style of Tarsila do Amaral's paintings, and Figure 2 (b) shows a painting of "Pampas" landscapes. Additionally, we investigated the potential use of AI to generate images of endangered species, specifically the red-bellied toad (*Melanophryniscus admirabilis* or "Sapinho-admirável-de-barriga-vermelha") and the little spotted cat (*Leopardus tigrinus* or "Gato-do-mato-pequeno"). The first is an endemic species with its habitat limited to a 700-meter stretch along the Forqueta River in the Arvorezinha municipality of RS. The latter originates from Central and South America and is endangered due to deforestation and habitat conversion to agricultural land [Silveira, 2018]. Figure 2 (c) depicts the "Gato-do-mato-pequeno" in Claude Monet style, whereas Figure 2 (d) represents the "Sapinho-admirável-de-barriga-vermelha" rendered in the style of a child's drawing.

In the case of "Araucárias", the main challenge lay in appropriately curating the dataset to ensure that other tree species were not present in the data, which could potentially bias the model. Many photos depicted a single tree in the center of the image, which tended to skew the model's preferences towards this configuration, resulting in less successful outcomes when the representation involved denser vegetation. Overall, this was one of the concepts where the model presented an easy learning process and demonstrated greater flexibility in our study. For the "Pampas", the primary challenge arises from the vast variability of images. Nevertheless, the model grasped that the concept pertained to a landscape and its fundamental characteristics.

Regarding the "Sapinho-admirável-de-barriga-vermelha", several challenges were encountered. Images often portrayed the toad in a close perspective, causing the model to lose its sense of scale, similar to the issue with "Cuia e Bomba de Chimarrão". The model tended to confuse the color of the toad's belly/legs with the ground, often depicted as red. Furthermore, there were difficulties in accurately representing the toad's limbs and toes. This concept proved to be particularly challenging to illustrate. The main challenges for the "Gato-do-mato-pequeno" concept revolved around its similarity to other animals already known to the model, such as domestic cats, ocelots, and others. To address this, we ensured the model had sufficient images of the distinctive features of this species in the training dataset and depicted them in their natural habitat.

4.3 Farroupilha Revolution

This section delves into the Farroupilha Revolution period in Brazilian history, especially for Rio Grande do Sul. Through images generated by Dreambooth, we immerse ourselves in the lives and actions of the characters who shaped this tumultuous period. In the first subsection, we present generated images that capture pivotal moments in the lives of Anita Garibaldi, Giuseppe Garibaldi, and Bento Gonçalves. The second subsection focuses on the battles that marked the Farroupilha Revolution. We explore everything from the proclamation of the Rio Grandense Republic to the Crossing of the Lanchões, highlighting the details of these historical events. Additionally, we underscore the often underestimated role of women in the revolution, demonstrating their vital contribution to the movement.

4.3.1 Historical Characters

The Farroupilha Revolution, which unfolded between 1835 and 1845 in the southern region of Brazil, centered on Rio Grande do Sul State. It was a movement that sought substantial reforms, including the quest for autonomy to reflect the region's unique characteristics and demands [Zalla and Menegat, 2011]. Led by iconic figures such as Anita Garibaldi, Bento Gonçalves, and Giuseppe Garibaldi, this episode transcended the boundaries of southern Brazil, leaving an indelible mark on the country's history. Anita Garibaldi is depicted in Figure 3 (a) with a sunset as a backdrop, Figure 3 (b) captured by opposing forces. Bento Gonçalves, depicted in Figure 3 (c), was one of the leading military leaders of the Farroupilha Revolution. His role was instrumental in shaping the course of the rebellion, contributing significantly to the pursuit of autonomy for the southern Brazilian region. Giuseppe Garibaldi, represented in Figure 3 (d), also known as "The Hero of Two Worlds" due to his participation in the unification of Italy and central role in Rio Grandense independence, is an influential character for the southern Brazilian people [Oliveira, 2012]. Alongside Bento Gonçalves, he played a crucial role in leading the revolution and advocating for the unique characteristics and demands of the region.

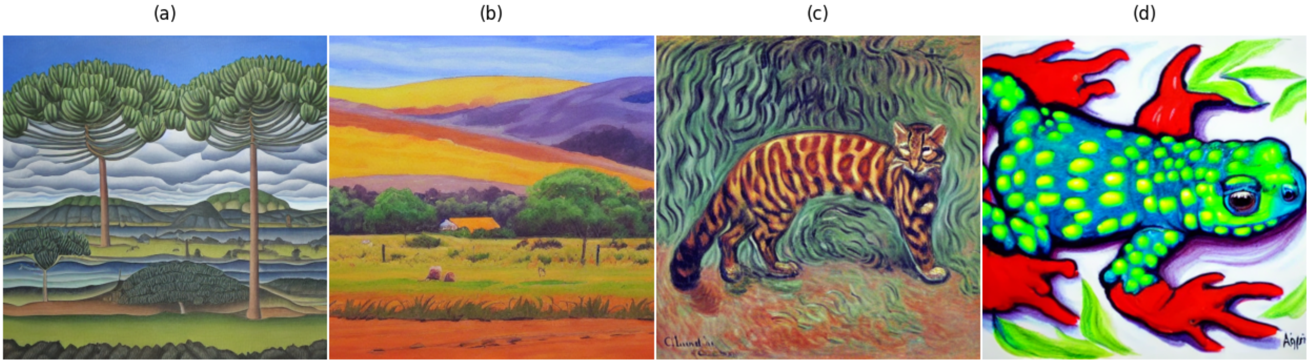

Figure 2. Images generated by Dreambooth to represent the fauna and flora of Rio Grande do Sul, Brazil. The prompts adopted were: (a) A painting of *araucaria*, Tarsila do Amaral Style. (b) A painting of *Pampa* landscape. (c) A painting of *gatoDoMato*, Claude Monet Style. (d) A painting of *sapinhoBarrigaVermelha*, kid's drawing style.

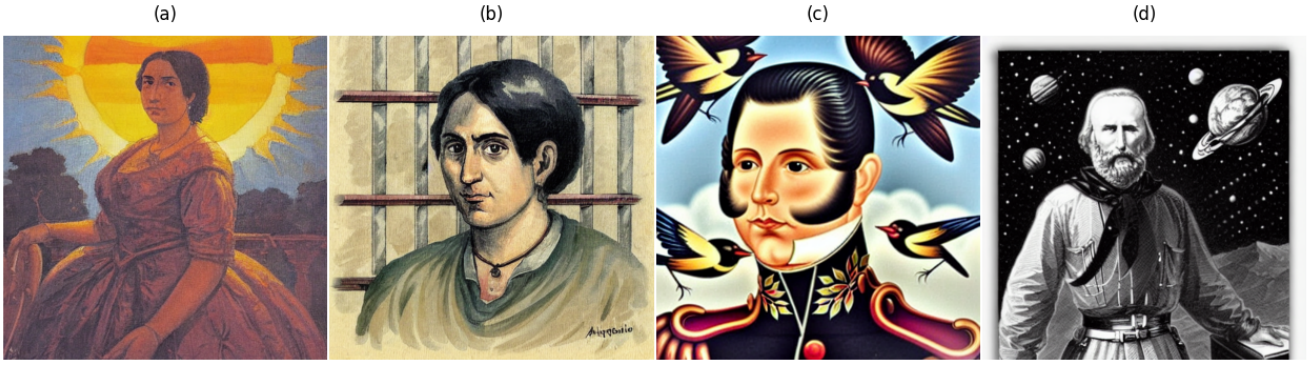

Figure 3. Dreambooth generated images to represent the Farroupilha revolution. The prompts used were: (a) A painting of *aniGaribaldi*, with a big sun behind her, (b) A painting of *aniGaribaldi*, kept in prison, (c) A painting of *bentoGonçalves*, with birds around him. (d) A painting of *giuseppeGaribaldi*, in outer space, with planets and stars around him.

Images of historical figures from the 19th century, such as Anita Garibaldi, Giuseppe Garibaldi, and Bento Gonçalves, are notably scarce in terms of variety. The existing photographs often provide limited perspectives, with some representations focusing only on the bust upwards. Furthermore, many historical images lack sufficient DPI and resolution, presenting challenges in capturing the filled essence of these figures. The limited visual records of these personalities create a scarcity of resources, particularly when considering the diverse and comprehensive representations. Despite these challenges, the model development centered on these historical subjects opens up new possibilities. These models, incorporating rare tokens like *aniGaribaldi*, *GiusepGaribaldi*, and *bentoGoncalves* within instance training prompts, demonstrate a remarkable ability to generate artistic images.

Figure 3 showcases the model's capacity to derive visually pleasing and conceptually rich images, even in scenarios with sparse historical representations. With fine-tuning configurations and inferences, the model navigates adversities posed by limited visual resources. Enable the creation of artistic depictions but also allow the integration of various elements, such as animals, into the narrative of the image. In essence, the ability to represent historical figures with scarce visual resources underscores the power of AI models in preserving and promoting historical narratives. By leveraging the potential of these models, we can bridge gaps in visual records, offering contemporary audiences a more holistic and engaging insight into the vital roles played by individuals such as those. This technological advancement can contribute significantly to preserving and celebrating the cultural heritage of nations.

4.3.2 Battles

The Farroupilha Revolution, a defining period in Brazilian history, was marked by numerous battles that played a pivotal role in the struggle for independence. These battles were the crucible in which the aspirations of the rebels clashed with the forces of the Brazilian Empire. They decided the fate of the revolution but also contributed to the shaping of Brazil's future [Sinotti *et al.*, 2015]. Some of the crucial events of the Farroupilha Revolution were the "Travessia dos Lanchões" or Crossing of the Lanchões, vessels driven by land, on wheels and pulled by cattle, and the Proclamation of the Rio Grandense Republic, a nation-state separate from the Brazilian state.

Figure 4 (a) captures the daring Crossing of the "Lanchões" under the leadership of Giuseppe Garibaldi, hauled the boats, or "lanchões," several kilometers over land, through dense terrain and across rugged landscapes, reaching the destination and crossing the Guaíba River, thereby outmaneuvering the imperial forces. Images serve as a visual testament to the audacity and resourcefulness that marked this episode in the Farroupilha Revolution. Figures 4 (b) and 4 (c) depict the iconic moment of the "Proclamation of the Republic" during the Revolution. It was a declaration that echoed the rebels' commitment to the cause of independence,

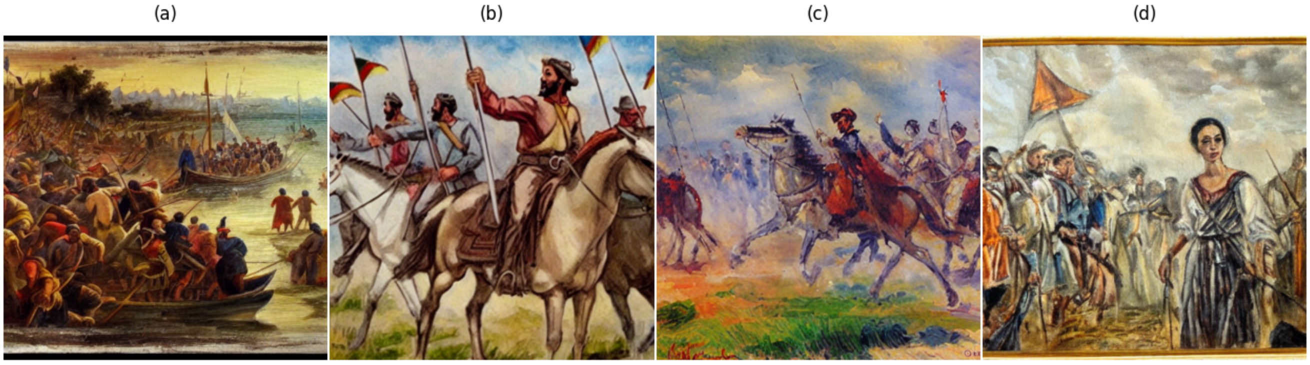

Figure 4. Dreambooth generated images to represent historical battles from the Farroupilha revolution. The prompts used were: (a) A painting of *batalhaFarroupilha*, crossing boats. (b) A painting of *batalhaFarroupilha*, republic ploclamation. (c) A painting of *batalhaFarroupilha*, republic ploclamation. (d) A painting of *batalhaFarroupilha*, a woman in the center of the battle.

asserting their determination to create a republican government. This act was a defining moment in the Revolution's narrative. The "Proclamation of the Republic" was a crucial event in the Farroupilha Revolution's historical context. It took place on September 11, 1836, when the revolutionary forces, led by Bento Gonçalves, declared the independence of Rio Grande do Sul from the Brazilian Empire and the establishment of a republic. Figure 4 (d) pays tribute to an often-overlooked aspect of the Farroupilha Revolution: the significant role of women in the conflict. Women played diverse and crucial roles during this period, including serving as dedicated nurses and providing essential medical care to wounded soldiers. They were instrumental in managing logistics and procuring and distributing vital supplies. Some acted as couriers, relaying messages and contributing to intelligence efforts. Additionally, women supported troops by maintaining camps, preparing meals, and ensuring soldiers' well-being. They also undertook the solemn task of burying fallen soldiers, ensuring a proper and respectful farewell.

Much of the visual documentation of these occurrences comes from paintings. Typically, the scenes depicted in these images consist of numerous mounted combatants, lacking a central subject in the visual samples. Similarly to the historical personality datasets, these concepts are scarce in terms of the quantity, diversity, and quality of images; however, DreamBooth generates satisfactory images. For battles generation, we trained our models with the Batalhas da Farroupilha dataset that includes images of generic battles (most of them), the proclamation of Rio Grandense republic, Crossing of the Lanchões, and also, images of the feminine representation of the Revolution.

5 Conclusion

In conclusion, our comprehensive exploration of DreamBooth's generative capabilities, framed within the context of Rio Grande do Sul has yielded promising outcomes. The fine-tuning process allowed us to capture the essence of Southern Brazilian culture, showcasing cultural elements, such as, "cuia e bomba de chimarrão", endangered species, including, "Sapinho-admirável-de-barriga-vermelha", and the traditional attire of RS, achieving the goal of generating images of local concepts and communities, historical moments and figures, and regional fauna and flora, of which, many of these concepts are scarce in high-quality data.

Throughout our experimentation, specific areas for improvement and best practices have come to light:

1. **Exploration of Diverse Training Configurations**: We recommend delving deeper into a broad range of training configurations to identify superior settings for image generation. Additionally, further exploration of hyperparameters and newer models, such as Stable Diffusion 2.1 and Stable Diffusion XL [Podell *et al.*, 2023], can help enhance the overall quality of the generated images.
2. **Diversification of Prompts**: By refining and expanding the input prompts, one can produce a more varied and evocative array of images, further enriching the user experience.
3. **Additional sources of imagery**: Future work could explore additional sources of imagery, such as local archives, museums, or community contributions. Collaborating with local historians or cultural organizations could further enhance the dataset's richness and improve the accuracy of specific cultural representations. These steps would allow for more comprehensive datasets, potentially resulting in even more detailed and authentic outputs.
4. **Representing Scale**: Additional refinement could entail incorporating images that present objects from multiple perspectives and contexts. This approach may enable the model to acquire a more profound comprehension of object scale, thereby enhancing the generation of representations that are both more realistic and contextually precise.

In conclusion, it is important to emphasize that the elements presented here are not inherently valuable as standalone works, but they serve as illustrative examples of how generative AI can be harnessed to represent crucial concepts for various communities. For instance, the chosen endemic animals are currently facing extinction, and the ability to generate representations from the available data in a repository could hold significant long-term importance. Similarly, the historical figures and battles depicted here have long faded into the past, with very few accessible images in public archives.

In this context, generative AI can play a vital role in creating more visual content related to such events and individuals for educational purposes, or other potential applications. Furthermore, we must acknowledge the potential for blending styles from various historical art movements, such as expressionism and modernism, or even the techniques of long-deceased artists like Van Gogh and Tarsila do Amaral. In essence, these examples underscore how diffusion models can be employed to generate representative art for local communities, allowing for the preservation and revitalization of cultural and historical significance.

Declarations

Acknowledgements

We acknowledge the support of the National Council for Scientific and Technological Development (CNPq).We would also like to express our heartfelt appreciation to our team of linguists who curated the dataset used in the training of our models.

Funding

This research is supported by the National Council for Scientific and Technological Development (CNPq/MCTI/SEMPI Number 021/2021 RHAE - 383035/2022-8) , Ministry of Science, Technology and Innovation.

Authors' Contributions

AZ and FM conducted the experiments. WC and MS provided oversight for the research. All authors contributed to the paper's composition, with MS and WC contributing to the study's conception. The final manuscript was reviewed and approved by all authors.

Competing interests

The authors declare that they have don't have any competing interests.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available by request.

References

Bernardes, A. D. (2021). O chimarrão como patrimônio imaterial gaúcho: os sentidos atribuídos ao desejo de preservação. Available at:http://www.monografias.ufop.br/handle/35400000/3409.

Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., *et al.* (2023). Improving image generation with better captions. *Computer Science*. Available at:https://cdn.openai.com/papers/dall-e-3.pdf.

Donahue, C., McAuley, J., and Puckette, M. (2019). Adversarial audio synthesis. DOI: 10.48550/arXiv.1802.04208.

He, Y., Yang, T., Zhang, Y., Shan, Y., and Chen, Q. (2023). Latent video diffusion models for high-fidelity long video generation. DOI: 10.48550/arXiv.2211.13221.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA. Curran Associates Inc.. DOI: 10.5555/3495724.3496298.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). Lora: Low-rank adaptation of large language models. DOI: 10.48550/arXiv.2106.09685.

Hughes, R. T., Zhu, L., and Bednarz, T. (2021). Generative adversarial networks–enabled human–artificial intelligence collaborative applications for creative and design industries: A systematic review of current approaches and trends. *Frontiers in Artificial Intelligence*, 4. Available at:https://www.frontiersin.org/articles/10.3389/frai.2021.604234. DOI: 10.3389/frai.2021.604234.

Kaur, N. and Singh, P. (2023). Conventional and contemporary approaches used in text to speech synthesis: A review. *Artificial Intelligence Review*, 56(7):5837–5880. DOI: 10.1007/s10462-022-10315-0.

Liu, D., Long, C., Zhang, H., Yu, H., Dong, X., and Xiao, C. (2020). Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8136–8145. DOI: 10.1109/CVPR42600.2020.00816.

Liu, V., Qiao, H., and Chilton, L. (2022). Opal: Multimodal image generation for news illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–17. DOI: 10.1145/3526113.3545621.

Lobato, W., Farias, F., Cruz, W., and Amadeus, M. (2023). Performance comparison of tts models for brazilian portuguese to establish a baseline. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. DOI: 10.1109/ICASSP49357.2023.10097264.

Mantovani, A., Morellato, L. P. C., and Reis, M. S. d. (2004). Fenologia reprodutiva e produção de sementes em araucaria angustifolia (bert.) o. kuntze. *Brazilian Journal of Botany*, 27(4):787–796. DOI: 10.1590/S0100-84042004000400017.

Oliveira, M. (2012). *Garibaldi: herói dos dois mundos*. Editora Contexto. Book.

OpenAI (2023). Gpt-4 technical report. DOI: 10.48550/arXiv.2303.08774.

Packhäuser, K., Folle, L., Thamm, F., and Maier, A. (2023). Generation of anonymous chest radiographs using latent diffusion models for training thoracic abnormality classification systems. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. DOI: 10.1109/ISBI53787.2023.10230346.

Pinaya, W. H. L., Tudosiu, P.-D., Dafflon, J., Da Costa, P. F., Fernandez, V., Nachev, P., Ourselin, S., and Cardoso, M. J. (2022). Brain imaging generation with latent diffusion models. In *Deep Generative Models*, pages 117–126, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-18576-2_12.

Ploennigs, J. and Berger, M. (2023). Ai art in architecture. *AI in Civil Engineering*, 2(1):8. DOI: 10.1007/s43503-023-00018-y.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., and Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. DOI: 10.48550/arXiv.2307.01952.

Raj, A., Kaza, S., Poole, B., Niemeyer, M., Ruiz, N., Mildenhall, B., Zada, S., Aberman, K., Rubinstein, M., Barron, J., Li, Y., and Jampani, V. (2023). Dreambooth3d: Subject-driven text-to-3d generation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2349–2359. DOI: 10.1109/ICCV51070.2023.00223.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation. DOI: 10.48550/arXiv.2102.12092.

Reed, J., Alterio, B., Coblenz, H., O'Lear, T., and Metz, T. (2023). Ai image-generation as a teaching strategy in nursing education. *Journal of Interactive Learning Research*, 34(2):369–399. Available at:`https://www.learntechlib.org/p/222304`.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. pages 10674–10685. DOI: 10.1109/CVPR52688.2022.01042.

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., and Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22500–22510. DOI: 10.1109/CVPR52729.2023.02155.

Silveira, F. (2018). Gato-do-mato-pequeno (leopardus guttulus). Available at:`https://www.ufrgs.br/faunadigitalrs/mamiferos/ordem-carnivora/familia-felidae/leopardus-guttulus/`.

Sinotti, K. G., Kontz, L. B., and Júnior, O. L. (2015). A revolução farroupilha: o massacre de cerro dos porongos. *Revista Contribuciones a las Ciencias Sociales*, (27). Available at:`https://www.eumed.net/rev/cccss/2015/01/porongos.html`.

Sutedy, M. F. and Qomariyah, N. N. (2022). Text to image latent diffusion model with dreambooth fine tuning for automobile image generation. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pages 440–445. IEEE. DOI: 10.1109/ISRITI56927.2022.10052908.

Takagi, Y. and Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14453–14463. DOI: 10.1109/CVPR52729.2023.01389.

von Platen, P., Patil, S., Lozhkov, A., Cuenca, P., Lambert, N., Rasul, K., Davaadorj, M., and Wolf, T. (2022). Diffusers: State-of-the-art diffusion models. *GitHub repository*. Available at:`https://github.com/huggingface/diffusers`.

Weber, T., Ingrisch, M., Bischl, B., and Rügamer, D. (2023). Cascaded latent diffusion models for high-resolution chest x-ray synthesis. In Kashima, H., Ide, T., and Peng, W.-C., editors, *Advances in Knowledge Discovery and Data Mining*, pages 180–191, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-33380-4_14.

Yıldırım, E. (2022). Text-to-image generation ai in architecture. *Art and Architecture: Theory, Practice and Experience*, page 97. Available at:`https://www.researchgate.net/publication/366594739_Text-to-Image_Generation_AI_in_Architecture`.

Zalla, J. and Menegat, C. (2011). História e memória da revolução farroupilha: breve genealogia do mito. *Revista Brasileira de História*, 31(62):49–70. DOI: 10.1590/S0102-01882011000200005.

Zhang, S. (2023). Dreambooth-based image generation methods for improving the performance of cnn. In *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pages 1181–1184. DOI: 10.1109/ICETCI57876.2023.10176568.

Zhou, D., Wang, W., Yan, H., Lv, W., Zhu, Y., and Feng, J. (2023). Magicvideo: Efficient video generation with latent diffusion models. DOI: 10.48550/arXiv.2211.11018.