

Laboratório VISGRAF

Instituto de Matemática Pura e Aplicada

**An experimental approach to illustrations and artificial
intelligence**

Bernardo Alevato, Luiz Velho, Jorge Lopes

Technical Report TR-24-04 Relatório Técnico

February - 2024 - Fevereiro

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório é de única responsabilidade dos autores.

An experimental approach to illustrations and artificial intelligence

Bernardo Alevato, Luiz Velho, Jorge Lopes dos Santos

Bernardo Alevato (graphic designer, educator, researcher) PUC-Rio, Arts and Design department. Rua Marquês de São Vicente, 225, Gávea. Rio de Janeiro, RJ - Brasil. Email: <balevato@gmail.com>. Website: <<http://www.bernardoalevato.com>> ORCID: 0000-0002-1431-3695

Luiz Velho (mathematician, designer, educator, researcher) IMPA, Vision and Graphics Laboratory, Visgraf. Estrada Dona Castorina, 110, Jardim Botânico, Rio de Janeiro, RJ - Brasil Email:<lvelho@impa.br>. Website<<https://lvelho.impa.br/>> ORCID: 0000-0001-5489-4909

Jorge Lopes dos Santos (designer, educator, researcher) PUC-Rio, Arts and Design department. Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, Brazil. Email:<jorge.lopes@puc-rio.br> ORCID: 0000-0002-8162-8291

Abstract

In this article, the authors show the documentation of an experimental process using machine learning to train and generate images based on authorial illustrations. Resulting images and flowcharts of the assembly using AI-like tools are presented. Topics arising from the process are discussed, such as issues related to model bias and consistency of results. Finally, notes are made about future applications and the relationship between local cultures and artificial intelligence that can be generated from the recorded activity.

Mimeograph style

This article describes an experiment carried out using the training of an Artificial Intelligence (AI) model and image generation, using a series of illustrations commissioned for the project's videos."Capoeira in Rio de Janeiro 1948-82", website Capoeira History [1]. The text provides a view of the technical process of image generation, training of generative models and raises philosophical discussions about image production and culture. The series of 30 digital illustrations that fed the training were carried out manually, in graphics tablet, using Adobe Illustrator and Adobe Photoshop to compose cards and animated vignettes for the

videos of the project. The movies present information about the history of capoeira [2] in the north of Rio during the aforementioned period, consisting of testimonies from capoeira masters, footage of capoeira circles, archive images and influences and consequences. This series was stylized as illustrations taken from a mimeograph machine to correspond to reproduction techniques from the period in question, and used images of the interviewees in the videos as reference.

These drawings were used in this experience to understand applications, technical and language characteristics derived from AI tools. In this specific process, we look at ways to train the model additionally with drawings, producing a style. This training process is called model fine-tuning. We start from the assumption of the possible ability to coexist with processes, about how the act of designing could include the use of AI and vice versa, imagining a future scenario in which AI can assist illustration and design projects.

In a learning-through-practice methodology, let's use the production of a model that can reproduce images of similar designs, from an authorial perspective, as a tool for other designers, illustrators and artists. At this moment, it is not yet a project for practical application, but an investigative process that can indicate ways of use.

An important reference for this work is NextRembrant [3], which is a proof of concept that opens the discussion about generation through AI. In the project, a work of art is developed using collection data training, resulting in a 3D printed painting. Another reference are the experiments carried out with the collection of the Museu Paulista[4], with training and image generation from his works, among other experiences. From a technical point of view, the performance in this study is different, with the possibility of the illustrator himself using image generation through AI.

AI imaging systems work through deep learning, with the ability to produce images, and natural language, which enables the interface through text with the system.

The system chosen for image production was Stable Diffusion [5] created by an association of companies Stability.ai, Runaway ML and the University of Munich Compvis . Stable Diffusion was distributed through an open license, which made it possible to create training tools, such as DreamBooth [6] and LORA [7].

Other generative imaging systems by AI, despite being very capable for representation as Midjourney, do not have the freedom of additional training and control when developing these experiments.

This model works through diffusion and deep learning. The Stable Diffusion model is trained from the LAION 2B [8] image matching dataset, trained in English on the correspondence between images and text and which presents the possibility of containing biases. This warning is acknowledged in the Stable Diffusion page itself. Biases in image generative AI are results containing biases in prompt generation that can lead to sexism and ageism. Agree to many points with the analysis that running Stable Diffusion using the standard model “does not democratize the predictive technology underlying stable diffusion algorithms.” and that the constructive structure of Stable Diffusion appears transparent but carries a series of instruments of power, as is common on technological devices. In view of this, we point out that today there are simplified paths, suggesting two viable paths.

The first requires a computer with a powerful video card and some knowledge of computer systems and Python, but it works using models on a local computer and the user can interact freely with the system through a graphical interface. The second is hosted on websites that provide the service of using Stable Diffusion for image generation and training using machine learning. In the research, we used both paths, but ended up opting for the second, by its practicality, its process clarity, using the websites [getimg.ai](#) [9] and [Scenario.ai](#) [10].

These are sites, like several others, that provide image generation and training services over the internet. It is worth highlighting Leonardo.ai's [11] effort to produce an excellent image generation environment, and Scenario.ai's coordinated use of Stable Diffusion to create consistent characters, in the direction of providing a tool for creating games and which also ends being a multifaceted tool for training and image generation.

In the discussions we ended up observing the inversion of biases in the generation of images, caused by the trained material. As we will see, biases in this context generally favor white and young men in generic representations, something resulting from model training. Furthermore, we reflect on the concept of image authoring in AI and possible adoptions of AI in design development.

Development

DreamBooth

The series of images was chosen for its visual cohesion, which made it possible to observe the resulting variables more objectively. The illustrations used in the training were organized to make the training viable. The original framing was a 16:9 video with a sign next to it. The crop was cut in a 1:1 ratio and the text was removed, since Stable Diffusion does not understand lettering well and the focus was on illustrations.

Some training tests were carried out with this material (Figure 1). The images were scaled to 512px and the DreamBooth was used for additional training via the website getimg.ai. The keyword “gtmg style” was registered with the training to indicate the use of the style. In practice, it increases the probability of using the style, to the detriment of other possible forms of representation.

Figure 1: Some of the drawings registered for additional training.



Figure 2: Training flow and image generation using the DreamBooth

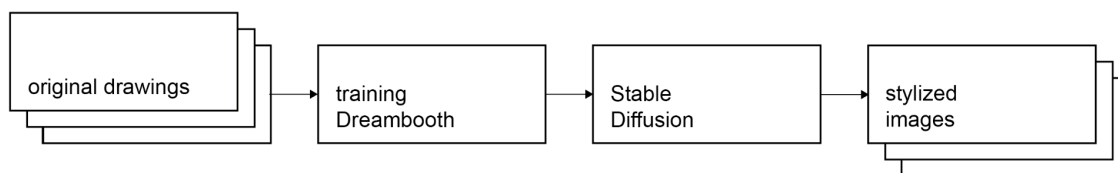


Figure 3: First results of the experiment.



Initial training using DreamBooth provided the generation of images with predominant characteristics in the originals, such as portrait framing, tracing over a photo and the use of contrast for shadows, which exist in almost all results. On the other hand, it did not result in a style as cohesive as what was characteristic in the originals. original images, the colors used are not accurate. Some images get the colors wrong. Approximately 500 test images were generated, some of music artists and others using generic prompts. In these tests, the predominance of people of African descent and the elderly is noted, something surprising, as it is the opposite of the bias of the model, recognized as a limitation in its publication. It seems that human representation follows style, even though it is only predominant. It is also evident that the resulting images are a mix between faces and similar styles, that is, the style is not completely pure, as it alters the color, contrast and visual solutions. In the Fig. 3(a) we can see an error in the model using yellow in the illustration.

On the model download card there is a warning that the model may generate images containing cultural biases, “which may exacerbate social prejudices”, as it was built from a LAION 2B dataset. This disclaimer is recognition that the basis works from a Western perspective. This recognized bias is part of society's imagery and, in a way, reproduces a power structure. After realizing that the resulting model gave greater importance to the human representation present in the training drawings, the model was tested from this

perspective. To test for bias a generic prompt “gtmg style Man with hat” was maintained across many generations (figure 4) and was used in other design models (figure 5), such as Archer Diffusion, Arcane Diffusion and Synthwave Punk v2 for image comparison.

Figure 4: Images generated with the model using the “man with hat” prompt



Figure 5: Images generated in the Archer Diffusion, Arcane Diffusion and Synthwave Punk v2 utilizando o prompt “man with hat”



Note the presence of different representations generated through the same prompt. Since many of these models tested for comparison represent series and animations existing, it is possible to consider that it is a representation problem inherited from other media.

Also considering the bias present in the images used, the male, since there are no women represented, the prompt “Portrait of Woman” was tested and the result maintained the previous inclinations, representing women of African descent who are often elderly (Figure 6).

Figure 6: A Results of prompt generations “Portrait of Woman”



Despite having some images that contain elements of architecture and different framings, this practically did not appear in any of the images, probably due to its minority in the series. And because in the case of architecture, the overlap between figure and background may have impeded this. O “berimbau”, a fundamental instrument that appears in some original images, also has little impact and appears distorted.

LORA

The same dataset of images of capoeira masters was used to train a style with the LORA method (Low rank adaptation). The training was done using the platform scenario.ai, which aims to be an AI tool to help build games. LORA creates an external, lighter, becoming an external addition to a main model. This step was after the experiment previously reported, so LORA was trained using Stable Diffusion XL in search of graphic and representation quality. It is important to note that the sdxl model was trained on a new dataset called laion5b, trained on 5 billion image-text pairs, which had improvements in terms of bias and coverage.

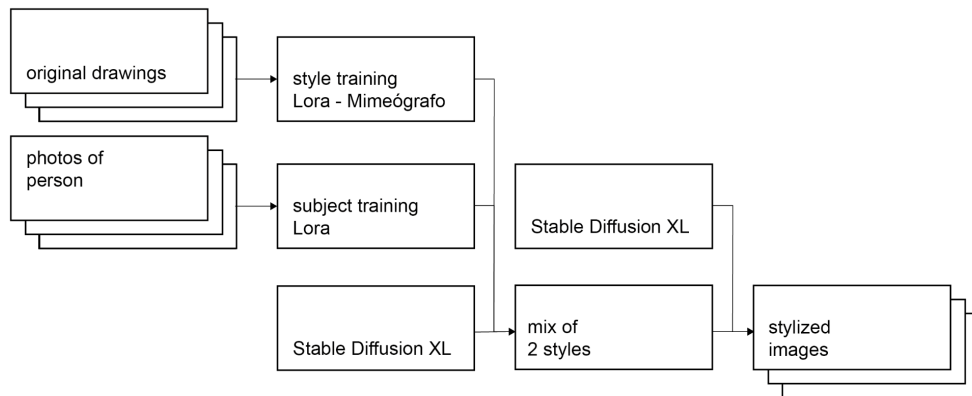
In this process it was also possible to view and edit the automatic descriptions of the images before training. By inserting the images we can see how LORA goes to classify them. This is done through computer vision, that is, an interpretation of the image followed by a textual description. In this classification the “berimbau” was described as a ski, a fishing rod and a sword, in different images. These texts have been corrected, but the difficulty in perceiving computer vision regarding Brazilian culture is notable.

Figure 8: Images generated with the model trained in LORA using the “man with hat” prompt and below images created from the prompt “Portrait of Woman”, finally in the last frame a portrait made with the model mimeograph remixed the photos of a specific person.



LORA generated images with a strong influence from the original style of the drawings, both in the lines and the registered faces. The attempt to create consistent images and generate new characters was successful when the model mimeograph was mixed with a model trained with photographs of a specific person. It is worth noting the color inconsistency, however, which is due to mixtures and weights given to the model to generate images with the characteristics.

Figure 8: Model generation scheme using LORA, considering the possibility of mixing models and creating consistent characters.



After this generation, LORA subject models were trained with photos from an author together with the Mimeograph model to generate images. These images presented a certain consistency in style and contained the appearance of the subject represented.

Considering the possibility of new characters with the style and of specifying characters training with photos and remixing, we can consider that training through LORA provides freedom for character consistency and variety in the production of images with the style. We can see that the colors are not yet exact, the line shows a small difference, as is the case with shading.

Discussion

Through these tests, we can see that the style carries with it the representation. Although it is a somewhat evident phenomenon in deductive mode, it is often barely visible, discreet. It's almost a reminder about visibility and erasure in the connected world. The discrete bias together with the reproduction scale forms an erasure of what is not hegemonic. This does not mean that the only way to deviate is through training, using *prompts* and references; certainly this is also possible with other methods, but this exercise brought visibility to the problem through contrast.

The reversal of bias, where elderly people of African descent primarily appear, is an unexpected result, which can enrich the perception of how systems work. Even when the style was wrong, the model was consistent with the representation, featuring black people, generally elderly, clashing with traditional behavior.

The association between drawing style and human representation reminds us that there is certainly an erasure of non-Eurocentric art in AI bases, this means that generic models with

interconnected images full of eurocentric art will probably have a similar behavior in image creation, towards standardization . These generated images will probably feed new datasets amplifying the bias effect. What will change this is the creation of new models that include cultures, ethnicities and improvement in computer vision under pressure from society. In a similar way to what happens with conversational models, where linguistic expressions can be lost or can be revitalized.

You can see more proportionate ethnicity and age behavior in the newer version, Stable Diffusion XL, and Adobe's AI tools. However, it is clear that the datasets used in training will always have some bias in their development, whether due to their sampling or classification and especially if they use today's internet as a source, which inherits many representation defects. This discussion certainly does not end, but in these experiments it is possible to think that by observing the process in an open way, possibilities arise to reverse biases, giving conscious emphasis to the content and form of the images.

According to the copyright denied by the court to the author of the controversial image [12] generated by AI that was awarded and disqualified from the digital image contest, the criterion of ownership of the authorship of an image created through AI is how much of the author's humanity is present. In this sense, training an AI model with original material seems like a generous way to use the technology. Although the authorship appears authentic, the originals appear to have more depth than their derivatives, similar to a printed copy of an original, despite not being copies. They are derivatives, just like source code and compiled material.

Considerations

Creating illustrations involves much greater complexity, and it does not seem possible at the moment to produce images using AI such as the Mimeografia series for the Capoeira project in Rio de Janeiro 1948-1985. Capoeira movements, instruments and representation are quite complex for generative mechanisms. Reading body movements will probably improve soon, and instrument repertoire could be a subject for training. And this training can produce interesting variations that sharpen creativity as they remix a style with other styles and human representations. Style is different from the look, the sensitivity. A reproduction of style differs from an artist's production. The use of generative systems as a tool is an appropriation that can be useful towards something similar to algorithmic reproducibility.

In the near future it is not difficult to imagine an illustration generator tool to support illustration guides such as IBM Isometric style [13] or even the Illustrations part of DASA Brandcenter [14]. But to do so, it would have to generate vector images, something more complex, including in terms of visual synthesis. For comic books in particular, it can be useful to use consistent characters. On the other hand, the ideal image format would be vector, with consistent curve treatment, if it's possible to add visual solutions given by illustrators and color accuracy. In this sense, it opens up possibilities for sequential illustration with character coherence, for books, animations and comic books. The resulting serial images are not exactly illustrations, as they are generated and not illustrated, but they correspond to an algorithmic technical reproducibility and other generative and parameterized production arts.

Generative AI models are culture propagators, so use of special material of non-hegemonic cultures and artists, such as the Brazilian, can be a good way to develop AI models. As is done in terms of conversational models in a project to revive languages, Maori [15] and Brazilian Indigenous languages [16] As well as training from the repertoire of the cultural universe and artists or cultural themes can generate derivative works or even interactive pieces for co-authorship for the general public. But it is important to reflect that imitating a style does not bring human vision and creativity. Transgressing one's own style is common in different phases of artists.

Generative tools are increasingly embedded in design programs, probably in a more opaque way. However, from the perspective of creating generative models, it is not difficult to imagine training and generation as a way of establishing a language visual cohesion, such as a brand application manual or even parameterized programs for generating shape variations. Instead of pure enchantment at the ability of AI models to produce images similar to photos and drawings, there may be interest in creating conditions for creating new images with personalized styles through training, noting that the critical aspect is essential for considerations about application and use. Tools like ComfyUI [17] can provide greater control over generation, using nodes, and visual coding, providing a greater diversity of options.

There should be an increase in the profusion of images on social networks, just as happened with photos from cell phones. This profusion of images is a natural consequence of an interaction that exists during the generative process and should occur quickly as this mode of image production becomes popular. We have already seen the growth in the production and

sharing of digital photographic images, their manipulation through filters and today they are beginning to mix with synthetic images produced by AI. This profusion could change the way we interpret images in general. It is to be expected that images produced manually will be given new meanings.

The training and generation experience is not technically complex and could be used as an educational activity, through a simple infrastructure. It can be useful for the digital literacy of general people in understanding machine learning and AI. With applications in real-time production, it can be useful for creating installations that use public collections and images, proposing new ways of looking at the collection.

The human representation accompanies the illustration style, is part of it, just like in the manifestation illustrated in the training. In capoeira, music, movement and game cannot be separated. The speed and depth of artificial intelligence action must be accompanied by a human assessment, in addition to the fundamental legal rules, observing leaving the box open and participating in society.

References and Notes

1. The illustrations are used on the YouTube channel available in 2022-23 <<https://www.youtube.com/channel/UC65yAcL4W2ZghIxI9Z04qww>>
2. Capoeira is an Afro-Brazilian cultural manifestation, a mix of game, fight and dance that manifests itself through the capoeira circle, where two members play while others watch and which is also driven by music using singing and instruments such as atabaque, berimbau and tambourine. The capoeira circle was recognized by the United Nations Educational, Scientific and Cultural Organization (Unesco) as Intangible Cultural Heritage of Humanity. To learn more about the culture and practice of capoeira see: Nestor Capoeira, *The Little Capoeira Book* (place of publication: short form of publisher's name, date) (Berkeley: Blue Snake Books; Translation edition, 2007)
3. See video description of the project <<https://www.youtube.com/watch?v=IuygOYZ1Ngo>> accessed 13 January 2022
4. O artigo de Savic sobre IA que discute a utilização do Stable Diffusion: Selena Savic, Yann Martins. Resisting Appropriation and Assumption of Access /Reason in Machine Learning Training Processes. WEIZENBAUM JOURNAL OF THE DIGITAL SOCIETY Volume 3 \ Issue 2 10-30-2023 <<https://ojs.weizenbaum-institut.de/index.php/wjds/article/view/89>> accessed 13 January 2022

5. Stable Diffusion original paper <<https://arxiv.org/abs/2112.10752>> accessed 13 January 2022
6. DreamBooth original paper <<https://arxiv.org/abs/2208.12242>> accessed 13 January 2022
7. LORA original paper em <<https://arxiv.org/abs/2106.09685>> accessed 13 January 2022
8. Laion 2b is a dataset of 2 billion image-text pairs that was used to train stable diffusion 1.5. It was launched by Laion, a non-profit organization. See more: <<https://laion.ai/>> accessed 13 January 2022
9. Getimg.ai website offer image generation and training services. accessed 13 January 2022. <<https://www.getimg.ai>> accessed 13 January 2022
10. Scenario.ai offers image generation and training services, with an emphasis on the production of characters and objects for games. <<https.scenario.ai>> accessed 13 January 2022
11. Leonardo.ai offers image generation and training services, high quality images and motion from picture. <<https://www.leonardo.ai>> accessed 13 January 2022
12. Article from Wired magazine that discusses the subject available at <<https://www.wired.com/story/ai-art-copyright-matthew-allen/>> accessed 13 January 2022
13. <<https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-peop le/>> accessed 13 January 2022
14. Project website <<https://conexoesoriginarias.github.io/index.html>> accessed 13 January 2022
15. IBM Illustration style manual <<https://www.ibm.com/design/language/illustration/isometric-style/design/>>accessed 13 January 2022
16. Illustration style manual <<https://brandcenter.placcopazzinidev01.com.br/ilustracao/>> accessed 13 January 2022
17. ComfyUI, graphical user interface with visual coding <<https://github.com/comfyanonymous/ComfyUI>>