

a monkey in a spaceship



pietro
ercolino
vizzardelli
barcucci

June 2023

Submitted by Pietro Ercolino Vizzardelli Barcucci

Supervisor: Haig Armen

Master's degree in Interaction Design at the Faculty of Design, The Estonian Academy of Arts

abstract

This master thesis is interested in transforming my own practice towards a deeper and more considerate engagement with technology. It is an attempt to experience what an alternative approach to artificial intelligence (AI) can lead to.

To uncover new ways of relating with technology, I focus on building a personalized tool that can empower my expression ability.

Situated at the intersection of art, design and philosophy, the study develops as a practical experiment that explores what is the connection between music and image, depicting a personal journey across the two mediums.

As I proceed to define my interaction with AI, I propose how addressing it in a very instinctive and irrational way can lead to a prolific collaboration despite of its limitations. This experiment unfolds into a narrative that speculates around the future of technology, as a tool capable of fostering creativity instead of automating it.

a monkey in a spaceship

1. introduction

1.1 why?

1.2 what?

2. methodology

2.1 autoethnography

2.2 explore

3. tools

3.1 artificially intelligent tools

3.2 tutorial

4. reflection

4.1 ai and creativity

4.2 the connection

4.3 a monkey in a spaceship

bibliography

1. introduction

1.1 why?

Why? This is usually the first question I ask myself.

Ironically, it is rare to find a clear or appropriate answer. Most of the time it remains there, pending. It is hard to find reasons, or to decompose any problem into the simpler causes that generated it in the first place.

I really like to leave this question hanging, but still, I find it useful to constantly ask it to myself, regarding what I do, what I think, see, hear or feel. Trying to reach the root of things helps me to place and relate them to myself.

Why am I doing this project? Mostly, I am doing it out of curiosity, interest, and understanding.

Many things for me start with curiosity, the idea of the unknown. Pure eager to witness how things are, or might be, and how they entangle with each other seamlessly.

The motion is initiated by curiosity, and then it is led by pleasure. Interest is a formal, rational form of pleasure. The topics tackled in this project are personally very close to me, and I like to let my mind wonder around them. Pleasure, or interest, can be broken down to different levels: from the very sensory, to high level sense of accomplishment, safety, or long-term gains.

Now, as the third factor, saying understanding is probably pretentious, because understanding would then suppose the ability to master the acquired knowledge, while most of the time, the more I know, the smaller I feel. Maybe experience would be a better word. With experience, my set of tools to decode and shape reality broadens. I might be extremely puzzled about what I am going through, but being aware of it, in a way, still generates memories of some kind. Memories that I will be able to evoke and abstract, comparing them to other experiences and then formulating a sort of personal experiential knowledge.

These three properties are usually very big drives in my day-to-day life (leaving aside necessities and bad drives such as guilt, responsibilities, ego, greed...). They are a big part of what keeps me in motion, both introspectively and through society. This motion comes naturally to me, or at least I have learnt to follow behavioral patterns that would benefit me in the long run. Nonetheless, following this approach in the thesis resulted in unexpected outcomes, and the difference is in the fact that I pushed myself to do this with consistency and method. This project has been a great opportunity to reflect and dig into myself.

1.2 what?

The nucleon of my research is to explore unforeseen creative realms. Navigating through the visual field following the lead of music.

As a partner in this research, I tightly worked with artificial intelligence, and learnt how to use it as a tool and partner through my exploration. I trained artificial neural networks to generate images based on my own drawings.

Music has been an ally throughout the whole process, working for me as a probe, influencing and indirectly piloting my thoughts and actions. I want to explore what the connection between image and sound is. Understand what this connection means to me. There are many ways to analyze the intersection of these two worlds, but none of them would be comprehensive enough to properly depict the issue because of the

elusive nature of its subjectivity. The richer examples surely belong in the art sphere, from all epochs and through any possible form.

For this reason, to contribute the topic, I will just portray what my path has been through it, and how this relatively short thesis study could serve as a metaphor to speculate about alternative possibilities of interaction with artificial intelligence, and the reasons why developing personalized tools is worth undergoing their slow building and learning phase.

2. methodology

2.1 autoethnography

Taking the nature of this project into account, the object is the self-research. Being it a personal, mostly artistic, sort of discourse, implies that the majority of knowledge built around it comes from the inner self.

The personal research acquires in this case a different importance and becomes a valid view on the argument without it being confirmed by any particular jurisdiction.

This characteristic makes it hard to generalize and to compare with existing practices, theories and literature. Resulting somewhat detached from institutional authority.

Autoethnography allows personal perspective to acquire documentation validity [1].

Personal perspective is biased, it only exist inside of a culture, inside fears and hopes, reflecting everything I have experienced and internalized so far. In a way, it has as little objectivity as it can. It just tells a story, a story that mostly makes sense to the subject, a story which final purpose is not to be judged and evaluated by others. Nevertheless, it is the duty of the researcher to make the research relatable and understandable if it's meant to be shared. It is part of the duty to express and make the personal experience navigable by others. To do so, it is necessary to compare and contrast subjective

experience with existing research and try to make it familiar to whomever is not coming from the same culture.

It is fundamental to record the process analytically and elaborate epiphany moments, since they are usually very focal steps in the development of introspective activity.

2.2 exploration

Writing and drawing are the main mediums in which I keep record and advance my research. I have also introduced the “self” as the source of the research, meaning that most of the material comes from or through myself.

Nevertheless, when it comes to the content of the research, or better, the exploration, I find it most prolific to look outside of myself. It is fundamental for me to take inspiration from the environment in order to deviate from my habits, mental models and comfort areas. When I try to seek inspiration introspectively, I find it hard to deviate from what is already there, to reach any novel idea that is not a mere recombination of other ideas or experiences that I internalized.

Especially when the research is aesthetic, or creative, I find it very useful instead to find hints in the pure unknown. To find it elsewhere, not in what is already in front of me or in my conscious brain.

I just find it easier and more powerful to get input from the external world. Not because it is more articulate and beautiful (there’s so much complexity in the human body and mind), but because I have better tools to explore it. Using senses feels faster and sharper than just reason and speculation when trying to explore something new. This is probably very subjective, it is just the way I found that better work in my practice.

The derive method (French for “drift”) was developed during the 1950s by Guy Debord, a member of Letterist International, an avant-garde collective based in Paris.

Dérive is defined as a “mode of experimental behavior linked to the conditions of urban society: a technique of rapid passage through varied ambiances.” It is a journey through (usually urban) landscapes in which the practitioners do not follow a structure and drop

their regular relations, following alternative conceptual, aesthetic, emotional, visual or social patterns during their exploration [REF2].

The way in which I have interpreted and performed this practice is to explore an unknown territory following a coherent theme or set of rules. This territory can be physical (literally wandering through a city), but also metaphorical, for example while drawing. The idea is to investigate a topic, a tool, or a city according to a set of guidelines that deviate from your usual flow. Or just a way to probe lands and topics following a defined key of interpretation.

This technique helped me to deviate from my usual path. To navigate physical and abstract lands through ways that I would otherwise never run across.

I was surprised by how deriving made me discover multiple cities in new ways. I remember one walk I had in Athens during the early stages of this project: while roaming around, any time I would be indecisive about which way to take, I would then choose the least inviting of the options, the opposite of what I rationally usually do. That walk ended up being extremely pleasant and inspiring. Usually, I get teased by similar elements while trying to decide where to go at a crossroad, taking the opposite decision of what I would normally do brought me to landscapes and situations that I would normally avoid.

Deriving has been useful throughout the aesthetic exploration as well. While drawing, I sometimes would force myself to follow a specific lead like the route of the unpleasant, unexpected or uninviting, as well as other themes.

Gathering new ideas and inputs is fundamental to create something novel.

Human brain is capable of abstracting knowledge, nevertheless, an epistemological empiricist would argue that anything it generate will always be a combination of knowledge and data acquired through sensory experience. Just like we cannot grasp how a four-dimensional world would look like, or cannot imagine a completely new color.

This perspective might be somewhat limiting, especially when it comes to the abstraction of ideas and many scientific and philosophic schools of thought are in opposition to the empiric limitation.

Truth or not, it is vital for me to experience something new to generate new knowledge or navigate new aesthetics. For this reason, I have chased and celebrated randomness, and tried to include it in my work in different ways.

Whilst drawing, I like to start with something random, and then elaborate on it. Some of you might know the feeling of standing in front of a blank page, and how it is intimidating when you must fill it with words or draw something. It is hard to breach through this feeling for me, and using something random breaks the tension. It could be just looking around, see what the eye catches and representing it, or just closing your eyes and let the hand go free for some moments. Usually I would try to follow the motion that music would push me to. Most of my drawings have started in this way, letting the pen dance for me.

Randomness can be intriguing, but far from pleasant most of the times. It just serves as a forerunner. Once I have the path set by randomness, I start applying layers of meaning. I start elaborating with criteria and knowledge that I developed throughout the time. I can now follow the aesthetic principles that I perceive as pleasant.

This second phase, the asynchronous analysis and shaping using conscious brain also has its limits and, as I explained, tends to always follow the same patterns.

The magic happens when you run the two phases in parallel. Whilst you rationally decide how to modify and shape the drawing, the hand will keep following an automatic motion and other random (or music-led) instinctive factors. I have found this two-steps technique to be most prolific and enjoyable. The key aspect is to keep a sustained rhythm while doing it and not interrupt the momentum.

3. tools

3.1 artificially intelligent tools

“A tool is a moving entity whose use is initiated and actively guided by a human being, for whom it acts as an extension towards a specific purpose. This definition is explicitly kinetic, yet it is open to abstraction> the entity can be physical or conceptual.”

There are many tools based on artificial intelligence to create images.

For instance, some recent popular applications like DALL.E [3] allow the user to generate images from a text prompt. These applications rapidly rose in capabilities and can now achieve incredible results.

What motivates me to create my own tools based on AI when there are publicly available technologies with entire teams of engineers behind?

I wanted to make a more specific, more personal tool. This motivated me to train my own Neural networks and not use already existing commercial products.

One of the problems of this large commercial AI tools is that you can clearly spot them. A good prompt will result in a good image, and the better you can describe what you want, the better the output will be. But with a little bit of experience, it's easy to spot when images are created like that.

I trained my own neural networks from scratch and did not use some already available models. The amount of resources that I can provide is not comparable to the ones available to a team of professionals. But this is the only way to build my tools exactly the way I want them to be. Detached from any form of data that is not relevant to this purpose.

The two main constraints to AI-based image generating tools are data and computational power. The training of a network can take an incredible number of resources. State of the art systems like DALL.E, which can generate images from a text prompt, are trained on hundreds of millions of captioned images from the internet and, according to the paper, it takes approximately 100'000 – 200'000 CPU hours to train the model [4].

You might have a glimpse now of how unbalanced can be a competition between a system like this and me, armed of my colored markers and a 5 years old laptop. Because of this, I had to compromise some aspects of the result. The reasons are exactly having a small number of samples to train the network on, and computational power. Ideally, tens of thousands images would form an adequate training dataset, whilst I manually drew each of my samples. Additionally, it takes a lot of computation to train a network: while trying to train the network for the first time, it was with great disappointment that I discovered that my computer would almost immediately crash and fail. I had to outsource the process and execute the training on an expensive virtual GPU, somewhere remotely.

In any form of art, the tools used are fundamental. Using tools gives us the ability of shaping matter and create artifacts. The usage of any tool has to be mastered, and this is what makes many artworks unique and valuable. Starting from cave paintings, all the way to digital art. Their characteristics, as well as their limits, are deeply intrinsic to the way the output will be.

Two key aspects of creation that will result in a valuable artwork are the form, or idea, that the artist is trying to achieve, and the technique that they use in order to achieve it. The implementation phase is strictly bound to the tools and mediums involved. For

instance, realistic paintings are appreciated not only for their composition and visual impact, but also because the viewer recognizes the level of skill required to portray a high number of details through the usage of brushes and paint. The skills of some artists can be considered unique in their kind, and that is what makes their work worthy.

The used tools can be kinetic, or practical effectors, like an extension of the body. This gives the artist a new means to perform tasks, tools used in fine arts usually fall under this category, like a brush in painting or a chisel in sculpture.

But tools can also be extensions of the mind, or information processors. This archetype of tools is more frequently used in new art forms like digital art. Software can be a tool. When tools are complex and abstract, it is harder to understand what their role is in the creative process.

Up to now, most of the AI-based art being produced has been made using artificial neural networks, a system that mimics the behavior of neurons to learn how to perform tasks.

When the goal is to replicate some kind of existing material, many networks are based on an architecture called Generative Adversarial Network (GAN). [5]

A GAN is constituted by two different entities: the generator and the discriminator. They start from scratch, and they improve together. The generator has no access to the training dataset of images whatsoever. All it does is try to put pixels together and present them to the discriminator. On the other side, all that the discriminator does is to guess whether an image it receives is fictitious or not, understanding if it comes from the dataset of real images or from the generator. They keep doing this simple guess game until the generator becomes good enough that its output cannot be distinguished from the training one.

In this way, everything that the generator does is not just a re-combination of elements memorized from the dataset, because it doesn't even know what is in there. It just starts with random "noise" and slowly learns what kind of elements, patterns, shapes and colors usually succeed in fooling the discriminator into believing that the image is real.

3.2 tutorial

I have used a variation of the GAN model that can be trained using a limited amount of data [6]. This is possible because the images get split into a lot of different variations of themselves, changing the colors, offset, zoom, and other parameters. For this reason, it means that the resulting output will be less faithful to the training samples and the variety that the results can achieve is somewhat limited, but it's the only way to overcome the limitation of having a small pool of images like in my case.

The reason why I chose to develop my own tool, and not to use pre-existing software, is to create a personalized workflow. Building it from scratch comes with obvious downsides, and requires effort, moreover, the results will hardly be competitive with other available commercial tools. Nevertheless, the advantages are uniqueness and flexibility. Personalized tools will be free from all the tacit knowledge, biases and aesthetics that are embedded in tools made by other people, or tools that are made to perform generic tasks or to achieve certain goals.

If compressed to its essence, the tool that I used can be explained very simply as a black box (an object whose internal functioning is unknown and unnecessary to know) that receives images as input and is then able to generate other similar images as an output. This object is implemented as a layered set of mathematical functions that can be broken down to very simple instructions. It comes to life inside of a deep learning framework called PyTorch, an abstract environment in which generic purposed software can be modified by adding code and making it specific to the user's application. The programming language used was Python.

Practically speaking, the main instructions that I have to communicate with this black box are for training the network, and to generate results once the training is complete. To execute the training of this kind of network, one would need around one hundred images to obtain consistent results. On a single GPU, training this model takes between 6 and 24 hours, according to its specific characteristics.

Like every tool, using AI for this purpose also needs to be mastered, and understanding its behavior comes with experience. This can only happen through the usage of the tool, and the subsequent exchange of information, or feedback, that one is able to gather from its utilization.

For these reasons, the main limitation I have found regarding the usage of this tool is the speed with which I can communicate with it, the pace at which I am able to receive [feedback \(image 1\)](#). Not only the training itself takes a relatively long time and computational resources, but most importantly because of the nature of the training material. Besides understanding the way that the AI learns, I also have to act and react to that by hand drawing all the images. Making multiple drawings per day for a prolonged amount of time has been a challenging, yet enjoyable task. With the purpose of creating pleasurable images and, most importantly, the underlying aim of conveying a consistent set of aesthetic elements for the AI to grasp and be able to replicate.

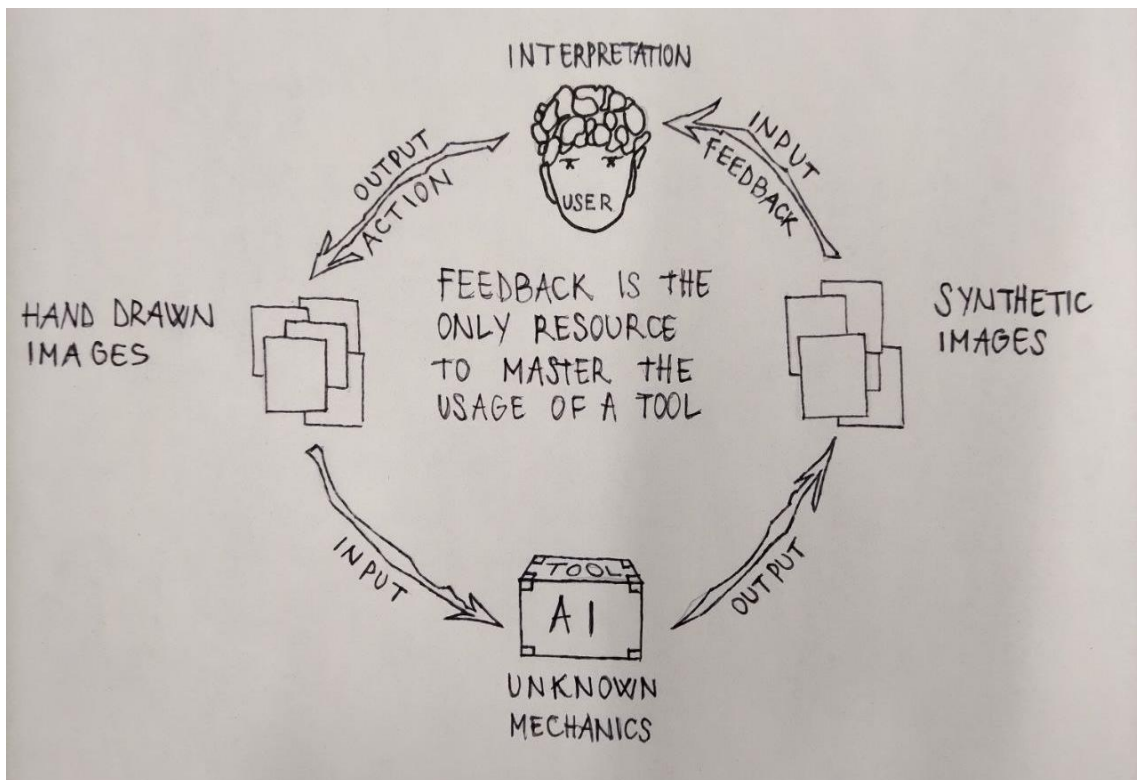


Image 1: A scheme representing interaction between the user and AI. Roles, actions, and affordances are highlighted.

Images should be homogeneous in their format. For instance, drawing on different materials would result in the network trying to reproduce this visual diversity of materials, this is going to require cognitive effort following the same trial and error format. By drawing on the same paper and size, more emphasis will be dedicated to other factors like shapes and colors. More generally it is easier for a network to converge and create good output images if the training dataset has a homogeneous set of traits, yet diverse in their manifestations, so that it can learn how to make abstractions of those traits and then be able to replicate them, making new artificial instances.

Good examples can be models trained on human faces. The results are extremely realistic exactly because there is a humongous amount of data to train it, and every human face has a distribution of shared traits that the network learns to replicate. [7] The last argument about homogeneity, together with a list of other precautions and small decisions, make part of my initial assumptions about the topic and the way in which I should operate with this simple, yet mysterious, tool. I had to strongly rely on decontextualized information and perspective I acquired through time regarding AI generated images.

There are mathematical ways to assess a network's performance. One example can be a metric like Fréchet Inception Distance (FID), which compares the distribution of the generated images with the samples belonging to the dataset [8]. FID can be useful when comparing precise data, like samples of lungs X-rays, but fails short when it is about aesthetics, because having images that are literally similar to the dataset doesn't necessarily mean that they capture the elements that result in them being pleasurable. When I started drawing images to train the network for the first time, the majority of the drawings were made with a sphere pen and had a very scribbly feeling. Some of them are abstract, but most of them incorporate elements from the environment around me ([Image 3a](#)). I was hoping for the AI to grasp, at least partly, some of the semantic definition of the elements I would draw. I was hoping that, making enough of

certain kinds of features and shapes, the network would learn to replicate them. I was quite wrong.

Being able to see digital “fake” images that slightly resembled my drawings was stupefying, but the result was not quite what I was planning.

The main difficulties lay exactly in the understanding of the training process. This perceptual gambling game deeply relies on tacit knowledge.

All I could use to improve the subsequent experiments is visual feedback and non-verbal / non-rational instinct to take any decisions about how to adapt my own input to the network, how to drive the training towards the direction that I want it to take.

This process usually happens during the learning phase with every tool. Humans are tools users, and the cognitive process of understanding and mastering the usage of new tools is one of the fundamental reasons behind our fast evolution into the dominant species on the planet.

For the subsequent prototyping iteration, I included bigger shapes filled with colors, and the network was surprisingly able to replicate those elements [\(Image 3b\)](#).

For the third and final iteration, I kept developing on colorful patterns and leverage on a certain type of spiky, thorn-like shapes [\(Image 3c\)](#).

It has been interesting to question and deconstruct my own ability of using experiential knowledge to adjust the way I interact with this tool. I had to deeply rely on sensorial intelligence and instinct, trying to be actively aware of it and the mental models tucked behind these instinctive reactions. Cause of that is the amount of feedback I can get back from the tool, which is just very limited, that is usually what makes tools users experts in their field.

Technically, the training dataset is the only thing that will affect the result, yet the network is non-deterministic. This means that the same model trained on the same dataset two different times will generate different results. This makes any trained network unique and not replicable. This characteristic makes it more valuable because

of its uniqueness, but also makes it hard to know when to stop, to accept it as it is, and not try to make it better.

In this case, I have decided to make one and one only execution of training for each of the iterations and not try again. This goes against the usual design process, that would assume a fine-tuning loop of feedback and testing until you reach the best results, but the main reason why I decided to impose this “one-shot” rule is because of the high resources necessary for the computation of the training. After running the train command on my own computer, it crashed within seconds due to the virtual memory getting overloaded. I needed much superior hardware infrastructure to continue. I outsourced the problem by renting a cloud top-tier GPU, which is expensive and consumes lots of electricity.

Beside the resources problem, accepting the network as it is after the training makes it a performative act. It pushed me to place more effort and foresight in the preparation phase: knowing that I do not have multiple tries made me face the process with greater prudence and focus.

This also made me develop a sort of ritualistic approach to the training. A moment you prepared so long for, doing the best to facilitate and create all the prerequisites for it to work flawlessly. Once it starts, there’s nothing else to do. Just a long wait. A wait that drips with hope and fear.

Some attempts have failed, for reasons to me obscure. All I got are strange error messages with no real documentation to check what the fault has been. Spending time to truly understand it has not brought to much advancement, I just tried again and hoped for every bit of code to take part in this complex dance without any bankruptcy. It’s a long wait, submerged within uncertainty, where you can only be a spectator.

Hoping to get some sense out of the verdict. Just like in a ritual, the practitioner can set the whole scene up, and be ready to receive signs from an external, greater entity.

There is no truth, no wrong, no beauty, nor disappointment. They only exist in the eyes and mind of the beholder, who is vividly trying to extrapolate meaning out of this obscure prophecy.

Once the ritual is over, all you get is a pile of images, nothing else. It's all there, for you to look, alongside the grace, the doubt, and the charm of interpretation. Free from any form of reason and causality, and the blank of prospects.

```
Help All changes saved
+ Code + Text
41% 20400/50001 [5:59:10<8:18:34, 1.01s/it]GAN: loss d: 0.35922 loss g: -1.21374
41% 20500/50001 [6:00:52<8:39:36, 1.06s/it]GAN: loss d: 1.84423 loss g: -3.28026
41% 20600/50001 [6:02:36<8:59:29, 1.10s/it]GAN: loss d: 1.56704 loss g: -2.26267
41% 20700/50001 [6:04:19<9:05:16, 1.12s/it]GAN: loss d: 1.03660 loss g: -2.14593
42% 20800/50001 [6:06:03<8:31:24, 1.05s/it]GAN: loss d: 0.75234 loss g: -1.34603
42% 20900/50001 [6:07:45<8:50:32, 1.09s/it]GAN: loss d: 3.02805 loss g: -1.96560
42% 21000/50001 [6:09:29<9:33:43, 1.19s/it]GAN: loss d: 3.21044 loss g: -1.58689
42% 21100/50001 [6:11:12<9:11:08, 1.14s/it]GAN: loss d: 2.90385 loss g: -2.18651
42% 21200/50001 [6:12:56<8:39:21, 1.08s/it]GAN: loss d: 2.09898 loss g: -2.21875
43% 21300/50001 [6:14:39<9:03:14, 1.14s/it]GAN: loss d: 2.56859 loss g: -1.54472
43% 21400/50001 [6:16:21<9:28:54, 1.19s/it]GAN: loss d: 1.65715 loss g: -1.97297
43% 21500/50001 [6:18:06<8:29:49, 1.07s/it]GAN: loss d: 0.73888 loss g: -1.37049
43% 21600/50001 [6:19:48<8:56:48, 1.13s/it]GAN: loss d: 1.40619 loss g: -1.65438
43% 21700/50001 [6:21:32<9:13:11, 1.17s/it]GAN: loss d: 0.64242 loss g: -1.43118
44% 21800/50001 [6:23:15<9:06:38, 1.16s/it]GAN: loss d: 2.98704 loss g: -1.86908
44% 21900/50001 [6:24:59<9:35:28, 1.23s/it]GAN: loss d: 0.85676 loss g: -1.44792
44% 22000/50001 [6:26:42<9:29:19, 1.22s/it]GAN: loss d: 1.54599 loss g: -1.90375
44% 22100/50001 [6:28:27<9:32:37, 1.23s/it]GAN: loss d: 0.96528 loss g: -2.97705
44% 22200/50001 [6:30:09<8:43:22, 1.13s/it]GAN: loss d: 2.05457 loss g: -1.52751
45% 22300/50001 [6:31:53<8:03:49, 1.05s/it]GAN: loss d: 1.18323 loss g: -2.49199
45% 22400/50001 [6:33:35<7:44:39, 1.01s/it]GAN: loss d: 1.70770 loss g: -1.97214
45% 22500/50001 [6:35:19<6:50:03, 1.12it/s]GAN: loss d: 2.72399 loss g: -0.96260
45% 22600/50001 [6:37:01<6:50:46, 1.11it/s]GAN: loss d: 2.51574 loss g: -3.28195
45% 22700/50001 [6:38:44<7:01:38, 1.08it/s]GAN: loss d: 1.57430 loss g: -1.13566
46% 22800/50001 [6:40:29<7:14:42, 1.04it/s]GAN: loss d: 2.63512 loss g: -1.96393
46% 22900/50001 [6:42:13<6:53:30, 1.09it/s]GAN: loss d: 0.83454 loss g: -2.56347
46% 23000/50001 [6:43:58<6:55:41, 1.08it/s]GAN: loss d: 0.92636 loss g: -1.62689
46% 23069/50001 [6:45:10<6:55:38, 1.08it/s]

Executing (6h 45m 32s) Cell > system() > _system_compat() > _run_command() > _monitor_process() > _poll_process()
```

Image 2: a snapshot of the training process.



Image 3(a).

Image 3 (a), (b), (c): In each of the images, the first two rows contain handmade drawings from the training dataset. Third and fourth rows show selected samples made with AI.

(a), (b), (c) respectively show the first, second and third prototyping iterations, the training dataset and the resulting images. The images should emphasize the aesthetic evolution of each iteration.

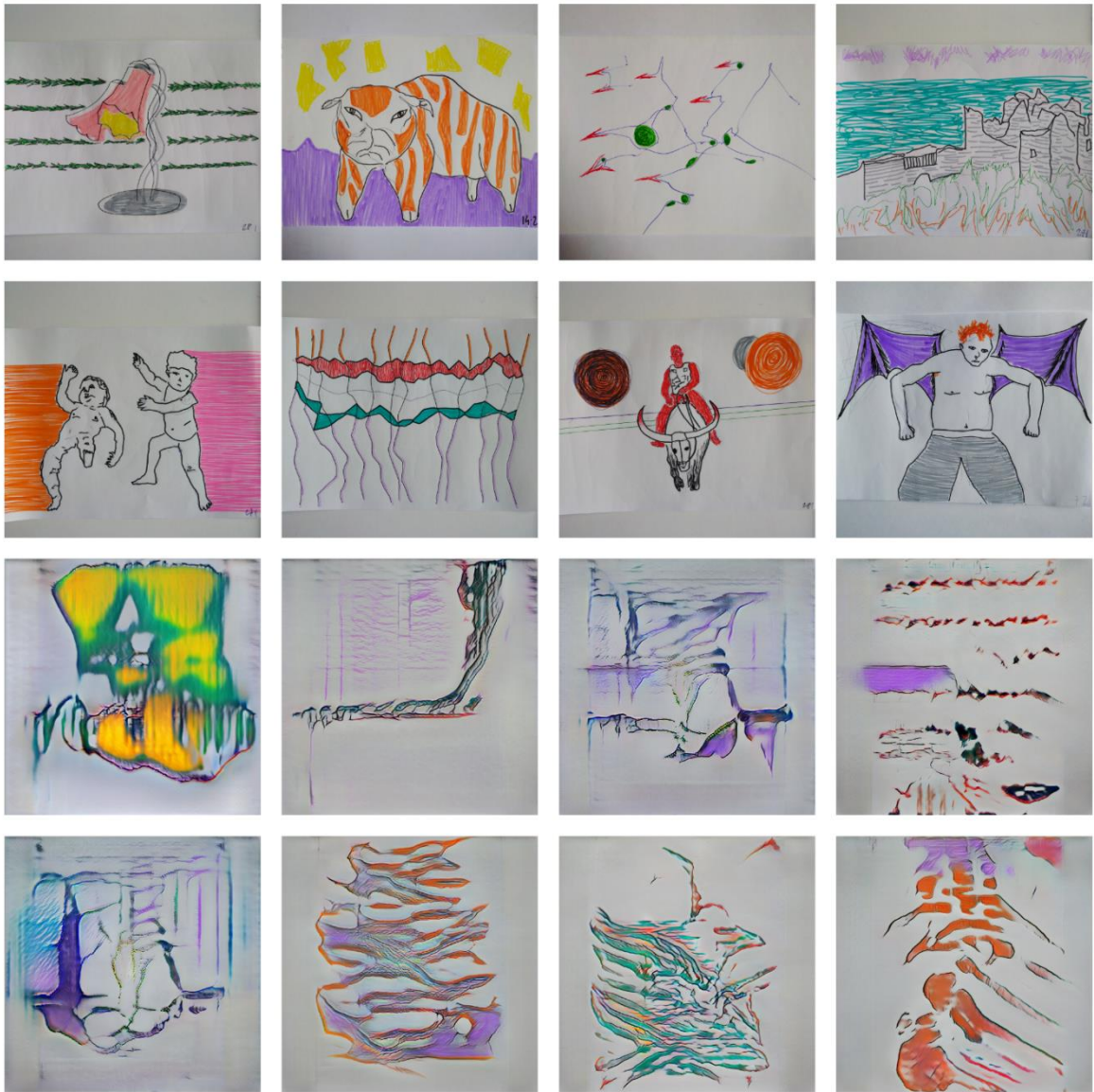


Image 3(b).



Image 3(c).

4. reflection

4.1 ai and creativity

“Is the result handmade? A lot depends on your attitude towards technology” [9].

Artificial intelligence is certainly a relevant topic now. The advancements that happened in the past ten years have increased exponentially. Especially in the running year, every couple of weeks there has been an incredible number of new applications exploiting AI being released regularly. The impact that these technologies bring is impressive, on multiple fields. The repercussions they will have on both daily life and bigger socio-economic and political systems are sometimes clear and sometimes very uncertain. It is exciting and scary at the same time. We have many examples in History of technological advance proceeding way faster than our collective ability to adapt and mitigate the consequences.

There are for sure many ecosystems that will be deeply shaken by these new technologies. It is fundamental to engage in meticulous and critical debate, in order to generate awareness and let space for intervention, trying to prevent negative repercussions.

The impact of AI on some of these fields requires a faster and stronger reaction and regulation compared to the creative field. Nonetheless, it is also crucial for the latter to have this kind of critical debate, for it to survive and exploit new opportunities, whilst preventing disinformation and misuse.

When new technologies emerge, they can potentially disrupt entire sectors. In the creative field the examples are many.

To better understand the relation between AI and art, a paragon with photography can be useful.

The Daguerreotype was the first camera to be created, in 1839. It was capable of reproducing records of the world. Back then, portraiture was one of the first practices that adopted this new tool, it offered a new economic way to make realistic portraits, even if it was quite long to take a picture and the subject had to tightly grip the chair and fix their head with a brace to have a good result.

Paul Delaroche, a painter, at a demonstration of the daguerreotype in 1839 said “From today, painting is dead”.

If photorealism could be replaced by a mechanical process, what would be the role of the artist? As a reflection of this progress, painting turned towards less realistic art movements like impressionism and other abstractions that photography could not allow. Meanwhile, the pictorialism movement, around the 1880s was trying to establish photography as an art form. Pictorialists gained much more control over the photograph process and approached it as they would approach other fine arts. The pictorialism movement culminated in 1910 in the “Buffalo Show” organized by Alfred Stieglitz, the world’s first photography exhibition. By then photography was established as an art. Coming back to our days, there is a frizzling hype around AI and its ability to generate content of all sources. Many people are enthusiastic and many are scared of the possible impact.



Figure 4: A tweet from Nathan Lands, founder of “lore.com”, one of the biggest product developer company that apply AI to business processes.

I would argue (and hope) that a shift similar to photography will happen.

Each time new technologies are involved, the emphasis is put on their standalone capabilities. It is justified, because each time they overtake certain human skills, or just offer new ways to consume content.

“The medium is the message” is a famous phrase coined by Marshall McLuhan during the ‘60s [10]. It proposes that a medium itself is the object of the study, rather than the message it is conveying.

This concept has been impactful in the studies of communication, but has been critiqued in many ways.

I think this is mostly true when mediums are in their early stages of absorption in society. Both mediums as means to consume content, and as frameworks in which creating content.

The same happened with movies and television: as long as it was new, people would watch it no matter what the subject was.

I hope that after this initial hype and threatful phase of AI, it will smoothly be incorporated into the creative field. Artifacts and artworks that are made only by AI will quickly lose their charm.

Some fields like the movie industry might get deeply affected, but it will probably evolve into something that still relies on human invention and artistic vision, moving away from tasks that AI will be better and better at. Just like painters dived into abstract forms after the introduction of the camera.

Artificial intelligence has the potential to allow the development of many new art forms. It should empower creativity and facilitate repetitive tasks. But the essence shall remain within the human behind it and the unicity of their intent and vision.

People will keep exploring and find novel ways to use AI in their practice. When consuming, analyzing, and appreciating this sort of artworks, the emphasis will shift to the skills and ideas of the human, rather than on all the steps that were automated through the process.

The degree of personal participation, more than any degree of independence from machine technology, influences perceptions of craft in work.

4.2 connection between music and image

If deconstructed to the essence, any image is just a construction of our brain. All we see is light, photons that are either reflected or emitted from objects in the environment.

When this reflection hits the rods and cones in our retinas, this signal is then processed by the brain, on multiple levels. From basic information to high level abstraction.

Similarly, in its essence, sound is just a vibration that travels through a medium. A vibration can be caused by collision, pressure or movement changes, and it causes the medium (liquids, solid or gas, like air) to compress and decompress in the form of waves. This wave, when it reaches our eardrums, is then converted into an electric signal that is processed by the brain, interpreting and reacting to it through an equivalent process as for visual signals.

When looking at the essence of these two physical phenomena, they are fundamentally detached and they do not interact directly.

The ability to process information from the environment is a crucial feature for most living beings. Light-sensitive molecules and mechanosensitive membranes (the most basic vibration receptors) were developed in single-celled organisms more than a billion years ago and evolved into the complex special senses that we have now [11][12]. They turned out to be a fundamental feature defining advanced forms of life, and they are indeed our main tools to decode and process the world, and our survival ability strongly relies on them. However, we do not only use our senses for surviving, but also as holistic tools for many abstract reasons, and the importance of art and music in modern society reflects our yearning to chase what we sensorially perceive as pleasurable.

This is just a lucky side effect of a very basic biological survival tool. We developed these skills out of necessity, and as the brain evolved in capabilities, we let space for the usage of optical and aural channels for the sake of their own pleasure. Just like romantic pleasure is a fortunate side effect of our biological imperative for reproduction.

Sight and hearing are vitally entangled to the human experience, and many people devote their life to their hedonistic fulfillment.

According to these premises, the only meeting point, or connection, is the human being.

**Leaving other life forms apart for simplicity and linearity of the argument*.*

Every human perceives sight and sound differently, and there is no direct way to understand what is the psychological process that others go through. We can only compare external knowledge with personal experience. Otherwise, a “scientific” way to pursue that could be analyzing electroencephalogram data and comparing brain activity. On the other hand, a different approach can be through expression and storytelling, which are very ancient social mechanisms to convey complex knowledge to other people, when it is not possible to verbalize or quantize it. Looking at the available literature, this phenomenon has certainly been addressed more artistically than in a scientific way.

There are many examples in art history expressing how these two domains merge, or how they prolifically served each other as a lead or inspiration. An example can be Wassily Kandinsky, who got inspired by classical music while making pioneering massive abstract paintings. Or Jean-Michel Basquiat, who got inspired by jazz and punk, incorporating musical symbols and other elements to works that often blurred the line between the two mediums.

Just like a musical piece can be composed as a sound translation of a certain landscape or a defined set of aesthetics, when inspecting the interconnectedness of visual and auditory elements, the richest examples express the visualization of music. Audiovisual artists have embarked all sorts of rules to portray music into shapes, colors and volumes, resulting in a dynamic abstraction that represents sound waves in another form, consumable visually.

Especially in the digital era, there are many signal processing tools to analyze sound and transform it into a different representation of itself. Some sort of reverse engineering of the musical notation. This symbolic representation of sound can be associated with visual elements in order to create moving or static iconographies.

A similar method can be achieved biologically: synesthesia is a perceptual phenomenon according to which some people are involuntarily experiencing a sensorial throughput in a secondary sensorial pathway. For example certain sounds can evoke colors, figures and patterns in someone's mind [13].

A parallel between music and sound can also be defined by cultural and historical context: art movements, as well as trends and social issues, can shape both musical and visual expression, tightening them together into a unified form, reflecting the context in which it exists.

Most probably, if you try to close your eyes and visualize an Opera or a Hip hop piece, they will be somehow bound to the elements of their milieus.

A semiotic approach can help to understand and bring these two realms together. Performances, music videos, album covers and fashion can indirectly contribute to establish identity of aesthetic and music genres, because they carry a meaning that will

be related to their themes, lyrics, etc. This can be seen as a subconscious high-level interpretation of the Synchresis phenomenon, that happens when we associate a sound and an image even if they do not correspond to each other, but happen at the same time [14].

According to my experience, throughout the process, music has been a great source of inspiration. It has been present since the beginning, because all the drawings I produced were made while listening to music, trying to convey on paper my reaction to sound. This means that the visual material I generated is somehow, partially, following the lead of music.

I usually try to develop a structure or strategy to lead me through the act of creation. An idea to follow, or a concept to lead me. Having a plan takes insecurity away, it is something to hang onto when I don't know how to proceed.

However, with the help of music, when I am there, in the moment, the plan gets lost. The emotional involvement in the act is stronger than the plan, and this is where the magic happens.

The rational plan is still there, and I can try to follow it as many times as I wish. But any emotion is unique, contextualized in space and time, not replicable. This is the reason why I value something that was created in this state of "rush", I value its uniqueness and awkwardness. I try to honor my errors as hidden intentions, as something that would not have come to life if not because I deviated from the plan.

The connection between image and music, for me, is all of the above, or maybe it is none. But I recognize it as most precious when it is characterized by this emotional singularity.

The idea of things being exact and objective has proven to be delusional. This mentality has fueled positivist rationality and scientific knowledge, characterizing it until the nineteenth century. It was science itself to prove it wrong: many discoveries, from quantum mechanics to Einstein's relativity, have shown that the subject (the observer) has an active role in shaping reality while investigating it, it only exist through the lens of

consciousness. This validates the personal point of view, making it impossible to detach from it and make objective statements.

I feel extremely privileged to have been able to spend a number of months, throughout this project, exploring what The Connection means to me. It acquired multiple layers of interconnectedness, merging on different phases and modalities, starting from the act of drawing, and culminating with the meta-analysis of the final images. My perception of it has changed and grown, being immersed in it was a great source of joy and inspiration. Once again, where words fail short to describe the depth and strength of emotional states, expressing it in other forms goes a little step further.

The drawings that I manually made reflect this moment of emotional connection. When I look at them, I have a sort of “paternal” feeling, because I have memories of the act of drawing them and recognize the process as something that directly came out of me. In a way, I identify with these drawings.

Differently, while looking at the images that I made with the help of AI, a strange feeling of depersonalization occurs. I think that these images are very similar to my own drawings, at least in some of their elements. These elements trigger the same paternal feeling, or identity association, however, this generates a sort of emotional gap because I have no physical memory of making them. It is almost paradoxical to see something so similar to what I did, but machine made. It allows me to look at my work with a different lens, taking the ego away, which makes you relate to your own work distortedly.

AI allows us to put a mirror on our own capabilities. It is able to automate certain tasks that we want it to do and perform them better (or at least learns how to because it is way faster). This allows us to make a very deep analysis of ourselves and those capabilities.

4.3 a monkey in a spaceship

Being a monkey in a spaceship means to be connected to one’s inner self. To embrace thy most controversial essence in this furious, chaotic, drifting world.

The monkey, ironically, is more human than the sapiens-sapiens, mostly relying on its instinct, experiential knowledge, perception, and emotions. It embodies the most fundamental essence of human beings, who have lost their primordial touch, they have passively learnt to suppress it and replace it with rationality. They have lost themselves in shiny material illusions, square angles, precise predictions and sequences of bytes. It is partially by choice that I happened to be a monkey in this metaphorical spaceship. I have tried to be the qualified sapiens conqueror, knowing all the instructions and having clear goals. Respecting all the rules and following the protocol. But my journey failed short.

The world pushes everyone to be a rational, a sapiens. It pushed me. This was not only oppressing me, but it is literally bringing humanity on the verge of mass extinction. This oppression indirectly pushed me to embrace my ancestral essence. I started to take a turn and ignore the ship's protocol and its precise instructions. I have seen that being a monkey brings me further on most of my explorations, bringing me to lands which existence I did not know, besides making the whole journey more exciting and enjoyable.

Defining the monkey can hopefully help to understand the attitude that characterized this project. I tried to stretch to the limit some behavioral patterns that I found healthy and helpful in my anthropological experience.

The spaceship represents technology, and its role as a medium through which one can explore unforeseen galaxies of future developments. A vessel through which we now communicate, work, pay, learn and grow into.

Technology, intended as an infrastructure that enables global interconnectedness, is arguably the most complex and fascinating between all human creations. The ability to process and exchange information at light speed has deeply revolutionized society and I have witnessed immense transformations throughout my short lifetime.

I believe that most of the change brought by technology does not facilitate happiness. The reason is just because of its misuse, led by greed and individualism. We live in the posthumanism era, technology is the main factor threatening our human essence.

Artificial Intelligence seems to be a golden goose for many of our contemporary challenges, the potential impact is humongous. AI can contribute to achieve 79% of the 169 targets underpinning the Sustainable Development Goals set by the United Nations [15].

I think it is useful to speculate and build frameworks that promote alternative approaches to AI. To not just use it as a sharp tool to optimize our demands, but rather use AI to empower the volatile beauty of the human condition.

In this project, with a relatively small amount of technical skills to set up the conversation, I was able to cooperate with AI in the most simple, instinctive way, and I am grateful for how it allowed me to discover, reflect and grow.

I hope that, not only in the creative field, there will be a thriving community that uses AI in the most humane way, aiming to explore the realm of expression, teaching smarter and smarter technology how to capture and empower our essence.

Being a monkey in a spaceship is a rebellious act. I revolt against the common idea and usage of technology, that is built upon clear causalities and precise calculus. We have now reached a level of technological complexity that allows abstraction, and artificial intelligence has the potential to be the launch pad of a new era of knowledge. This is the time to uplift the role of technology into something that can truly empower human beings.

bibliography:

- [1] Ellis, C. et al. 2011, Autoethnography: an overview. FQS, Forum Qualitative Social Research.
- [2] Debord, G. 1956. Translated by Ken Knabb. "Theory of the Dérive". Les Lèvres Nues
- [3] OpenAI, 2023. DALL.E, assessed 05/05/2023, <<https://github.com/lucidrains/DALLE2-pytorch>>
- [4] OpenAI, 2023. DALL.E training process, assessed 05/05/2023, <<https://openai.com/research/dall-e-2-pre-training-mitigations>>
- [5] Goodfellow, I. et al., 2014. Generative adversarial nets. In Advances in neural information processing systems. pp. 2672–2680.
- [6] Liu, B. et al. 2021. "Towards faster and stabilized GAN training for high-fidelity few-shot image synthesis." International Conference on Learning Representations.
- [7] Unreal Person, 2023, assessed 28/4/2023, <<https://www.unrealperson.com/>>
- [8] Yu, Y. et al. 2021."Frechet Inception Distance (FID) for Evaluating GANs." China University of Mining Technology Beijing Graduate School.
- [9] McCullough, M. 1996. Abstracting crafts. MIT press
- [10] McLuhan, M. 2001. The medium is the message. Gingko Press.
- [11] Plachetzki, D. C., Degnan, B. M., & Oakley, T. H. (2007). The origins of novel protein interactions during animal opsin evolution. PLoS One, 2(10), e1054.
<https://doi.org/10.1371/journal.pone.0001054>
- [12] Kung, C., Martinac, B., & Sukharev, S. (2010). Mechanosensitive channels in microbes. Annual Review of Microbiology, 64, 313-329.
<https://doi.org/10.1146/annurev.micro.112408.134106>
- [13] Cytowic RE (2002). Synesthesia: A Union of the Senses (2nd ed.). Cambridge, Massachusetts: MIT Press. ISBN 978-0-262-03296-4. OCLC 49395033
- [14] Chion, M. (1994). Audio-Vision: Sound on Screen. Columbia University Press.

[15] Vinuesa, R., Azizpour, H., Leite, I. et al. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat Commun* **11**, 233 (2020).

<https://doi.org/10.1038/s41467-019-14108-y>

[16] Scoates, C., 2013. Brian Eno: Visual Music. Chronicle books.

[17] Gladwell, M. 2005. Blink: The power of thinking without thinking. Little, Brown and Co.