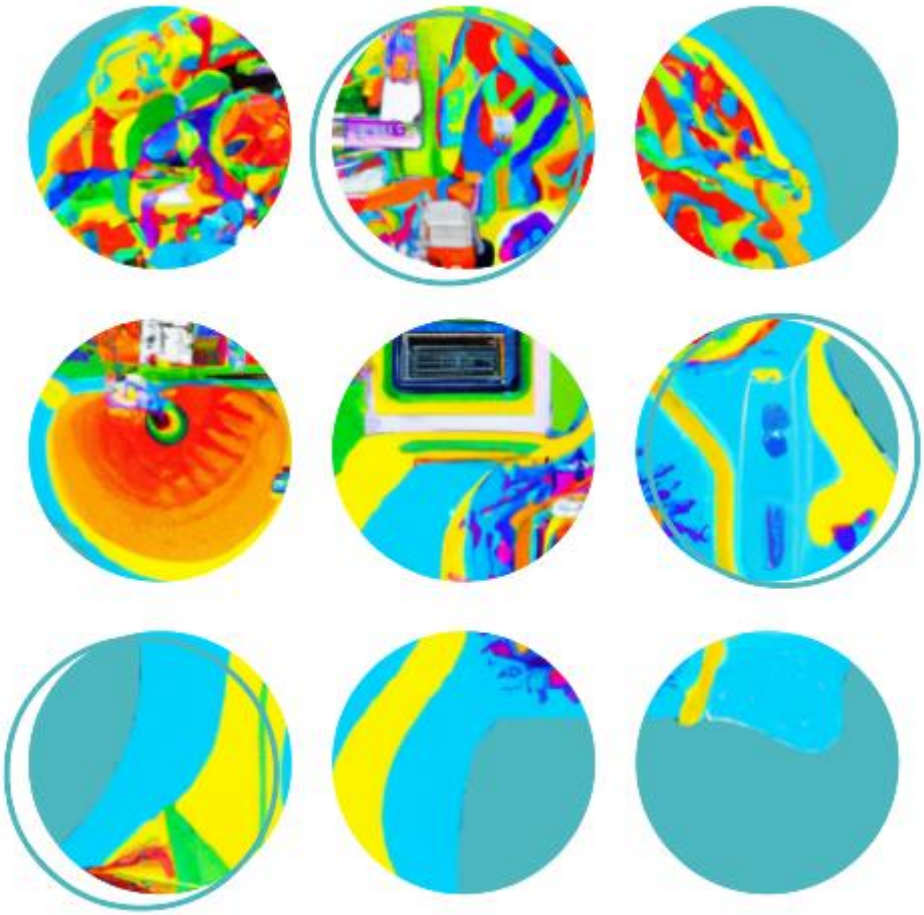


AI and Cultural Heritage

Between Research and Creativity

Workshop Proceedings

Edited by Antonella Guidazzoli and Maria Chiara Liguori

CINECA

AI and Cultural Heritage
Between Research and Creativity
Workshop proceedings

Edited by Antonella Guidazzoli and Maria Chiara Liguori

Foreword by Francesco Ubertini

Cineca

AI and Cultural Heritage. Between Research and Creativity

Workshop proceedings

Edited by Antonella Guidazzoli and Maria Chiara Liguori

www.cineca.it

visitlab@cineca.it

ISBN – 9788886037457

DOI: 10.1388/WORKSHOP-AICH-01

SUMMARY

FOREWORD

Francesco Ubertini – President of Cineca

Cineca and Artificial Intelligence, a pairing that can be considered natural. Indeed, the training of neural networks finds its ideal partner in high-performance computing. When we presented the Leonardo pre-exascale supercomputer to the public and to President of the Republic Sergio Mattarella at the Bologna Technopole at the end of November 2022, we also talked about the opportunities that this machine will offer to research and innovation in the field of Big Data and Artificial Intelligence.

A link between supercomputing and cultural heritage may seem unusual, but Cineca's support to this field dates to the late 1980s, thanks to the activities of the VisIT Lab, dedicated to scientific visualisation. The laboratory's mission is to support the research community with advanced data visualisation tools, an activity that is not limited to the simple return of information obtained from research results, but opens up to communication and, above all, to synthesis, interpretation and support for interpretation. Archaeology was the first branch in Cultural Heritage with which cooperations were established. One of the most significant projects was the Casa del Centenario in Pompeii, carried out by the University of Bologna in the early 2000s, with which I myself, as an engineering researcher at the time, found myself collaborating due to the modern and multidisciplinary nature of the project.

Just as VisIT Lab acted as a link with the Cultural Heritage for Cineca at the time, it later enabled the collaboration with projects with a strong artistic component. Cineca, thanks to its location at the heart of research and its computing resources that act as a catalyst, also finds a valid position in such contexts, all the more so when artificial intelligence, with its ubiquity, is also spreading in these fields.

The organisation of the workshop on 20 January 2023, dedicated to a reflection on the intersection between artificial intelligence, cultural heritage and art, in the light of the VisIT Lab's activities over time, is therefore an almost inevitable landing place.

The positive reception of the meeting by the public has led us to publish not only the recordings of the workshop, but also the proceedings, and to envisage other future events, also linked to the Researchers' Night.

INTRODUCTION - THE GREATEST ENEMY OF KNOWLEDGE

Pierdomenico Memeo – Science writer, outreach expert

I was frightened.

"Robots don't feel fear. They don't feel anything.
They don't get hungry, they don't sleep–"

I do. I have even had dreams.

"Human beings have dreams. Even dogs have dreams, but not you.
You are just a machine. An imitation of life.
Can a robot write a symphony? Can a robot turn a canvas into a masterpiece?".

Well, can you?

Jeff Vintar and Akiva Goldsman, *"I, Robot"* (screenplay)

There is a subtle thrill in not knowing something. The feeling that there is a vast landscape of ideas, concepts, information, relationships that we are just about to enter. It is the sound of the wind before the journey, the gleaming of the waves before the dive, the smell of a new book before reading the first line. We leave the comfort of the known and jump into something else.

We have grown accustomed to attaching a negative meaning to the lack of knowledge, and this is a natural consequence of our knowledge-based society. If knowledge is power, then being ignorant is being impotent. But just as a cup is useful when it is empty, ignorance has value because it can be filled with knowledge. And only by acknowledging that we do not know something, we unlock the potential to learn, to explore, and to grow.

These Proceedings are full of knowledge, built over years of study and work by enterprising people who trail-blazed the undiscovered country of a future unravelled. But we will be able to fill up our cups only if we recognize our own ignorance, feel the thrill, and jump.

****

Only a few years ago, **Artificial Intelligence** was nothing but a **buzzword** for most people. It had to do with "intelligent computers" and was more science fiction than anything else. The general public was vaguely aware of advances in specific fields of artificial intelligence, such as the all-too-famous **chess matches** against various human Grandmasters.

But away from the spotlight, something truly revolutionary was brewing, blooming from two converging factors: on the one hand, systems such as **neural networks**, **fuzzy control systems**, and tools for **mathematical optimization**; on the other, the tremendous amount of data created and accumulated by the progressive expansion of digital systems. The former provided the seeds for this new paradigm; the latter, the soil in which it could grow.

Slowly, quietly, these new tools crept into many fields, but they were **hardly recognized as AI**. Then, in the 2010s, faster processors, algorithmic improvements, expanding cloud-computing infrastructures and ever-increasing datasets enabled advances in machine learning, particularly in deep learning methods. Much-sought after, but elusive fields such computer vision, object recognition, image generation and language models were finally within reach.

All of a sudden, around the start of the next decade, **artificial intelligence was everywhere**. In fact, it had been growing for the last 30 years, but the realisation came as a shock to many people. What was an obscure, hermetic, even cryptic field of study instantly became topical.

Public discussion about the workings and ethics of artificial intelligence became commonplace and much effort was poured into public awareness of the use, limitations, and biases of artificial intelligence in web search engines, social media algorithms, job recruitment, medical insurances, bank loans, and many others. **CINECA**, with its public outreach programs, already highlighted some of these topics, for example during the **2021 European Night of the Researchers**, during the online workshop *I Sing the Body Electric: AI and Gender Bias* with Anna Elisabetta Ziri, Senior Manager at PwC and AI expert, and Antonella Guidazzoli, Headperson of CINECA VisIT Lab.

In **2022**, two major public breakthroughs have enticed, and oftentimes worried, the public:

In the second half of the year, three major platform for text-to-image generation were unveiled, to the delight and sometimes concern of the public: **Midjourney** (Midjourney Inc, July 2022), **DALL·E 2** (OpenAI, September 2022), and **Stable Diffusion** (Stability AI, December 2022). Although their workings and business strategies were vastly different, all three contributed to highlight the potential of text-to-image generation and were widely commented in social media and discussion forums.

In November 2022, **ChatGPT** was released: a chatbot developed by OpenAI using their own pre-trained generative transformer language models based on deep learning techniques. While chatbots were not a novelty, ChatGPT gathered much attention for its precise understanding of

questions, detailed responses, and articulate answers across many knowledge domains. At the same time, it very uneven factual accuracy, and sometimes completely made-up replies, were a source of deep concern.

This string of events moved the discussion on artificial intelligence from the depth of academia and industry to the spotlight of public, even political, discourse. Although undoubtedly necessary, this quick change of scene means that we must be **very, very careful with our attention**, because it would be all too easy to fall into flimsy misconceptions and inconsistent simplifications. As historian and educator Daniel J. Boorstin is credited to have said: **The greatest enemy of knowledge is not ignorance - but the illusion of knowledge**.

The aim of these Proceedings is therefore to take a moment to see where we were, where we are, and where we are going. We do this with the help of **academics, computer scientists, software developers, cultural entrepreneurs, communication experts and digital artists**. We approach this topic with an open mind, trying to build a common ground between different expertise.

We have gathered the contributions of professors, researchers, critics, and artists, and gave them free rein to take us into their own works, with the purpose of coming through the other side with a new perspective.

We cannot thank them enough: in the next pages you will be able to deepen the concepts presented during the workshop.

And of course, I would be amiss if I did not thank Antonella Guidazzoli, Headperson of CINECA VisIT Lab, and Maria Chiara Liguori, digital historian, without whom nothing of this would have happened, for giving me the opportunity to host the workshop and write these few forewords.

Now, with me– jump.

THE INTEGRATION OF AI IN CULTURAL HERITAGE AND ARTISTIC AND CREATIVE EXPERIENCES: INSIGHTS FROM CINECA VISIT LAB

Antonella Guidazzoli, Maria Chiara Liguori – Cineca

Over the last twenty years, Cineca has gained significant experience in the application of information technologies to scientific research and cultural heritage, using high-performance computing (HPC) and advanced graphics tools. In this environment, VisIT Lab, which stands for Visual Information LAB, has endeavoured to remain at the forefront of creating visualisation solutions to build new bridges between research, communication and art.

In recent years, artificial intelligence has emerged as a disruptive force and can undoubtedly be considered one of the most sophisticated tools that technology has to offer today. So sophisticated that it is likely to form the basis of the next great industrial revolution. Its potential is far from being realised, but it is beginning to spread faster and faster, in an exponential progression that follows the law of accelerating returns hypothesis of Raymond Kurzweil [1]. Its use is becoming more pervasive, and soon there will be no sector of human activity that will remain excluded. In the course of 2022, AI has become more and more public, with the popularisation of GANs in applications such as Midjourney, DALL-E or Stable Diffusion, and natural language processing tools such as the all the rage ChatGPT. There are now tools like Wonder Studio that automatically animate, light and compose CG characters in live-action scenes, and Google and Microsoft are integrating AI into all their products. At the same time, at Cineca's VisIT Lab, more and more European projects that ask for our participation want to integrate AI to achieve their goals, such as the creation of real-time avatars with realistic renderings for immersive collaborative environments. After all, supercomputers are the ideal environment for training neural networks and processing their results.

As a lab, the first time we really became aware of the level that AI had already reached was at Siggraph Los Angeles 2017. As an event dedicated to computer graphics and all things that merge advanced technologies and graphics, it could not have been a better place for such a revelation. At the Nvidia booth, a screen connected to a webcam showed passers-by their real-time morphing into the style of one of the painters on whom the AI running the application had been trained (Fig. 1).

Fig.1: VisIT Lab at Siggraph 2017 interacting with AI generated images.

A deeper change in perspective came for us with the I-Media-Cities project, in which the Lab took part from 2019 (EU Horizon 2020 - imediacities.eu). I-Media-Cities is the result of an ambitious and innovative research project funded by the European Union. It was a collaboration between 9 EU film archives, 6 research institutions, 2 large digital expert centres and a business model expert to develop a new platform providing access to digital films and photographs of European cities. The goals of the project were to discover new approaches to research on digital audio/video content, to improve the overall accessibility of European cultural heritage, and to stimulate collaboration between archives and researchers. In order to achieve these goals, the a/v content was processed through a pipeline using different types of algorithms capable, for example, of selecting the different shots within videos, to perform the task of object detection and automatically enriching the shots with the retrieved information. At the same time, the researchers connected to each film archive could manually add more tags [2; 3].

If the first experience presented AI to us mainly as a surprising amusement, giving us a glimpse of the enormous potential of this tool, with I-Media-Cities it was possible to touch the effects of this potential, and it was only a first taste. As we have said, the transformations are increasingly rapid and profound, even in these two areas, Cultural Heritage and artistic and creative experiences, to which we feel particularly close and which we have chosen to make the focus of this workshop. However, the profound transformations taking place in these two contexts are having different but converging consequences.

As in the case of the I-Media-Cities project, Cultural Heritage benefits most from AI as a valuable support tool, capable of automating and accelerating repetitive processes by applying it to large amounts of data. Think of the IDEA project and the Isabella d'Este Virtual Studiolo, a project that VisIT Lab has been following since 2016 [4]. Part of this project is an archive with the digitisation of the epistolary of the Marchioness of Mantua, kept at the State Archives of Mantua. In this case, the several thousand letters belonging to this archive could be processed with a text recognition tool suitable for handwritten content, followed by the application of sentiment analysis algorithms.

In the artistic and creative field, however, the effects have been much more problematic and are often perceived as a threat, sparking heated debates that eventually expand to include and question what actually characterises what is truly human. Visual artists have been the first to raise their shields, and in early 2023 a class action lawsuit was filed against some applications that use generative AI, accusing them of using datasets that collect copyrighted images without paying due compensation and at the same time jeopardising the artists' own careers [5]. On the other hand, those who have embraced these new technologies emphasise their nature as mere tools: at least so far, artificial intelligence lacks artistic intentionality. Art is a combination of technical realisation and intentionality. AI helps with the technical realisation part, but it is the artist who drives this possibility with his intentionality. Some critics provocatively ask whether today's artists are capable of providing this to the tool, given the results they have produced in recent years [6]. However, according to other artists, such as the musician Peter Gabriel, many, but not enough, are coming to grips with AI technologies; and it is good that they are doing so, and quickly, because given the speed of AI developments, it is better to try to work with it than to be overwhelmed by [7]. As early as 1972, the science fiction writer Philip K. Dick reminded us that it will always be good to have a human behind the wheel of the technological machine [8].

It is therefore inevitable that art and artists are increasingly considered important by the scientific community, not only to try to better communicate discoveries to a wider audience, but even to better understand the data on which scientific projects operate and to obtain new and unprecedented interpretative perspectives. And it is the artist who is asked to collaborate with AI and use it as the tool that it is, and it is the artist who is therefore asked to sit in the driver's seat of the technological machine.

In such a context, we realised that, given the relevance of the issue, it was not possible not to participate in the collective reflection and that it was even important to try to promote spaces of confrontation ourselves. From all this came the desire to organise a workshop. What we did on

January 20, 2023, is only the first of a series of meetings that we would like to organise in the future, again within the framework of the Night of the Researchers, given the incredible evolutionary speed of AI.

References

[1] R. Kurzweil, *The Age of Spiritual Machines*, 1999.

[2] Caraceni, S. *et al*., "I-media-cities, a searchable platform on moving images with automatic and manual annotations," *2017 23rd International Conference on Virtual System & Multimedia (VSMM)*, Dublin, Ireland, 2017, pp. 1-8, doi: 10.1109/VSMM.2017.8346274.

[3] Teresa-M. Sala, Mariona Bruzzo (eds.), *I-Media-Cities Innovative e-Environment for Research on Cities and the Media*, 2019.

[4] Shemek, D. et al., Renaissance Remix. Isabella d'Este: Virtual Studiolo, in DHQ, Vol. 12, N. 4, 2018, http://www.digitalhumanities.org/dhq/vol/12/4/000400/000400.html

[5] Noone, G., Will 2023 be the year of the AI lawsuit? February 2, 2023, https://techmonitor.ai/technology/ai-and-automation/will-2023-be-year-of-ai-lawsuit (checked on March 23, 2023).

[6] Caliandro, C., L'Intelligenza Artificiale sa fare meglio degli artisti?, in Artribune, 14 febbraio 2023, (checked on March 23, 2023),

https://www.artribune.com/progettazione/new-media/2023/02/intelligenza-artificiale-androidi

[7] Parker, L., Peter Gabriel on the future of AI: 'We might as well just grab the algorithms and dance with them, rather than fight them', March 14, 2023, (checked on March 23, 2023),

https://www.yahoo.com/entertainment/peter-gabriel-on-the-future-of-ai-we-might-as-well-just-grab-the-algorithms-and-dance-with-them-rather-than-fight-them-203833247.html

[8] Dick. Ph.K., *L'androide e l'umano*, ne La ragazza dai capelli scuri, a cura di Carlo Pagetti, Fanucci Editore, Roma 2014.

AI FOR ART: A LOOK AT OBJECT DETECTION AND TRANSFORMERS

Giorgio Pedrazzi, Donatella Sforzini, Gabriele Fatigati – Cineca

Abstract

In the field of the visual arts, the integration of artificial intelligence (AI) has garnered significant attention in recent times. Specifically, the application of AI for the classification and generation of artwork has become increasingly popular. We delve into the cutting-edge utilization of object detection and zero-shot detection models, which have the impressive capability of identifying and categorizing objects without the need for prior training. With the aid of high-performance computing, the processing of thousands of images can be accomplished at a lightning-fast pace. By exploring examples, we demonstrate the vast potential of these advanced technologies in the field of art.

State of the art AI–Digital Humanities

As highlighted in a recent review of the state of the art in the digital humanities by She and Cetinic [1], AI technologies are having a strong impact on research and creative practice in the visual arts. Two aspects of the use of AI in this field emerge:

- AI is used for the automatic cataloguing and annotation of digitised artwork collections;
- AI is used for creative purposes and to generate new works of art.

Our focus is on the first point (cataloguing and annotation). Three main methods of interest have therefore been identified:

- Classification
- Object Detection
- Zero-shot detection

In classification, the task is to assign an artwork to a specific category based on predefined criteria, such as author or style. Models are trained on a pre-classified dataset to learn how to classify new artworks. In Object Detection, the goal is to identify and locate objects within the artwork, such as, for example, people, animals, or religious symbols. Models are trained on annotated images to learn how to detect and localize objects accurately.

Classification and Object Detection (OD) require the provision of annotated examples consistent with the purpose of the analysis. There is a list of datasets that can be used in the different tasks [1; 2]. For Object Detection, an annotated dataset of the European project 'Saint George on a Bike' SGoaB has been used for training a model.

Zero-shot detection (ZSD), on the other hand, is a type of machine learning task that involves detecting an object or concept in an image without any prior training. Using advanced language models, transformers can analyse images and understand their contents with a high degree of accuracy. This technology has taken the world by storm with its incredible ability to detect and classify objects.

Object Detection and Zero-shot detection proved to be the most relevant techniques in our work considering that they also include the possibility of using their results for subsequent classification.

Overview of tested tools

We have identified several reference tools for the various tasks mentioned above. Here is a brief description of each:

- Detectron2 (OD) enables the detection of objects in new images based on pre-annotated examples, which are not easily obtainable in an artistic context.
- CLIP (ZSD) allows for the assignment of a probability that a particular class is present in an image, given a vector of arbitrary classes. This multi-modal model is based on 400 million images, each of which is associated with its own description. The limitation of this model is that it does not assign probabilities to individual detected objects but instead considers the entire image [3].
- OWL-ViT (ZSOD) is an extension of CLIP that can determine the bounding box for each detected object and its corresponding class. This tool overcomes the limitations of previous systems by enabling the detection of the same object multiple times in an image and assigning a probability of belonging to each class included in the query vector [4].
- Other tools such as RegionCLIP and ViLD (ZSOD) have been tested as alternatives to OWL-ViT, but they have presented some technical problems.
- LabelIMG is a manual annotation tool that allows for the creation or editing of annotations for an image.

Object detection

We used the dataset published by SGoaB consisting of 3946 images (paintings and some photos) manually annotated coming from 8 image source databases (COCO, Europeana Collection, IconClass, Pharos, WikiArt, Wikimedia Commons plus 2 other non-specific sources named "None" and "Unknown" in which 50 classes/object were noted).

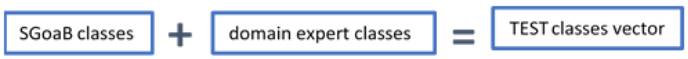
First, we performed a pre-processing phase in which we did data analysis and data preparation of dataset to be passed to Detectron2, CLIP and OWL-ViT. We also created procedures to make the three systems talk to each other and be able to compare the available meta-information with the results. Then, in the second phase we performed the training of an Object Detection model.

Due to the high heterogeneity of the classes predefined by the project, some are very general (crucifixion, landscape...) and contrasted with others very specific (zucchetto, trumpet, sheep...) and the low number of annotated images the model obtained was not good enough and the accuracy is still not satisfactory. However, the alternative of creating a new dataset with a different class is very time consuming as it requires manual annotation.

For this reason, we started looking for AI models that make an automatic annotation without class limits.

Zero-shot Detection: annotation examples with CLIP and OWL-ViT

During the experimentation we carried out a Zero-shot Detection using two tools: CLIP (Contrastive Language-Image Pre-training) and OWL-ViT (open-world object detection with vision transformers). As already mentioned before, the second is an extension of the first.




```
text_vector =
["angel","apple","arrow","banana","banner","bird","boat","book","camauro","cat","chalice","crown","crown of
horns","crozier","deer","devil","dog","dove","dragon","eagle","elephant","horn","horse","key of heaven",
"knight","lance","lily","lion","mitre","monk","nude","orange", "palm", "person", "prayer","rooster",
"saturno","scroll","serpent","sheep","shepherd","skull","stole","sword", "tiara ","tree", "trumpet", "zebra",
"zucchetto", "Christ", "Virgin Mary", "saints", "evangelist", "bow", "armour", "halo", "the baby Jesus", "hat",
"clamshell", "cross", "flower", "fruit", "jewel", "lute", "mantle", "cow", "cloud", "landscape", "shield", "snake",
"mirror", "violin", "crucifixion", "hand", "head"]
```

Fig. 1 – Text vector used for the CLIP and OWL-Vit annotations.

In this phase, to the set of classes derived from the annotation of the SGoaB dataset, we added further terms identified during periodic comparisons with a team of domain experts, the digital humanists of the MIC (Italian Ministry of Culture) for a total of 75 terms (Fig. 1).

Obviously, this is only one of the possible vectors that can be constructed and used for the annotation and the same images can be analysed using different more or less numerous vectors.

Theoretically we could construct a set *n* of vectors, each one specialized on a specific aspect of interest and execute a workflow of classifications in sequence to obtain *n* partial classifications. Then this *n* partial classifications could be merged into a unique macro-classification of the image.

Our idea of future work is to build vectors containing the IconClass classes, an iconographic classification system, exploiting the hierarchy of the system to stop at a specific level of descriptive detail [5, 6].

According to the IconClass classification, each code is represented as a sequence of numbers and letters which correspond to the descriptions of the classified objects. These descriptions become more and more detailed as the alphanumeric string increases. Let's see an example with the IconClass code 73 D 35.

The first number, 7, identifies that the object we describe is concerning the Bible. If we look at the first two digits, 73, then we have a more detailed description and we know it is concerning New Testament and so forth. In particular, the hierarchy would be as follows:

7 Bible

73 New Testament

73 D passion of Christ

73 D 3 capture and torture of Christ

73 D 35 tortures of Christ

During the experimentation we also acted on another parameter: the probability threshold.

Both CLIP and OWL-ViT provide as output the name of the identified object and the probability value with which the IA algorithm has identified it. This allows us to consider acceptable all the automatic annotations generated by the AI with a probability value higher than a certain threshold. The following examples show some of the characteristics of the experimentation described above.

Fig. 2 The image we use is a photo of "The Crucifixion, by Cranach, Lucas, the Elder (1472-1553). Oil on wood" A: Original image with boxes derived from SGoaB annotation; B: image with OWL-ViT mod google/owlvit-base-patch32 annotation; C: image with OWL-ViT mod google/owlvit-large-patch14 annotation.

Example 1: Since in the public dataset of SGoaB each image corresponds to an xml file containing, for each object, the coordinates of the bounding box, in the first example we compared the reconstructed boxes of SGoaB (Figure 2A) with the results obtained from the elaboration with CLIP (in two versions, v1 and v2) and with OWL-ViT (2 versions). CLIP v1 annotation differs from CLIP v2 annotation because the vector used is different. In CLIP v2 we added "crucifixion", "hand", "head" and the IA algorithm found "crucifixion" with a confidence of 94.2% (Table 1).

CLIP v1 annotation		**CLIP v2** annotation	
virgin_mary	29.8	crucifixion	94.2
cross	15.9	virgin_mary	1.7
crown_of_thorns	13.6	cross	0.9
key_of_heaven	7.9	crown_of_thorns	0.8
the_baby_Jesus	6.6	key_of_heaven	0.5
Total	73.9	***Total***	98.1

Table 1: Clip v1 annotation vs Clip v2 annotation

To test the two models of OWL-ViT we processed the same image using the same vector (Table 2). We chose large-patch14 because it annotates more objects and the same objects with greater confidence (even if it does not detect "crucifixion").

OWL-ViT mod **google/owlvit-base-patch32** (figure 2B)		**OWL-ViT** mod **google/owlvit-large-patch14** (figure 2C)	
crucifixion	10	Christ	25
Christ	12	cloud	26
Christ	15	cloud	32
tree	8	crown	28
person	9	hand	25
monk	8	hand	28
person	10	head	30
		head	32
		head	26
		person	35
		person	33
		stole	26
		tree	32

Table 2: OWL-ViT models google/owlvit-base-patch32 vs google/owlvit-large-patch14

To evaluate the overall performance, we considered the annotations of SGoaB as our ground truth and compared it with the results of Owlvit-large-patch14 with threshold 0.15 (see Table 3). Owlvit-large-patch14 correctly identified about 60% of the objects manually annotated in SGoaB.

		Annotazione SGoaB		Annotazione Owl-ViT	
filename	name/object	name/object	Num. Name/object	name/object	Num. Name/object
00000046.jpg	*Christ*			Christ	1
00000046.jpg	*arrow*			arrow	1
00000046.jpg	*cloud*			cloud	10
00000046.jpg	*crown*			crown	2
00000046.jpg	*crown of thorns*	crown of thorns	1		
00000046.jpg	*crucifixion*	crucifixion	2		
00000046.jpg	*deer*			deer	1
00000046.jpg	*hand*			hand	5
00000046.jpg	*hat*			hat	1
00000046.jpg	*head*			head	6
00000046.jpg	*landscape*			landscape	5
00000046.jpg	*monk*			monk	2
00000046.jpg	*nude*			nude	2
00000046.jpg	*palm*			palm	1
00000046.jpg	*person*	person	2	person	2
00000046.jpg	*stole*			stole	6
00000046.jpg	*tiara*			tiara	1
00000046.jpg	*tree*	tree	1	tree	5
00000046.jpg	*zucchetto*			zucchetto	1

Table 3: SGoaB annotation vs OWL-Vit google/owlvit-large-patch14 annotation.

This means that the algorithm could be a useful tool to support manual annotators being able to suggest the objects to be catalogued.

Example 2: we tested the model using a language other than the training language for the vector of terms (specifically Italian) obtaining a decrease in performance (Figure 3).

Example 3: we tested the model on colour and grayscale images (thinking the lack of colour was a limitation). In the example (Figure 4) the number of objects identified is greater in the grayscale painting than in the colour one.

Fig. 3 - English text vector vs Italian text vector annotation.

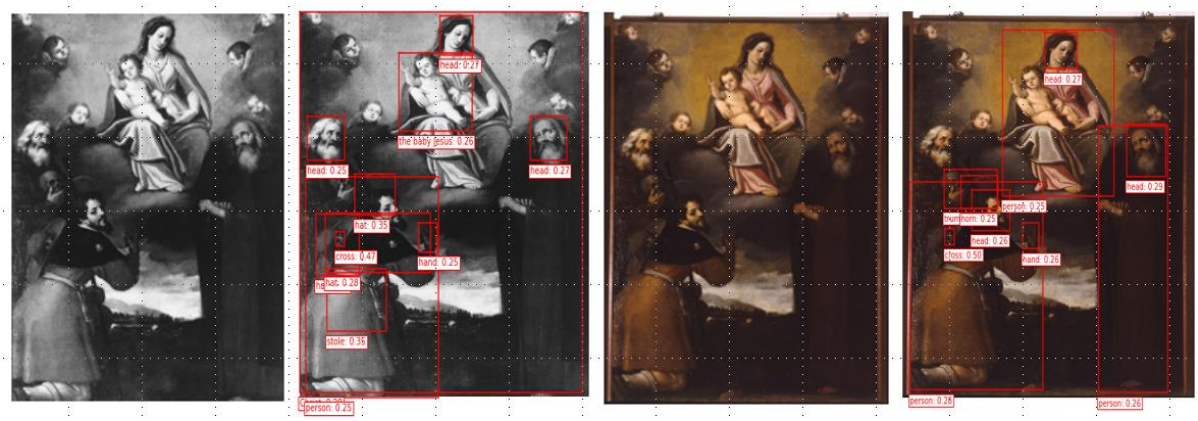


Fig. 4 - Grey scale image vs colour image annotation.

Cineca workflow: demo and massive application on HPC

Deep learning tools are very time and memory expensive, and requirements are usually quite high to achieve a good execution time. The training phase can take a long time if it is neglected. Various factors impact the execution time, such as the choice of one neural network over another, the learning rate, the number of classes to make predictions on, the number of bounding boxes, etc. All these aspects must be carefully analysed to port the neural network to an HPC machine.

Since on HPC a lot of computing nodes are used, we need to use a scheduler, like Slurm, to launch and manage our application.

Tools like Detectron2 and OWL-ViT can be benefit of the enormous computing power available with supercomputing machines. A porting on Marconi100 of both were made.

Detectron2 on Marconi100

Detectron2 is a tool built from Meta (Facebook) to detect objects in images. It can draw a bounding box on selected objects, adding a number about the inference quality. The higher that inference, the more reliable the prediction. Detectron2 can run on multi-GPU and multinode, to accelerate training phase. A lot of pre-trained models are already available on *model zoo* section. Such models, usually has tons of layers like 150 and more, and are trained over thousands of images.

Building from scratch a huge model like that, can be very time consuming from a programmer point of view, so it is strictly suggested to start from there.

The following is a typical workflow to prepare and run Detectron2 on HPC system:

1. Dataset setup and images selection (training, validation)
2. Bounding box creation on classes to be detected (tools: LabelImg)
3. Conversion in COCO format (suggested with Detectron2, tools: fiftyone)
4. Image augmentation (optional) in case of few images available
5. Code implementation- Model net choice
6. Code implementation- Training parameters setup (learning rate, batch size, etc.)
7. Code implementation-Run on GPU (multi-GPU)
8. Performance analysis (loss function, AP metrics, etc.)

The following is a snippet on how to launch a multi-GPU and multinode training with Detectron2 (Figure 5):

```
def main():

  cfg = setup()

  os.makedirs(cfg.OUTPUT_DIR, exist_ok=True)
  trainer = MyTrainer(cfg)
  trainer.resume_or_load(resume=False)

  return trainer.train()

if __name__ == "__main__":

  num_gpus_per_machine = int(sys.argv[1])

  starting_time = time.time()

  print("Process started...")

  hostnames, master_ip, master_port = get_hostnames()

  node_address, node_rank = get_node_rank(hostnames)

  num_nodes = len(hostnames)

  print("master_address: ", hostnames[0])
  print("node_address: ", node_address, " rank: ", node_rank)
  print("num_nodes: ", num_nodes)

  launch(main,
         num_machines = num_nodes,
         num_gpus_per_machine = num_gpus_per_machine,
         machine_rank = node_rank,
         dist_url = 'tcp://' + master_ip + ':' + master_port)
```

Fig. 5 - Detectron2: main code of multi-GPU configuration.

By using built-in *launch* method, the user can select number of computing nodes (machines in Detectron2) and number of GPUs per node. Additional source code is needed in order to define number of nodes and node ranks.

OWL-ViT on Marconi100

OWL-ViT is a tool part of Transformer deep learning framework. Unlike Detectron2, OWL-ViT can detect and drawing bounding boxes as output also on classes never seen during training

phase (**zero-shot detection**). Such outputs can be used to feed another net like Detectron2 or OWL-ViT itself, to add another training phase and improve quality predictions. A branch of original OWL-ViT is developed, to add new functionality, in particular about images saving. Several optimizations have been made to the original code, to achieve best performance. At the moment, OWL-ViT support the inference only on a single GPU, so it is necessary to split the dataset in more instances if we need to parallelize such phase.

For each input images, OWL-ViT execute a processing on GPU (device) by using the input text queries (Figure 6):

```
# Process image and text inputs
inputs = processor(text=text_queries, images=image_rgb, return_tensors="pt").to(device)

# Set model in evaluation mode
model = model.to(device)
model.eval()

# Get predictions
with torch.no_grad():
  outputs = model(**inputs)
```

Fig. 6 - OWL-ViT: main code to configure evaluation mode.

Conclusions

Our exploration of the state-of-the-art in the field of AI and digital humanities has highlighted the usefulness of AI in cataloguing and annotating digitized artwork collections. We have specifically focused on object detection and zero-shot detection, which have the capability to identify, locate, and classify objects within artwork. Object recognition gives unsatisfactory results due to lack of large collections of annotated examples. Zero-shot detection, on the other hand, can automatically annotate artwork without predefined classes, making it a useful alternative for the automatic annotation of artworks. We have tested several tools, including Detectron2, CLIP, and OWL-ViT, and demonstrated their capabilities through examples. An initial evaluation of the results shows that, although there are still errors, these AI tools can support the categorisation work of domain experts and eventually achieve a higher degree of precision in automatic annotation.

References

[1] Cetinic, E., & She, J. (2022). Understanding and Creating Art with AI: Review and Outlook. *ACM Transactions on Multimedia Computing, Communications and Applications*, *18*(2), pp. 1-22, https://doi.org/10.1145/3475799.

[2] Milani, F., Fraternali, P. (2021). A dataset and a convolutional model for iconography classification in paintings. *Journal on Computing and Cultural Heritage (JOCCH)*, *14*(4), ACM, pp. 1-18, https://doi.org/10.1145/3458885.

[3] Radford, A., Wook, J., Chris, K., Aditya, H., Gabriel, R., Sandhini, G., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2019). CLIP: Learning Transferable Visual Models from Natural Language Supervision. In *OpenAI*.

[4] Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., Wang, X., Zhai, X., Kipf, T., & Houlsby, N. (2022). Simple Open-Vocabulary Object Detection with Vision Transformers. *arXiv*.

[5] Cetinic E. Towards (2021). Generating and Evaluating Iconographic Image Captions of Artworks. *J Imaging. 2021 Jul 23;7(8):123*.

[6] Sack, H., Posthumus, E., & Brandhorst, H. (2022). The Art Historian's Bicycle Becomes an E-Bike. *ECCV: European Conference on Computer Vision*. Springer.

MULTIMODAL AI FOR DIGITAL HUMANITIES

Lorenzo Baraldi – Università degli Studi di Modena e Reggio Emilia

Deep Learning has been the most important revolution in the AI field of the last ten years. By combining large-scale data, computing power and scientific innovation, it has powered image classification networks, language models, and multi-modal architectures, giving them levels of efficacy which were unthinkable just a few years ago. Much of this innovation has also been driven by the emergence of new operators: starting from convolutional and recurrent neural networks, which have been the basis for the first successful models in Computer Vision and Natural Language processing, the last few years have been characterized by the emergence of the attention operator, which has been employed for developing most of the large-scale Foundational models like GPT [1] and Vision Transformers [2]. The AImageLab research group of the University of Modena and Reggio Emilia has been working actively in such topics for five years and has currently a focus on Foundational Models and Multi-modal AI.
The idea behind the operator is technically simple (i.e., taking a weighted summation of input slices, according to a learnable similarity function), easy to implement and to be optimized, and incredibly powerful from a learning perspective. The operator, indeed, has turned out to be so general to be applied to almost any kind of input data, ranging from images, language, 3D data, sensor data, etc. Apart from the famous large-scale language models, like GPT-3, Switch or the recent GPT-4, self-attentive architectures are also driving research in the field of multimodal AI, which combines both Vision and Language. An example is CLIP [3], which is contrastively trained to match sentences and images, and exhibits impressive zero-shot classification capabilities and incredibly powerful visual features. On the same line, a relevant example is that of Flamingo [4], which works in autoregressive and multi-modal manner so to enable a dialogue between the user and the system where both images and text can co-exist.

While all these approaches have been trained on (large) quantities of natural images, when it comes to applying the recent trends of AI to artistic data, life is not easy. One of the first issues one encounters, when trying to employ existing image models to artistic data, is the different distribution of low-level features. For instance, paintings have visible brush strokes, different colors, and different textures from a picture taken from the reality. All these create a distribution shift in low-level cues, which ultimately damages the final prediction given by the network. A possible way of solving this is that of transforming the input painting so to be closer to the low-

level distribution of real data, i.e. by creating a more realistically-looking version of the painting. This is the idea behind the Art2Real approach [5], in which a cycle-GAN is employed to transform artistic images into more realistic ones. As reaching a realistic result is fundamental here, we employ the Cycle-GAN framework together with a patch-level retrieval strategy, with which our network can copy from a database of real images and ensure realism (Fig. 1).

Fig. 1: Translation results generated by the Art2Real approach.

A second domain of artistic data in which we have applied Deep Learning is that of the recognition of handwritten documents (HTR). In this case we have an even larger domain gap with respect to "standard" networks, as the visual appearance of handwritten lines is completely different from that of the pictures found in standard image datasets. For this, we have developed custom networks (both CNN-based and Transformer-based) and collected large datasets for training them. A relevant example in this case is the Ludovico Antonio Muratori dataset [6], a line-level HTR dataset of Italian ancient manuscript, featuring the typical challenges of HTR on historical data. As neural networks are powerful and data greedy, we also augmented this dataset by employing a synthetic generation approach, in which we generate novel images of text lines by

starting from ASCII strings. This is also coupled with architectural studies on the right operator to employ: for instance, we proposed to employ a deformable convolution operator in place of the traditional convolution, as it can "follow" the writer's strokes and ultimately exhibits greater performance online-level HTR.

Vision, by itself, is a complex but solvable modality. Recent advances in the field, indeed, have demonstrated that we can classify natural images, detect and segment objects, and also deal properly with videos and moving objects. While there is still much to research in this area, we have also turned our attention to a more multi-modal setting in which, rather than just focusing on the Vision modality, we also integrate the textual modality. An example of a task in which language is added to vision is that of image captioning, i.e. the task of automatically generating natural language descriptions of input images. Clearly, the task is of fundamental importance, as not only it replicates an inherent human ability, but it also allows us to translate images to text, thus bringing them to a modality in which existing tools for search and analysis can be employed.

On this line, we have developed several state-of-the-art architectures over the years. The most successful one has been, probably, the Meshed-Memory Transformer [7]. Here, we start from the Transformer architecture, which employs the attention operator to build an encoder-decoder structure, where one modality can be translated into another. Then, we add two architectural innovations: firstly, we employ a mesh-like connectivity between encoder layers and decoder layers. In this way, we promote an exchange of low- and high-level features between the two modalities, enhancing the modality transfer capabilities of the architecture. Secondly, we endow the self-attention operator with a memory. In our design, the memory is implemented via learnable vectors, whose values can be adjusted via stochastic gradient descent like that of any other learnable parameter. In this way, though, we endow the attention operator with the capability of learning a-priori knowledge – something which would otherwise not be possible, given its exclusive dependency between inputs and outputs. The resulting model, which we called Meshed-Memory (or M2, for short) has been a state-of-the-art approach in image captioning for years and has inspired many subsequent works on the matter.

Architectures like M2, however, can just describe ordinary objects when they are trained on regular, medium-scale datasets like COCO [8]. To give our captioners the capability of describing any kind of object and more of the situations that exist in the real world, we need to scale the model and scale the amount of data on which it has been trained. This has been the driving philosophy behind the Universal Captioner architecture [9], in which we addressed the task of generating human-like descriptions with in-the-wild concepts. To this aim, we trained on web-

scale automatically collected datasets, for a total of 36.4 million image-text pairs. From an architectural point of view, instead, we employed a fully-attentive encoder-decoder and scaled it to reach a larger size. Beyond simply scaling data and models, we also adopted a few tricks to cope with the noisiness of the training data and maintain a fluent descriptive style. Firstly, we extracted textual keywords with a large-scale cross-modal model, so to translate our image with a sequence of tags. Then, we employed a binary stylistic token which gives our network the capability of distinguishing hand-collected from web-based image-caption pairs. At generation time, our network can generate human-like descriptions on standard datasets, reaching state-of-the-art performances, and it also showcases a very good zero-shot generalization capability on less common datasets. Most importantly, it has the capability of naming long-tail concepts like proper nouns of places, famous people and brands (Fig. 2).

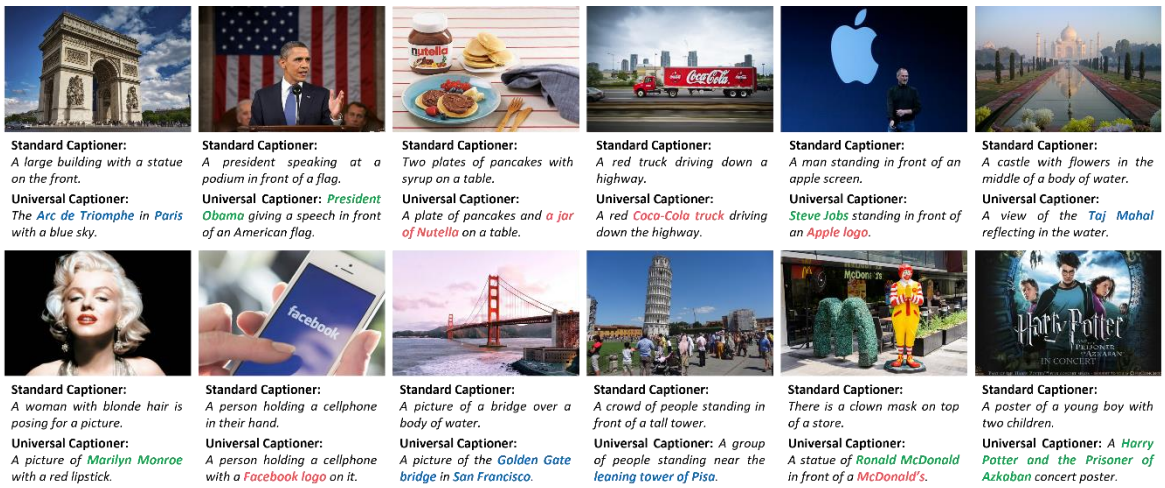


Fig. 2: Sample image descriptions generated by the Universal Captioner approach.

Training such an architecture would not have been possible without the computational support of CINECA, and without a powerful supercomputer like Marconi100. Overall, the experience of the last few years demonstrates that joining large-scale data, computing power and architectural innovations is a winning combination for developing accurate and production-ready models for multimodal AI, also when it comes to applying them to specific domains like that of the Digital Humanities. The best, however, has yet to come, and it is easy to imagine that the future is holding even more success for large-scale models. At AImageLab we are investing in this direction, as it is also testified by two appointments that we are organizing: the new master course on Scalable AI (organized in conjunction with NVIDIA) which will be taught in the new Master Degree on Artificial Intelligence Engineering and the ELLIS PhD School on Large-Scale

AI, where we will host top-level scientists, in Modena, for giving lectures to PhD students from all over Europe.

References

[1] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. "Improving language understanding by generative pre-training", 2018.

[2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. "Transformers for image recognition at scale". ICLR 2021.

[3] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. "Learning transferable visual models from natural language supervision". ICML 2021

[4] Alayrac, J. B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., ... & Simonyan, K. "Flamingo: a visual language model for few-shot learning". NeurIPS 2022.

[5] Tomei, M., Cornia, M., Baraldi, L., & Cucchiara, R. "Art2real: Unfolding the reality of artworks via semantically-aware image-to-image translation". CVPR 2019

[6] Cascianelli, S., Cornia, M., Baraldi, L., & Cucchiara, R. "Boosting modern and historical handwritten text recognition with deformable convolutions". International Journal on Document Analysis and Recognition (IJDAR), 2022.

[7] Cornia, M., Stefanini, M., Baraldi, L., & Cucchiara, R. "Meshed-memory transformer for image captioning". CVPR 2020

[8] Sarto, S., Barraco, M., Cornia M., Baraldi, L., & Cucchiara, R. "Positive-Augmented Constrastive Learning for Image and Video Captioning Evaluation", CVPR 2023

[9] Cornia, M., Baraldi, L., Fiameni, G., & Cucchiara, R. "Universal captioner: Long-tail vision-and-language model training through content-style separation". arXiv 2023.

FOOD DATA DIGESTION: A MULTIDISCIPLINARY METHODOLOGY BETWEEN ART, CULTURE AND ARTIFICIAL INTELLIGENCE

Alessia Tripaldi – Sineglossa
Federico Bomba – Sineglossa / UNIBZ

Introduction

Food Data Digestion (FDD) [1] is a two-year research and production project curated by Sineglossa, in collaboration with Play With Food, and supported by Compagnia di San Paolo.

In FDD the concept of food is interpreted in the sense of "nourishment", and more specifically of the relationship between data and Artificial Intelligence, of data as nourishment that gives shape and "identity" to an AI. The aim is to foster the **integration of artistic research and scientific research in relation to new technologies**, so as to 'nourish', precisely, each of the two sectors through the mutual contamination of skills, visions and experiences.

The project's macro-objectives of impact were:

- concerning **art**, the macro-objective is to increase artists' knowledge of what Artificial Intelligence is and how it can be used, in order to offer them both **new research opportunities** and **new media** through which they can bring their "voice" to the attention of the public and governance bodies;
- concerning **technology**, the macro-objective is to stimulate new visions on the possible functions and developments of AI, in order to foster an **ethical, aesthetic, sustainable and inclusive technological innovation**, aimed not only at economic gain, but at people's wellbeing, in a tech for good perspective.

The first year of the FDD project saw the creation of the work **And We Thought** by the artist Roberto Fassone, together with the creative technologist Andrea Zaninello and the artistic director of the project, Federico Bomba: the process of conceiving, realising and disseminating the work is the subject of this document, which aims to report on the FDD methodology, assessing its strengths and margins for improvement (Fig. 1).

The reflections and information reported here derive from qualitative evaluation tools - interviews and questionnaires - disseminated throughout all stages of the process.

In the following paragraphs, to outline a methodological framework that can be replicated in other contexts, the focus will be on the two main steps that led to the creation of the work.

Art and tech residency

For being truly the result of contamination between skills, artists and technologists must have the time and the creative freedom to envision and produce an AI-based work of art *together*. Giving an artist a "closed" AI device, already directed towards a specific function, means limiting their creative capacity, the possibility of imagining innovative and visionary declinations of AI, as well as the acquisition of new skills, the understanding of what an AI is and the exploration of its endless possibilities. Similarly, giving a technologist a 'closed' artistic project, in which they have the role of technically executing an idea, means cutting them off from the creative part of the process, from the possibility of exploring AI through the disruptive gaze of an artist.

In the first months of the FDD project artist Roberto Fassone, creative technologist Andrea Zaniello, AI AiLai, and artistic director Federico Bomba were all involved in an artistic residency. The presence of an artistic director in the role of mediator is another element of great importance for the success of an 'ecosystemic' process: **the mediator is a figure capable of 'translating' the languages of the actors involved** and guiding them in a relationship of mutual listening.

Fig.1: Federico Bomba and Roberto Fassone; ph. Alain Battiloro.

The FDD residency's central theme was food. The artist selected for the project, although having already produced works based on digital technologies, had no previous knowledge of AI. This element, which is of particular importance for considerations of the impact of the methodology, will be discussed again later on.

During the first phase of the residency, the artist and the creative technologist focused on an in-depth study of the concepts of AI and machine learning.

Starting from the stimuli collected in the first phase, the artist proposed working on the **concept of "hallucination"**, a scientific term used in natural language processing to define the generation by machines of contents that do not produce the effect for which the AI has been programmed: in Fassone's vision, the hallucinogenic substances should have become the AI's food. This is a meaningful insight regarding what is intended as art's disruptive potential. In science, the phenomenon of hallucination is studied to be avoided, to direct machines towards efficiency. On the other hand, in Fassone's work the distortion, the "error" of the machine, becomes the focus of research, an innovative and unexplored avenue that was immediately met with the enthusiasm of the creative technologist. From the subsequent meetings and exchanges between Fassone and Zaninello, **AiLai was born, the Artificial Intelligence fed with reports of psychedelic journeys produced by thousands of people following the ingestion of hallucinogenic mushrooms**.

From the accounts of both, an informal and collaborative relationship has emerged: 'I created a pre-language model and gave it to Roberto, who started playing with it,' Zaninello recounts: 'At a certain point he called me and asked: Why does it invent names or places, where does it get them from? I explained to him that there is random initialisation in this language model, so that each generation of AiLai is unique'. It is precisely the AI's ability to surprise with its own 'inventions', what Fassone calls 'revelations', that has been at the heart of the art project and the works derived from AiLai's inventions, which will be discussed below. A process of four-handed exploration, of co-design between an artist and a technologist, which led to the creation of *And We Thought* [2] the "multimedia art project that investigates the unexpected in machine learning to explore the human mechanisms of creativity and knowledge" (Fig. 2).

Fig.2: And We Thought III; Roberto Fassone, AI LAI, LZ. Visual by Roberto Fassone.

Exhibition

The first public exhibition of And We Thought was hosted in Turin in June 2022 at the Combo spaces, as part of the Play With Food #Cantieri2022 festival and re-proposed at Ars Electronica 2022 and Artcity 2023 (Fig. 3; 5). And We Thought is a **multimedia work**, composed of:

- three short films inspired by one of the stories produced by AI (Fig. 4);
- an artist's diary including all the stories generated to date;
- a series of posters dedicated to the most poetic and unconventional texts generated by AiLai;
- a rap album taking its title from a user-generated story.

Fig.3: And We Thought, Roberto Fassone, AI LAI. Exhibit at CANTIERI 2022 Turin, ph. Alain Battiloro.

Fig.4: And We Thought III_Roberto Fassone (channeling Led Zeppelin), The Road, video still.

This multifacetedness of languages is the first relevant element to focus on: as already mentioned, the experience of co-designing with a technologist and an AI has opened up scenarios for the artist that he would not have imagined at the beginning of the process (a point when, by his own admission, he felt somewhat bewildered by the creative possibilities offered by an Artificial Intelligence). Fassone's exploration of the concept of hallucination and AiLai's 'revelations' have, on the other hand, initiated a **new line of research** for him. This is probably the most fascinating aspect to highlight regarding the artistic component of the project, which

has gone beyond the immediacy of the output generated by the AI (in this case, the 'hallucinated' texts), trespassing into other languages that derive from the artist's experience, from his vision, stimulated by the collaboration with the AI without being limited by it.

Fig. 5: And We Thought III, Roberto Fassone, AI LAI, LZ. Exhibit at Alchemilla (Bologna) during Art City 2023. ph©RolandoPaoloGuerzoni

A methodology between art, culture and artificial intelligence

What, in conclusion, are the characteristics of an approach aimed at facilitating contamination between artistic and scientific research?

Starting from the experience and data collected during the FFD - And We Thought project, we have identified the key concepts of an art&tech methodology, in order to stimulate reflections and suggestions on the topic and to facilitate the emergence and implementation of similar experiences by cultural organisations, training organisations and scientific research centres [3].

Multidisciplinarity. An art&tech project is, by its very nature, based on contamination between disciplines. For the disciplines involved in the process to really contaminate each other - and thus learn from each other and influence each other - it is necessary to ensure the coexistence and balance of the different skills and approaches involved.

With respect to the production of an art&tech work, it is important that the process of generating the work is based on co-design between artist and technologist, in order to involve both parties in the conception and realisation of the work.

With respect to art&tech training, it is important to propose different points of view on the subject matter, in order to educate on the language and use of new technologies through a plurality of visions: technical, artistic, social, ethical.

In the case of FDD, the team of trainers was composed of an artist who had already had experience of art&IA, a technologist who had already collaborated with an artist, and an art director who had already designed and directed projects involving contamination between art and Artificial Intelligence.

The presence of a cultural manager in the role of mediator represents a further support to the multidisciplinary nature of the project, since he can "translate" the languages of the parties involved and guide them in understanding and listening to each other, thus fostering the interpenetration of skills.

Concreteness. For the art world to be able to integrate new technologies into its research and works, artists need to be given the opportunity to "touch with their hands", to experience the use of technologies in a practical manner right from the training phase.

To produce an art&tech work, the artist needs to be able to interact with the technological component, exploring its functioning and potential and thus actively contributing to its development. In the case of FDD, for example, the creative technologist created a pre-model with which the artist could 'play', discovering through direct experience the possibilities of AiLai.

For art&tech training, the approach of the training course must be based on a theoretical-practical one, in which the transmission of notions is side by side with the experimentation of tools used in one's field of research. The aim is not only to make artists understand what certain technologies are and how they work, but also - and above all - to stimulate reflection on how those technologies can be used in the artistic world.

As mentioned above, a multidisciplinary team of teachers contributes significantly to this "plural" approach that approaches technology from multiple perspectives.

Communities and audiences. By imagining the art&tech methodology from an ecosystemic perspective, one should consider a third element which goes together the world of art and the world of science: the communities destined to receive the process. Involving communities in the ecosystem is what helps an art&tech project to produce an impact in the territory - local, national, international -, integrating technical, aesthetic and ethical research on new technologies with the needs and challenges of the contemporary world.

The levels of community involvement can be different, and of varying degrees. In some cases, specific interest groups (e.g. citizens, activists, policymakers) may be brought into the process of co-designing the work. Speaking of Artificial Intelligence, for example, stakeholders could be involved in the discussion on data (what data to use, how to use it, how to disseminate it) for the realisation of an art&tech work that helps to promote a cause, to analyse the needs of a citizenry, to improve the quality of life in a given geographical area.

In other cases, as was the case in FDD for example, the communities of reference correspond to the public that will benefit from the work.

Whatever the modality, the involvement of the audiences is an essential component of an ecosystem approach to art&tech, not only for the social impacts that a project can produce, but also for the education of people in the new languages which are being investigated and created. A work of art based on a new technology is a powerful medium for raising awareness among citizens about the risks, limits, advantages and opportunities that that technology represents.

For this reason, even when communities are not directly involved in the co-design process, it is important that the dissemination of the artwork is not limited to an exhibition, but, as was the case in FDD, includes moments of dissemination where the actors in the process share their experience, or where the public can interact with the artwork itself, contributing to its development and learning about it at the same time.

Process. The difference between a goal-oriented and a process-oriented approach is that in the former case the focus is on the final result, in the latter on the process leading to the result. For the implementation of an art&tech methodology it is essential to put the process before the result: this does not mean that certain aspects of the final output cannot be established and known from the outset, but that the path to reach that output must leave room for deviations and unexpected discoveries. To put it in a word, an art&tech methodology must be based on the concept of exploration.

Since the ultimate goal of a methodology that combines two different worlds is to produce innovation - of thought, of models, of action - it is necessary to create a fertile environment for innovation, an environment that interprets uncertainties as spaces to explore and failures as opportunities to learn, rather than as obstacles to productivity. A process-oriented approach, which does not rush at the result or consider the outcome as definitive, which does not pretend to move within impassable boundaries, which is willing to change direction according to external stimuli, is a fertile environment for innovation.

From a practical point of view, this translates into attention to two aspects. The first is the space/time dedicated to the process, which does not necessarily have to be long, but must provide the right amount of space for the parties involved to get to know and surprise each other. The second is the dimension of discussion and dialogue between the parties: the more moments dedicated to the exchange between points of view - and the more points of view - the more likely it is that one will come across an unexplored path along the way.

References

[1] Food Data Digestion is produced and curated by Sineglossa, in collaboration with Play With Food festival and Free University of Bozen-Bolzano, supported by the Compagnia di San Paolo Foundation as part of the call for proposals "ART~WAVES. For creativity, from idea to scene". To know more: https://sineglossa.it/en/projects/food-data-digestion/

[2] To explore And We Thought artwork and interact with AI LAI visit https://andwethought.it/

[3] Sineglossa has been running research on STEAM (Science – Technology – Engineering – Art – Math) methodologies for years. It is currently a partner in two European projects promoting STEAM education: https://sineglossa.it/en/projects/westeam/ https://sineglossa.it/en/projects/steam-process/

DIGITAL AI FOR IA: ARTIFICIAL INTELLIGENCE FOR INTERPRETATIVE ARCHAEOLOGY

Maurizio Forte – Duke University

Felipe I. De Castro, Irakli Pkhovelishvili, FollowFox.AI

AI and Virtual Archaeology

Virtual Archaeology (1996) [1], Cyberarchaeology (2008) [2], and AI-Archaeology (2022) [3] represent significant turning points in the evolution of the digital depiction of the past. If Virtual Archaeology was "model-oriented" (3D computer graphics), Cyberarchaeology aspired for the interactive simulation of the past as a hyperreal world (Fig. 1).

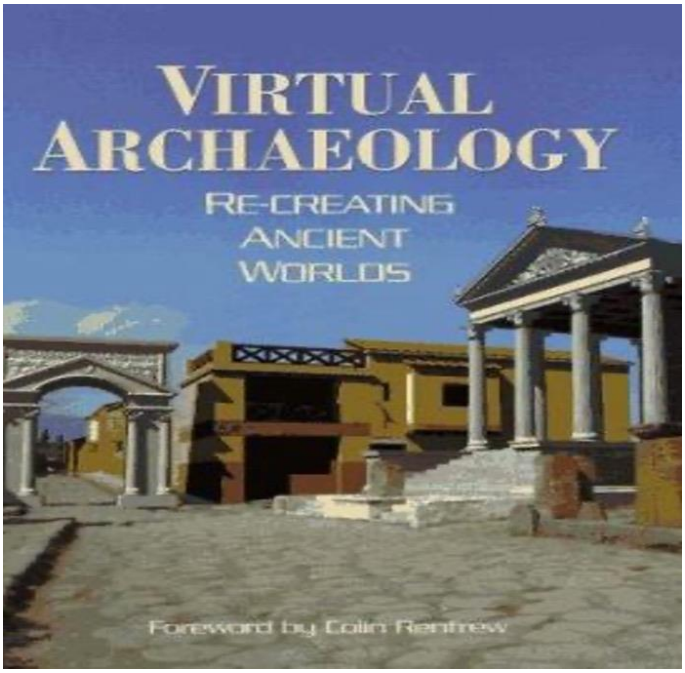


Fig. 1 - The first book on Virtual Archaeology (M. Forte, A. Silioti eds.)

Hyperreality is strongly related to Baudrillard's concept of the Simulacrum [4], which he defines as something that replaces reality with its representations. The present world, according to Baudrillard, is a simulacrum in which reality has been substituted by false pictures to the point where it is impossible to distinguish between the real and the unreal. In this context, he made the controversial and provocative comment, "The Gulf War did not occur," emphasizing that the 'reality' of the Gulf War was portrayed to the public through media portrayals. This emphasis on the diversity of media and the power of different digital ontologies can explain the meaning of this statement: the meta-reality constitutes the basis of the current human knowledge, because of the potential impact of its representations. In provocative terms: the past is a multiverse.

This approach is relevant also in the reconstruction of the past as an endless process generated by multiple simulations/representations. In this case, the uncertainty of the interpretation is managed by different and multivocal perspectives and not by a utopian idea of fidelity/authenticity/objectivity. In archaeology, material culture is the *medium* and this materiality requires a complex effort in order to rebuild its affordances and contexts (Figs. 2 – 3).

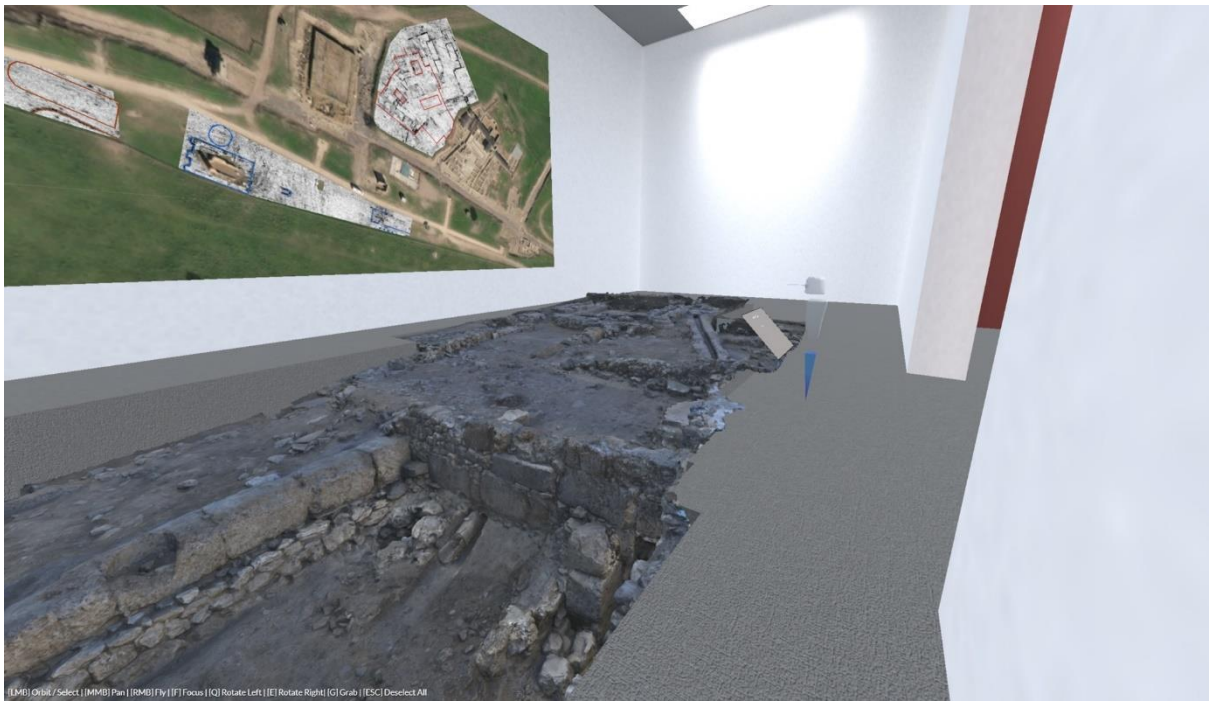


Fig. 2 - A cyberarchaeology approach in a VR collaborative environment. In this case, an archaeological excavation is virtually represented in the Mozilla Hub platform (elab. A. LoPiano).

In digital archaeology this complex workflow generates data in different format and according to specific ontologies, often in 3D (Fig.4).

The ability to mimic processes has played a crucial role in all of these study fields. AI archaeology, in fact, opens new and endless perspectives in the interpretation and reconstruction process. The application of AI in archaeology is not new [5; 6] and it is a well-established research field in particular in the automatic classification of objects and monumental structures, but also in remote sensing sites detection [7]. What is really new is the availability of very fast and open-access tools for infinite visualizations and graphic simulations, starting with a simple textual syntax (text to image) or by image variants (image to image, image to videos, image/video to model). In two words, "Deep Learning" (DL).

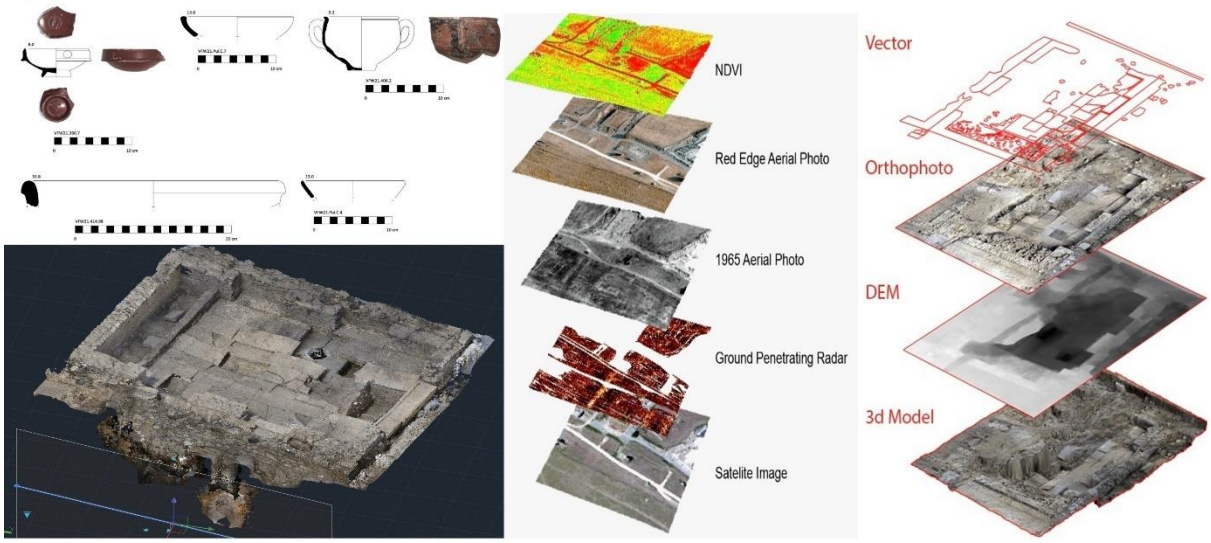


Fig. 3 - The digital ontologies of archaeological research in a representation of artifacts, sites, layers and models.

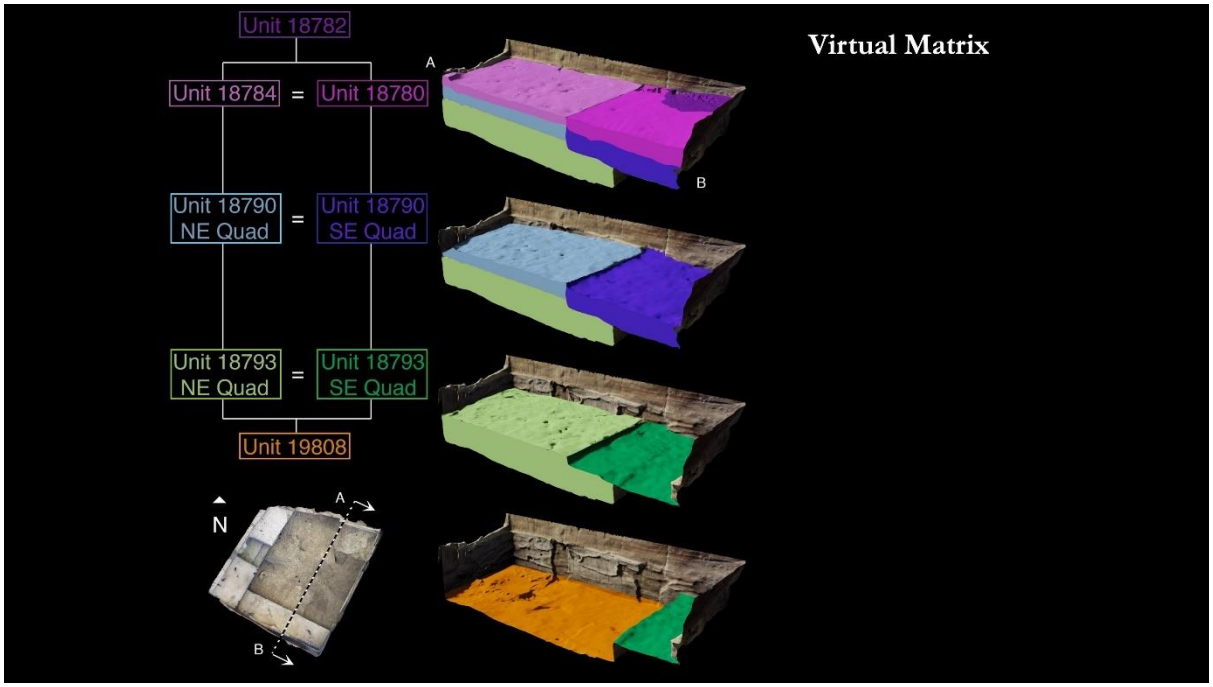


Fig. 4 - Virtual excavation of stratigraphic layers of a Neolithic house at Catalhoyuk (Turkey) – 3D Digging Project (2010-12). All these archaeological deposits are removed during the excavation and, actually, they "exist" just virtually. The sequence of visual representation is conceptually close to the idea of the past as a multiverse.

We anticipate a near future in which artificial intelligence can generate visuals, models, and complex simulations just from simple coding or textual prompts. Within an unlimited framework of simulations, generative AI in VR-real-time situations could reassess digital consistency research issues. In accordance with this methodological approach, we "perform the past" as opposed to recreating it, given the regime of ambiguity with which we contend and the subjectivity of our archaeological and historical interpretations. In other words, the stronger our capacity to reinterpret a place, monument, landscape, or relic, the more virtual/digital situations we generate.

The generative potential of AI visualizations multiplies and diversifies the content; in other terms, it increases substantially the number of iterations and visualization. Given the uncertainty of the reconstruction, the research work in AI archaeology should be focused on the relationship between textual narrative (the "prompt") image training, and visual simulations.

The artificial intelligence software Stable Diffusion generates an image in response to a textual request. Its training was conducted via Stable Diffusion employing a database of more than 5 billion photos with their appropriate descriptions, allowing the AI to comprehend the concepts linked with the images themselves. On top of it, Stable Diffusion is a totally open-source code, based on an open-source database (LAION-5B), and capable of running locally on a consumer-grade machine.

In the near future, AI archaeology will be perceived as a generative-autopoietic system that generates its own hypotheses, digital codes, models, and syntaxes between human and artificial brains.

From AI to Art: a brief background

Artificial Intelligence (AI) has historically been a loosely-defined term, having been applied to systems as simple as formulaic responses to a hardcoded setup (as in a computer playing tic-tac-toe), and as complex as conversational chatbots such as LaMDA, ChatGPT or Bing Chat, which passed the Turing test, as reported by several media. We subscribe to the definition that an AI is any artificial agent that receives percepts from the environment, maps percepts sequences to actions, and performs them accordingly [8, preface]. A subset of AI systems that gained prominence in the past few years are those that rely on Machine Learning (ML), defined as AI systems that exhibit the capacity to improve performance based on training and experience; of special note are those that are designed using multilayer neural networks, an approach called "Deep Learning" [8, chapter 1].

The advent of Deep Learning AIs, coupled with the steady advance in graphics computing power and computer vision, have revolutionized the role that AI can play in art. The first AI capable of better-than-human performance in image categorization (identification of specific concepts within photos and images) was AlexNet, a Deep Learning AI that won the 2012 ImageNet Large Scale Image Recognition Challenge. AlexNet was a multilayer artificial neural network composed of 650,000 neurons and totaling 60 million parameters [9]. Since then, there has been increasing interest in the use of Deep Learning AIs in the fields of image processing and handling, culminating in the advent of DALL-E in 2021.

OpenAI's DALL-E was the first public release of a *generative* AI – that is, an AI that is capable of creating original content, as opposed to just processing existing images. Launched in January 2021, DALL-E was a pioneering development over OpenAI's research on Large Language Models (LLMs), which are AIs capable of reading texts and predicting future words in a sentence based on their relative frequency within a very large dataset of texts (and whose most well-known example currently is ChatGPT).

OpenAI's plan was to correlate a textual description with an image and then apply the probabilistic method used to forecast a word in a phrase to the problem of creating a new image. Instead of predicting a word in a text, however, DALL-E would start with a random noise image, and, through a number of iterations, slowly change the original image so as to create an image that matched a textual prompt given by the user. In order to learn how to do this, DALL-E had 12 billion parameters and was trained over a proprietary dataset of 250 million pairs of images with corresponding textual descriptions, which were scraped from the internet by OpenAI's team [10]. Other generative AIs quickly came about, including Google's Imagen and Parti, Midjourney, and an improved version of DALL-E called DALL-E 2. By August 2022, a startup called Stability AI launched a new generative AI: Stable Diffusion.

Stable Diffusion is another revolutionary step in an already very active field. Thanks to a new paradigm in Deep Learning called Latent Diffusion Models (LDMs), Stability AI designed Stable Diffusion as a compact AI with only 900 million parameters, capable of being run locally in PCs equipped with consumer-grade GPUs. This is different from all other generative AIs made public until today, which are large models that can only be run in cloud-based servers and under the purview of their original developers [11]. In addition, Stability AI launched Stable Diffusion as a 100% open-source project: Its source code, training datasets (LAION-5B, containing over 5 billion image-text pairs) and trained model weights – whose creation requires significant computing power and time – are available for public use.

Launching Stable Diffusion as a compact and open-source AI resulted in an explosion of interest among developers, artists, and researchers, as indicated by the engagement with the AI's primary code repository on GitHub. As of February 16, 2023, in less than a year since launch, the original Stable Diffusion repository registered 41.5k stars (akin to GitHub "likes"), eclipsing other highly successful open source projects such as WordPress (leading CMS powering blogs and websites worldwide; 17.1k stars), WebTorrent (peer-to-peer file transfer protocol; 27.5k stars) or the Ethereum blockchain protocol (41.2k stars), and approaching that of the Python programming language (50.7k stars).

Accordingly, opening the technology to the crowd has brought about an avalanche of new tools and creative algorithms leveraging on the Stable Diffusion AI, ranging from algorithms to create animations using a sequence of AI-generated images (Deforum), to complete web-based user interfaces (AUTOMATIC1111 Web UI), or even to creating music snippets by way of generating audio spectrogram images and interpreting them as audio files (Riffusion).

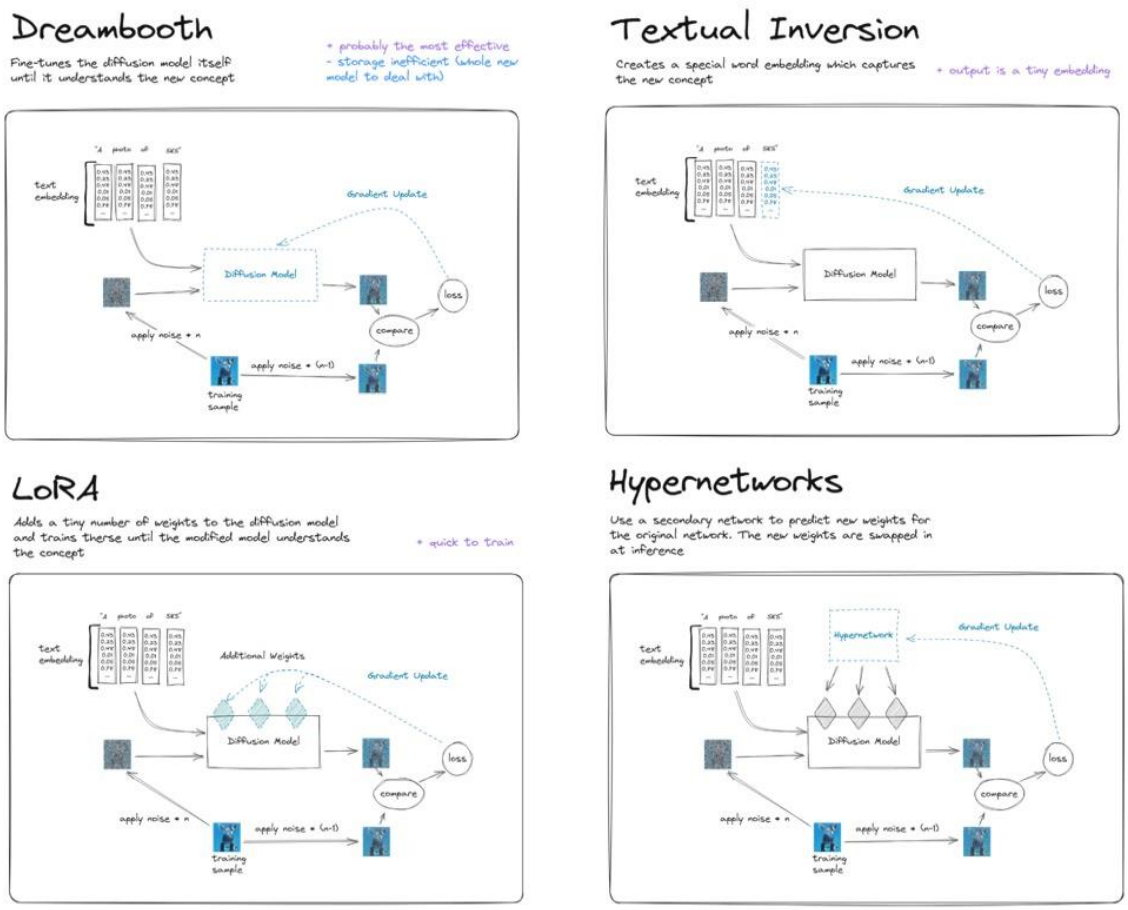


Fig. 5 - simplified diagram of the four fine-tuning algorithms. Credit goes to Reddit user use_excalidraw

Particularly interesting are algorithms like Textual Inversion, DreamBooth, Low-Rank Adaptation (LoRA), and HyperNetworks that let users teach the AI. While these algorithms work using distinct modeling paradigms (Fig.5), all of them ultimately relate to exposing the AI to a set of input images with their corresponding textual captions, so as to fine-tune its response to user input; in effect, what they all do is train the AI's 'memory' to associate a specific instance whenever its corresponding name presents itself in a given prompt. These so-called fine-tuning algorithms allow the generative AI to produce results that are remarkably consistent across generative iterations, allowing for the recalling of persistent entities upon user request.

With that knowledge in hand, our approach becomes evident: we chose a compact, yet powerful generative AI that is flexible and capable of being tinkered with by a regular user. We taught this AI what a specific concept is by ostensibly showing instances of it in a small dataset of photo-caption pairs of our own design. Then, we instructed the AI to generate original images of the exact concept we had taught it, in a feedback loop that generates intuition in us as to how that concept might seem.

Limitations of the Original Models

The AIs mentioned above are extremely powerful and capable but have their own limitations. When it comes to applications such as archeological interpretation, we need to highlight a few of such limitations.

- Absence of certain images/subjects in the original dataset: taking Stable Diffusion as a reference – 5 billion images is a lot, but far from every possible image. This means that the original AI was not trained on domain-specific images that require specificity and are not widely present on the internet.
- Under/over-representation: even if specific archeological images of certain examples and monuments exist on the dataset, they were most likely underrepresented compared to more popular objects. Coupled with the fact that the original AI training process was not weighted to balance this out (that is, each image was repeated or 'shown' an equal number of times without regard to what subject it shows), the end result was an AI that is extremely powerful, accurate, and flexible for creating images of very popular things like Hollywood stars or monuments like Eiffel tower, but that struggles with less popular subjects such as Etruscan Tumuli.
- Incorrect or insufficient text captions: even in the cases when the database used to train the original AI has enough images of a specific subject, it's not guaranteed that those

images are properly labeled or captioned, especially whenever archeological terms are concerned. The problem is amplified by the fact that images from the database were filtered by another AI model on caption accuracy that might not know such specific terms. So, an image of an Etruscan Tumulus tagged only by the term "Etruscan Tumulus" has a high probability of being filtered out. And in those cases when it was not filtered out, it might have ended up with a non-domain-specific tag, such as a simple visual description of the image. As a result, the original AI struggles with generating images when domain-specific terminology is used.

All of these limitations can be addressed by the fine-tuning process: the domain-specific images can be collected manually, or even new ones can be taken for this particular purpose, the issue of underrepresentation can be adjusted by increasing the number of repeats in the fine-tuning process, and text captions can be manually generated to teach the AI specific and accurate terms about the images.

Case Studies

The application of AI tools in our first experiments included: Etruscan tumuli, Roman statues and environmental reconstructions of Roman landscapes. This experiment is just a glimpse at the possibilities that customized AI models can enable, especially as the community and users master how to optimally harness those powerful capabilities.

Etruscan tumuli are aristocratic funerary monuments (7th-6th cent. BCE) in the shape of large mounds that show the symbolic power of these families and the visibility of the burials in relation to the city space and funerary contexts. The reconstruction and visualization by Stable Diffusion (figs.) and a specific animation created in Deforum show a generative simulation of tumuli in different landscape settings and environmental conditions.

As the original Stable Diffusion AI could not reliably create accurate images of Etruscan Tumuli, we elected to teach the AI what an Etruscan Tumuli was by fine-tuning it, so we exposed it to a small dataset of high-quality additional data in form of text and image pairs. A few key parameters used for the fine-tuning:

- The official Stable Diffusion 1.5 as a starting point
- EveryDream 1.0 as a DreamBooth-based fine-tuning tool
- The training data was 20-25 images of Etruscan Tumuli available on the Internet (see in Figs. 6 -7 an example of the dataset we utilized).

- Each image was tagged with descriptive text, including a unique identifier "olis" which is known to be a rare token for the Stable Diffusion model. Basically, after training, the AI will associate "olis" with images that look like Etruscan Tumuli.
- A few different experiments were run in order to gain intuition on the optimal number of training steps (that is, how many times each training image is shown to the model in the teaching process), as well as in what image types and variety contribute best to the quality of the end result.
- The relatively optimal resulting AIs were selected by trial and error based on the quality of the output that was observed.

Fig. 6 - example of dataset of image-caption pair used to train Stable Diffusion AI to generate Etruscan tumuli images. For the avoidance of doubt, note that all pictures shown in this image are actual photographs of Etruscan tumuli mounds.

It's important to highlight that the speed of the development of this tech is extremely high and since the experiments were run a few weeks ago, there have been a few minor to major releases and breakthroughs related to Stable Diffusion that are certain to push further what can be done with this AI. Most notably, the advancement of image-to-image generations using [12] ControlNet. The groundbreaking aspect of ControlNET is its answer to the spatial consistency problem. ControlNet solves this by offering a mechanism that enables Stable Diffusion models to employ extra input conditions that tell the model precisely what to do. Previously, there was no efficient way to instruct an AI model which parts of an input image to retain.

Fig. 7 - Visual reconstruction of Etruscan tumuli by typological and environmental features. The simulation stressed the importance of this shape (mound) as important landmark.

The experiment over the statue-portrait (Fig. 8) of the emperor Trajan shows the AI capacity to humanize a statue, to generate visual empathy, in other terms to generate a human and not just a portrait. This also recalls the fact that the marble statues in Roman times were colored and tended to be "photorealistic": we consider them as one of the earliest "media".

The last experiment concerns a visual representation of paleobotany and environmental analyses conducted by Duke University and the University of Modena and Reggio in the archaeological site of Vulci, Viterbo (10th cent. BCE-4th cent. CE). In this case the research goal was the visual reconstruction of the ancient environment in Etruscan and Roman times by using textual syntax coming from the samples' paleobotany analyses. For the AI visual reconstruction we used this scientific narrative [13] A*mong herbaceous plants, as expected there are many anthropogenic indicators,*

cultivated or wild plants that grow well near places frequented by humans (in particular, trampling areas: plantain). Of interest, the presence of cereals in the reservoir is high. The highest amount can be attributed to the Avena/Triticum type (oats/wheat), which is more closely related to domesticated plants. Also attributable to the fields is the finding of poppy, an archaeophyte weed. The wet environment with water stagnation is evidenced by the water lentil, Lemna, and typhus.

Fig. 8 - Bust of the Roman Emperor Trajan (Glyptothek, Munich). Face "humanization": 512-depth-ema model from Stable Diffusion 2.0 release.

All this articulated list of plants, flowers and cultivations is summarized in a series of very accurate images of vegetation and environmental reconstructions (Fig. 9). The AI visual simulations, obtained in a range of a minute by using DALL-E 2 (online), show a very accurate result validated by a team of specialists.

Fig. 9 - Visual reconstruction of the Etruscan and Roman environment of Vulci (6th cent.BCE - 3rd cent. CE).

Conclusions

The visual generative power of AI is theoretically endless and puts humanity towards a new heuristic challenge: hyperrealism, more real than real. What kind of knowledge can we transmit to a future where the information will be hybridized between human and artificial intelligence? Can we envision the past as we do in our contemporary societies?

Fig. 10 - This image indicates the power that Stable Diffusion has in reimagining archaeological sites as if they belonged to completely different environments. Each landscape presented here was generated by the AI over the same random seed image, changing only one element in the prompt. The prompt offered was: "high-quality landscape view of Etruscan tumuli mounds, in the background, olis", with the negative caption: "low quality". Counterclockwise from top left, the element used was: "desert", "sea", "wheat fields", and "snowy fields".

The realm of hyper-realism generated by neural networks can generate visual simulations in a very short time and by browsing online archives of billions of images and models. The concept of a past as a "multiverse", in relation to the generative power of AI, is a fascinating and thought-provoking idea that questions our understanding of history and archaeology and the very nature of reality

In fact, in the last decade, the theoretical approach in cyberarchaeology [3] introduced the idea of "potential past", with an emphasis on the relativistic idea that the past is fluid and perceived differently by different societies through space and time.

The first experiments in Stable Diffusion and Deforum demonstrate the success of an AI approach for the simulation of sites, artifacts, and ancient landscapes (Fig. 10). Re-imaging sites, monuments and landscapes (i.e., by different backgrounds) can suggest new interpretations and simulations. All these variations are part of the interpretation process, at the intersection of space (for example a necropolis) and time (the evolution of a shape or a morphology). Environmental settings and photorealism challenge our contemporary view of the past as a *multiverse*.

The idea to "blend" them, rather than to reconstruct them, is revolutionary because it uses the generative and transformative tool of a neural network, rather than just a single hypothesis coming from a traditional scholarship. Things and objects are transitional and coevolve in something else: they are syntax, images and models; this is the best representation of the past we have in our hands. The multiplication of contents builds new knowledge and pushes our brains to contemplate and compare diverse options, worlds, and contexts. From this point of view, even a simple taxonomy can recreate a complex visual narrative, because of the metaphysic power of the simulation.

In conclusion, Deep Learning AIs are an advanced form of artificial intelligence that can revolutionize the way we process and analyze large sets of data, especially with regard to generative archaeology. The first experiments, still in an embryonic phase, show unexplored research perspectives in the visualization and simulation of complex datasets and models. Perhaps it might be a first step for the systematic integration of Deep Learning AIs in the archaeological process.

References

[1] M. Forte, A. Siliotti (eds. by) *Virtual Archaeology*, (forward by Colin Renfrew) Thames & Hudson Ltd, 1997.

[2] Forte, M., *CyberArchaeology* (Edited by Maurizio Forte), BAR International Series 2177, Oxford, 2010.

[3] Forte, M., *Cyberarchaelogy: a Post-Virtual Perspective*, in D.T. Golberg and P. Svensson (eds.) Humanities and the Digital. A Visioning Statement, MIT Press, 2015, 295-309.

[4] Baudrillard, J. (1994). *Simulacra and Simulation* (S. Glaser, Trans.). University of Michigan Press.

[5] Jamil, A.H.; Yakub, F.; Azizul Azizan, A.; Roslan, S.A.; Zaki, S.A.; Ahmad, S.A. A Review on Deep Learning Application for Detection of Archaeological Structures. J. Adv. Res. Appl. Sci. Eng. Technol. 2022, 26, 7–14.

[6] Davis, D.S. Defining what we study: The contribution of machine automation in archaeological research. Digit. Appl. Archaeol. Cult. Herit. 2020, 18, e00152.

[7] Orengo, H.A.; Conesa, F.C.; Garcia-Molsosa, A.; Lobo, A.; Green, A.S.; Madella, M.; Petrie, C.A. Automated detection of archaeological mounds using machine learning classification of multi-sensor and multi-temporal satellite data. Proc. Natl. Acad. Sci. USA 2020, 117, 18240–18250.

[8] Russel, S. J., Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed). Pearson.

[9] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2017-05-24). ImageNet classification with deep convolutional neural networks. Communications of the ACM. 60 (6): 84–90. doi:10.1145/3065386. ISSN 0001-0782. S2CID 195908774.h

[10] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. arXiv:2102.12092v2.

[11] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B (2022). High-Resolution Image Synthesis with Latent Diffusion Models. ArXiv:2112.10752.

[12] Zhang, L., & Agrawala, M. (2023). Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543.

[13] Text elaborated by the research team of Annamaria Mercuri, University of Modena and Reggio.

CRATIVITY IN THE AI ART ERA. NEW SCENARIOS

Rebecca Pedrazzi – Iulm AI Lab / notiziarte

On October the 25th 2018 Christie's first auctioned a work of art realized with Artificial Intelligence: the portrait of "Edmond de Belamy", sold for 432,500 US Dollars.

Few months later, March the 6th 2019, Sotheby's put up for auction Mario Klingemann "Memories of Passersby I"[1], an innovative work presenting us with new aesthetic scenarios and the "morphing" characterizing the first artistic experimentations with A.I.

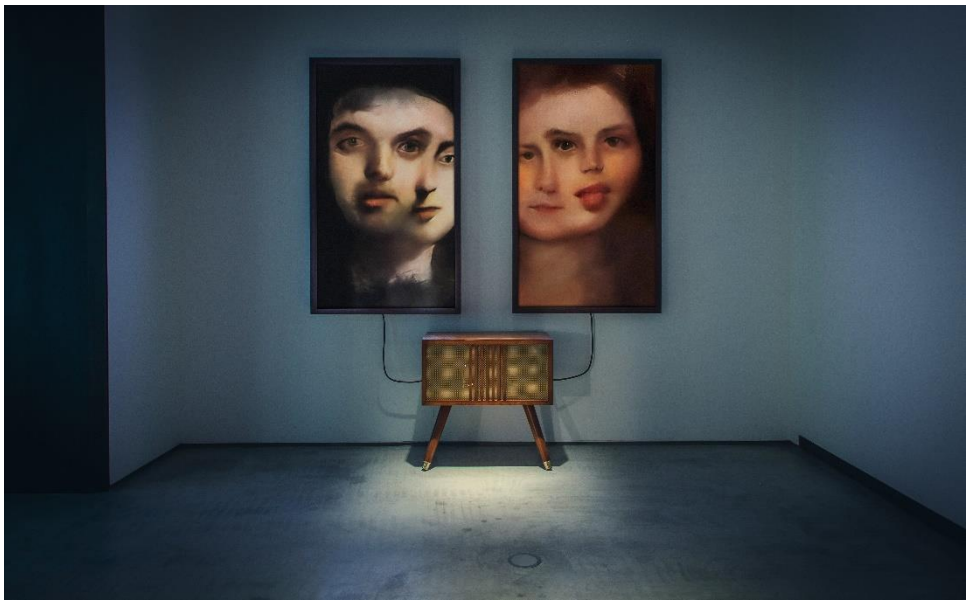


Fig. 1 - Mario Klingemann - Memories of Passersby I, 2018 - Courtesy of Onkaos.

And the question is: were we ready for the AI Art? Excluding professionals, the answer sounds negative.

[1] The work "Memories of Passersby I" consists of a tailor made cabinet in chestnut wood connected to two screens endlessly displaying people portraits generated through a complex system of neural nets and through a data set of thousands of portraits dating from the XVII to the XIX century. Composition of multiple GANs, two 4K screens, custom handmade chestnut wood console that hosts the AI brain and additional hardware. Wood console: 70 x 70 x 40 cm. Each 65" screen (3840 x 2160) custom framed: 145 x 82.9 x 3.8 cm (without frame) / 152.2 x 89.2 x 7.1 cm (with frame). Installation dimensions: 233 x 208 x 55 cm. Edition of 3 + 2 AP.

We are in a cultural historical time when the word "hype" is greatly headed on towards specific topics: the hype for metaverse, blockchain, NFT, AI tools such as text to image and GPT-3. And we have also great technological innovations not only in the world of Art but also in the world of Sciences. A.I. has led us to important advances in research, just think of its use to identify the molecular structure of proteins.

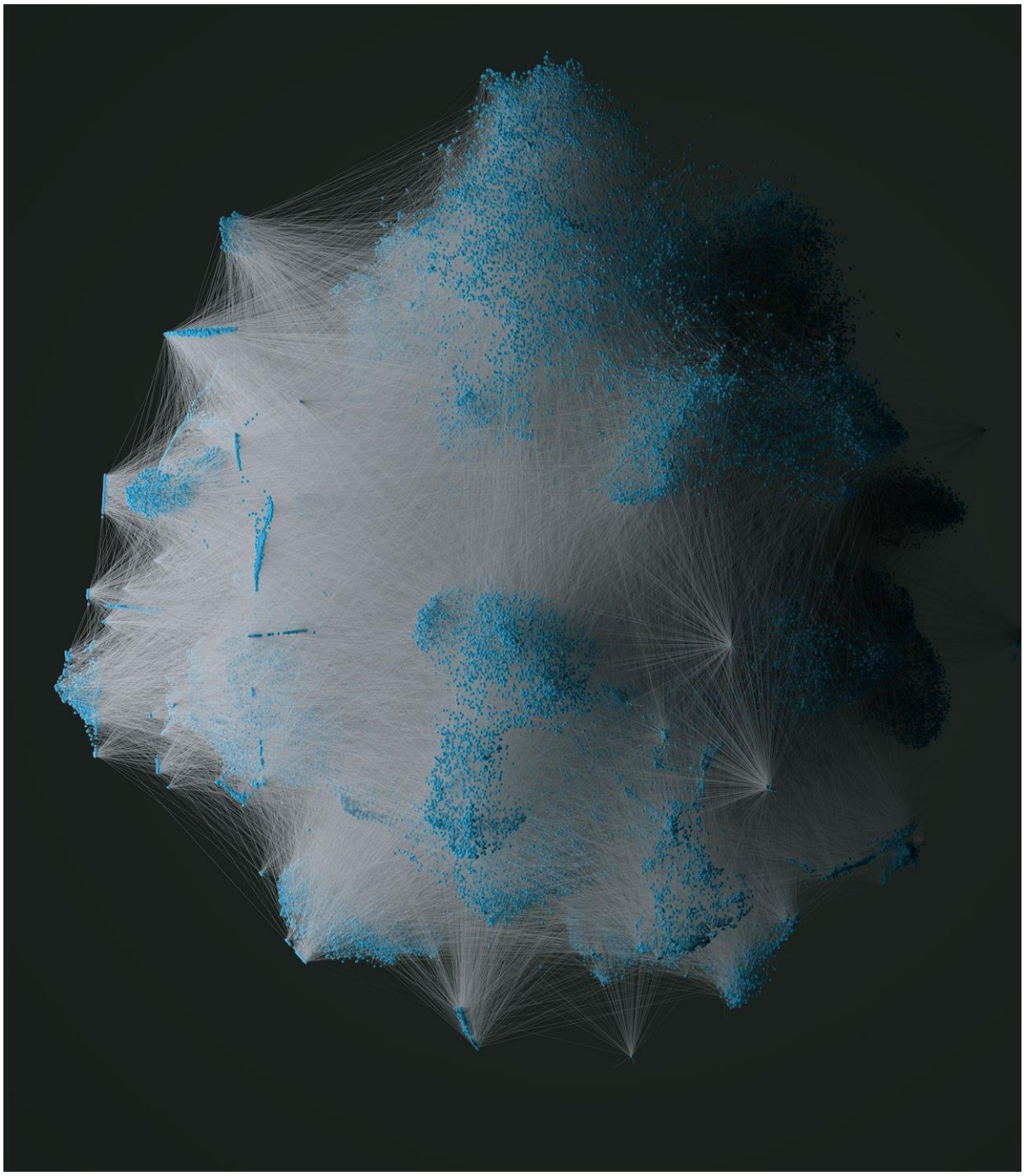


Fig. 2 - Mike Tyka e Refik Anadol, Archive Dreaming, installation view, Salt Galata, 2017. Courtesy Mike Tyka

As an art critic and historian, I first invite you to hold tight certain guiding concepts and to draw on our philosophical tradition and aesthetics: the most suitable means of interpreting the technological innovations in the world of Art.

Fig. 3 – Giuseppe Ragazzini, 8281 Lovers frame nft; courtesy of the artist.

We need to have a clear understanding of what a work of Art is, and in this respect, we can mention John Ruskin: "Fine art is that in which the hand, the head, and the heart of man go together". Art is the expression of an idea, of a thought. Within this thought there is all the story of the artist: culture, social time, experimentation, conscious and unconscious ego. Their life. To

Ruskin, the idea comes from the soul, from the heart. But to create the idea itself is not sufficient: a mental process is needed to shape life into reality, into a tangible form generally given by a manual or functional work. In this sense the term τέχνη, art in ancient Greek, indicates dexterity and skill: the art of the making. Nowadays we have a τέχνη 3.0: AI Artists are living a technological era, the era of web 3.0 and of the fourth industrial revolution, the era of a constant research and innovation.

Listen to Mike Tyka, a scientist and pioneer artist in the use of AI. He tells us:

"It was not the artists who invented the AI, it was the scientists, but then that's a new tool, a new thing you could use and the artists explored it" [1, p. 157].

As for art criticism and history, today we have the chance to deal directly with the artists. Surveying and interviewing the AI Artists is one of my firm commitments: I was the first in Italy to interview Mario Klingemann and Mike Tyka, and I think it interesting that during these interviews the word that kept being repeated was "creativity".

An artist graduated in philosophy, Giuseppe Ragazzini, tells us:

"The opportunities that such technologies and their many languages provide us with are a wonderful chance for artists. On this point of view, contemporary artists are living a unique time having at disposal amazing technologies to realize works that were previously unimaginable or requiring a huge economic cost" [2].

Vincenzo Marsiglia, a polyhedric artist who, coming from traditional mediums, got to present creative experiments such as Hololens 2, tells us: "Fifteen years ago, I started walking on with and versus technologies. This path represents an important part of my artistic production and it allows me to establish a relationship between my works and the users. In our contemporary world the artist must get to a rapport with technologies because they are a basic requirement, our actual creativity" [3].

Now let us imagine showcasing the history of Art, all the artists one after the other as if we were composing a jigsaw puzzle, one piece after another (as for A.I. we would refer to a Knowledge Graph). The first piece of the jigsaw is given by the work of a man who just using a stone drew a hunting scene on a cavern wall, and piece after piece we finally get to Botticelli's Primavera. We stop to admire this masterpiece, but if we want to understand its historical value, we must consider all the previous pieces of this jigsaw: Filippo Lippi, Pollaiolo, Verrocchio, and so on. Only in this way we will be able to comprehend why an artist has realized that work in that specific time and place.

Fig. 4 - Vincenzo Marsiglia, #UNRITRATTOPERUNIRCI, 4.4.2020, iPad, applicazione per iPad, proiettore sound ocrasunset, misure variabili - courtesy of the artist

Before getting to our present times, I would send out a "j'accuse" because many of these jigsaw pieces were ignored, and only recently it was given academic relevance to "the era of machine art" which actually started in 50s/60s when some mathematicians and artists began using algorithms to realize works of art. Just to name two of them: Frieder Nake, pioneer of the computer art whose first exposition dated 1965, and Vera Molnar, founder of the "Groupe de Recherche d'Art Visuel" and "Art et Informatique" in the first 60s, who only in 2022 was called for at the 59th Venice Biennial.

We also remind that it was January 1963 when the front page of the "Computers and Automation" magazine read: "The brush is an electron beam; the canvas, an oscilloscope; the painter, an electronic computer." And it was a month later when it was run the first "Computer Art Contest" on the topic of machine automatism.

Nowadays we have got to such a technologic level that we commonly speak about robots and machines applications to artistic realizations, but one of the first machines programmed to realize artwork in an autonomous way was AARON, created in 1968 by Harold Cohen. When asked "Is the computer being creative?", Cohen answered: "I think creativity is a relative term. Clearly the machine is being creative, the program is being creative. … I don't think it is currently as creative as me in writing the program" [4].

The machine may be creative – from the Latin "crĕare": to bear something from nothing – but already during the last century attention was fixed on the value of the human element. Questions were posed, some answers were given, about the role of the machine, about creativity in the sense "computer vs man".

To understand which way we have got to the AI Art in the modern sense, a piece of our jigsaw is represented by Desmond Paul Henry, a British artist pioneer to experiment with machine-generated visual effects who in 2001 created an aesthetic application of fractal mathematics. This way we can relate to the first experiments the AI Artists made in 2015 with DeepDream, a computer vision program created by Alexander Mordvintsev using a convolutional neural network to find and enhance patterns in images and to bring to the hallucinatory state proper of the fractal structures. Science and Art. Once more it was a scientist to register the potentialities of AI in the artistic world: Mike Tyka was among the first to explore them out of DeepDream and GAN[5].

Why up today so many artists use AI to produce artworks? Because today they have at their disposal a huge amount of data which they can use to generate their own data set. And being part of our present time, data can be classified as part of our history and culture.

Thus, though AI is bound to the future, artists are always linked with history and culture, and they focus on them as key features to explore new creative paths and aesthetics.

Back to data set, it is 2017 when Mike Tyka and Refik Anadol create Archive Dreaming, an installation of immersive projection employing machine learning algorithms to search and sort relations among 1,700,000 documents of the SALT Research collections.

And it is Refik Anadol who takes us to the discovery of the data sculptures by means of immersive installations such as his Machine Hallucination — NYC: a data universe of New York City created by deploying machine learning algorithms on over 100 million photographic memories of New York City. A new aesthetic is born: the dreaming vision of New York re-interpreted by and with AI.

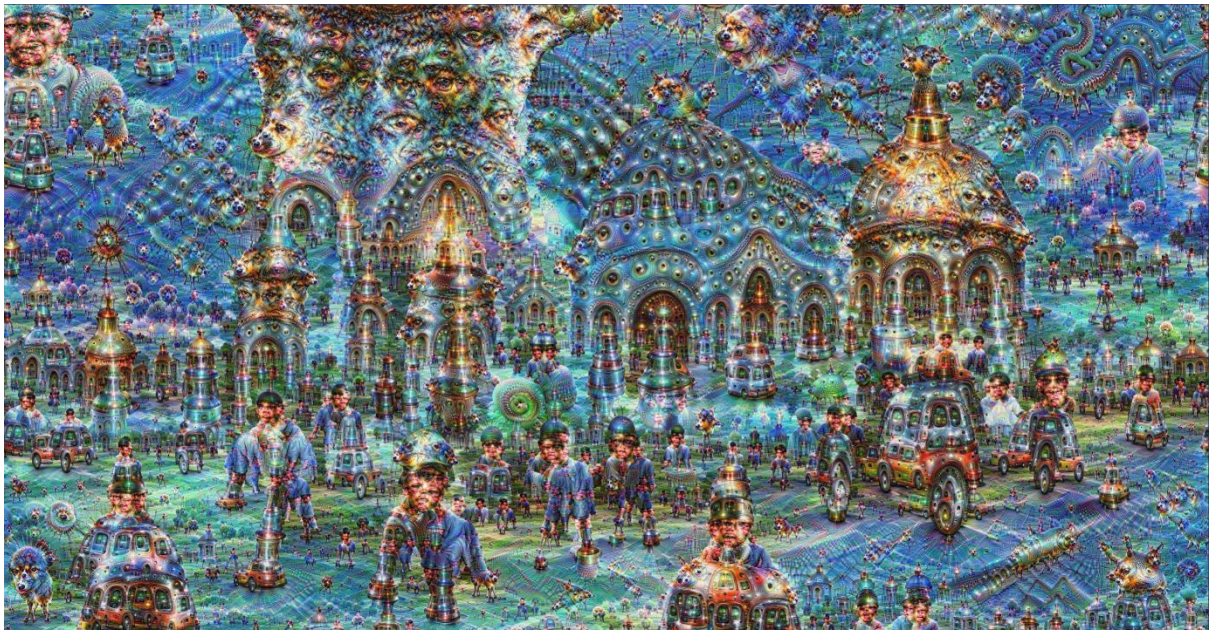


Fig. 5 - Mike Tyka - Die Ankunft, Neural net, Archival print, Edition of 10, 36x66" - 2016 - courtesy of the artist.

Not long ago we could not have easy access to instruments employing AI to create images, texts or music, but thanks to tools such as text to image, text to music or GPT-3 we have started to use AI even daily. With such tools we can generate images by writing a text, anyhow this way we do not generate an artwork: to make art the idea is needed, the creative process is needed, and consigning the making of a work of art to the mare description of an image is not at all sufficient.

On this topic we have valuable reflections of the artists themselves: Andrea Crespi says: "Technology must be at the service of the artist to help him express his creativity at its best and can't be used as a shortcut to delegate the machine with your own work and personal research. You have to be the artist and not to pull up the artist if you want to be part of history" [6].

This is a reason to mention the immersive site-specific work "Mapping the NFT Revolution" by Mauro Martino (AI Artist e Principal Research Scientist at the MIT-IBM Watson AI Lab where he is founder and manager of the Visual Artificial Intelligence Lab): this project, presented in September 2022 at Meet Digital Culture Center in Milan, reveals how just by a text to video the artist got to realize images of artistic beauty and suggestion.

Fig. 6 - Image from: Mapping the NFT Revolution by Mauro Martino - In collaboration with Andrea Baronchelli, Luca Stornaiuolo, music by Philip Abussi . Courtesy MEET Digital Culture Center | Fondazione Cariplo.

And what will the artists be doing after AI has become a tool accessible to everybody? They pass over and start experimenting with something not yet accessible. Mario Klingemann tells us: "This home field has been explored and now everybody is doing it. … with this new project I hope I'll go on working by myself in some little solitude" [1; pag. 243].

At this point me too, as an art historian, I have got to give my answer to: AI, is it creative?

And my answer is: yes, positive, I think machines are creative insofar they have already created many things in so many and different fields from books to music, but I can't compare the creativity of the machine with the creative process of the artist: behind the creation realized by the machine, by the GAN, by AI in general terms, there is a complex creative process involving the human element, the human factor which is the *conditio sine qua non* for the creation given by the machine itself.

Up today the machine is not a sentient being that – eureka! – creates all by itself something such as images or music melodies. A machine, a GAN not trained on a data set chosen and given by a person – the human factor – neither creates images nor music melodies

Up today only the creative process of the artist offers the magic of depicting an idea and make a work of art, while AI [7] may just take us to discover new horizons, may arouse our curiosity

leading us to constant, new aesthetic research: new studies on digital pigmentation have already been started, as well as investigations on the use of the light for AI artworks.

As for the human factor, there is another point to underline: while presenting the use of the latest technologies, algorithms, potential and limits of the data set, AI Artists also participate major reflections on themes of great relevance such as the Climate Change, and this also comes within the historical and social evidence of Art.

Everything is evolving, recently we observe a new trend in the world of AI [8]: some traditional artists would experiment with these super technologies and collaborate with data scientists on artistic projects. This leads to new synergies between artists and robotic tools, and to happiest experiences such as the case of the IA artists Davide Quayola and Sougwen Chung.

By way of conclusion, I would like to mention John Dewey: "Science states meanings; art expresses them". My personal call is to be always armed with a critical and philosophical mind: this way we will be able to handle all the instruments needed to comprehend the great technological revolution leading to these new scenarios in the world of Art.

References

[1] R. Pedrazzi, *Futuri possibili. Scenari d'arte e Intelligenza Artificiale*, Jaca Book, 2021.

[2] R. Pedrazzi, *L'intervista a Giuseppe Ragazzini: tra sperimentazione, filosofia e Intelligenza Artificiale*, in notiziarte, 4 ottobre 2022, https://notiziarte.com/2022/10/04/lintervista-a-giuseppe-ragazzini-tra-sperimentazione-filosofia-e-intelligenza-artificiale/

[3] R. Pedrazzi, Tra installazioni e Intelligenza Artificiale: l'intervista all'artista Vincenzo Marsiglia, in notiziarte, 27 luglio 2022, https://notiziarte.com/2022/07/27/tra-installazioni-e-intelligenza-artificiale-lintervista-allartista-vincenzo-marsiglia/

[4] *Harold Cohen and AARON: Ray Kurzweil interviews Harold Cohen about AARON*, in Can Computer create?, 1987, https://www.computerhistory.org/revolution/computer-graphics-music-and-art/15/231/2306

[5] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio (2014). "Generative Adversarial Networks" – https://arxiv.org/pdf/1406.2661v1.pdf

[6] R. Pedrazzi, *Intervista ad Andrea Crespi: tra sperimentazione, percezione e rivoluzione digitale*, in notiziarte, 10 gennaio 2023, https://notiziarte.com/2023/01/10/intervista-ad-andrea-crespi-tra-sperimentazione-percezione-e-rivoluzione-digitale/

[7] M. A. Boden, L'Intelligenza Artificiale, il Mulino, Bologna 2019

[8] A.I. Miller, The Artist in the Machine. The World of AI-powered Creativity, The MIT Press, Cambridge MA 2020

ARTISTIC EXPRESSION THROUGH AI AND THE ARRIVAL OF THE NEW RENAISSANCE

BLAC.ai

The concept of the New Renaissance is one that we are all experiencing as we go through the revolutionary process of technology and human expression. As we navigate through this disruption, we face challenges as a global community, particularly with AI's bad brand that has been promoted by Hollywood for decades. In reality, people are using AI to express themselves through the human creative process in ways never done before. In this publication, we will explore the historical relevance of this topic, highlighting examples of my journey, and exploring how AI and artistic expression mark the arrival of the New Renaissance.

Globally, we are in a state of massive transition from one thing to another, and we are currently in a global state of chaos. Whether it be war, Art, relationships, communication, or the economy, everything is changing, and it is pretty crazy. Undoubtedly, in a few years to come, things will continue to change. Within this chaos, I believe that we have three major attributes contributing to this global change: AI, the introduction of Web3, and the way it operates, which is a wild arena of things that no one can quite wrap their head around. We have a lot of global changes where governments and economies are in a state of somewhat chaos, but also humanity is too in terms of how we communicate and how we work. It is something that we need to keep in perspective when we go through these phases and understand where we are coming from versus where we are going, while still trying to stay in the understanding of what it is to be in this moment right now.

We are in a state of disruption, and the classical definition of disruption is a radical change to an existing industry or market due to technological innovation (Fig. 1). This is the literal moment that we are involved in right now; and AI is at the pinnacle of it. It is about to disrupt almost every industry on the planet. I have been involved in many things in my career, and this is the largest catalyst for change that I have ever been a part of. As someone involved in the community of artists, with the focus on Art, self-expression, and contemporary fine Arts, this is an unbelievable process. We are in the middle of creating the New Renaissance, with a disruptive force not seen in some time.

Disruptive Art is progressive Art. The cycles that we have gone through as humans on this planet and in the universe as a whole have been from chaos to order through Art. We are a byproduct of this, and our consciousness operates in a level of chaos and order as well.

Something becomes traditional, something becomes standard, and then something disrupts it. That disruption then progresses us further, which is the state in which we are in right now. This occurred with the original Italian Renaissance, and we can see how it has happened over and over again throughout history.

Fig. 1 We are in a state of disruption.

One of the most disruptive Art movements was the Dadaist movement, which is important to speak about in the context of AI. Dada essentially created and put out into the ethos of the Art world the idea that because something is said to be Art, it is Art. This is extremely relevant when we are speaking about the way that Art and creative tools are being utilized to create things. The core of that is breaking expectations, and that is the road we are on with AI and Art.

Surrealism breaks the mindset of what is able to be created, and abstract expressionism questions the nature of fine Art. The rise of street Art is also significant, as it challenges the status quo legalities and makes people think about Art in a new way. Every single time new and progressive Art challenges the status quo, it makes people hesitant and fearful, but it also makes them embrace it and move forward in a way that has never been seen before. This creates the next disruptive Art movement that eventually becomes mainstream and traditional, and the cycle continues.

We are in the middle of this cycle right now, and there has always been backlash every single time. It is not only focused on photography or photoshop, but it is also seen throughout history. For example, during the Renaissance, sculpture Artists were upset about realism painting as a style because they felt as though it removed them from the market. It is important to keep this in mind as we continue to progress with AI and new creative tools that progress Art forward.

In the AI Art community, everyone has a unique story and background. My background is as an Artist and entrepreneur, having worked with hundreds of companies and gained a natural instinct for recognizing new ideas. When I was exposed to AI Art a year ago, I saw the potential and created my own proprietary AI. Over time, the technology evolved, and I focused on pushing its limits and breaking conventions. Through my own expressive purpose, I have achieved a level of expression that pushes the boundaries of what is possible with AI. These pieces are not solely created by AI, but infused with my creativity and expression (Fig. 2).

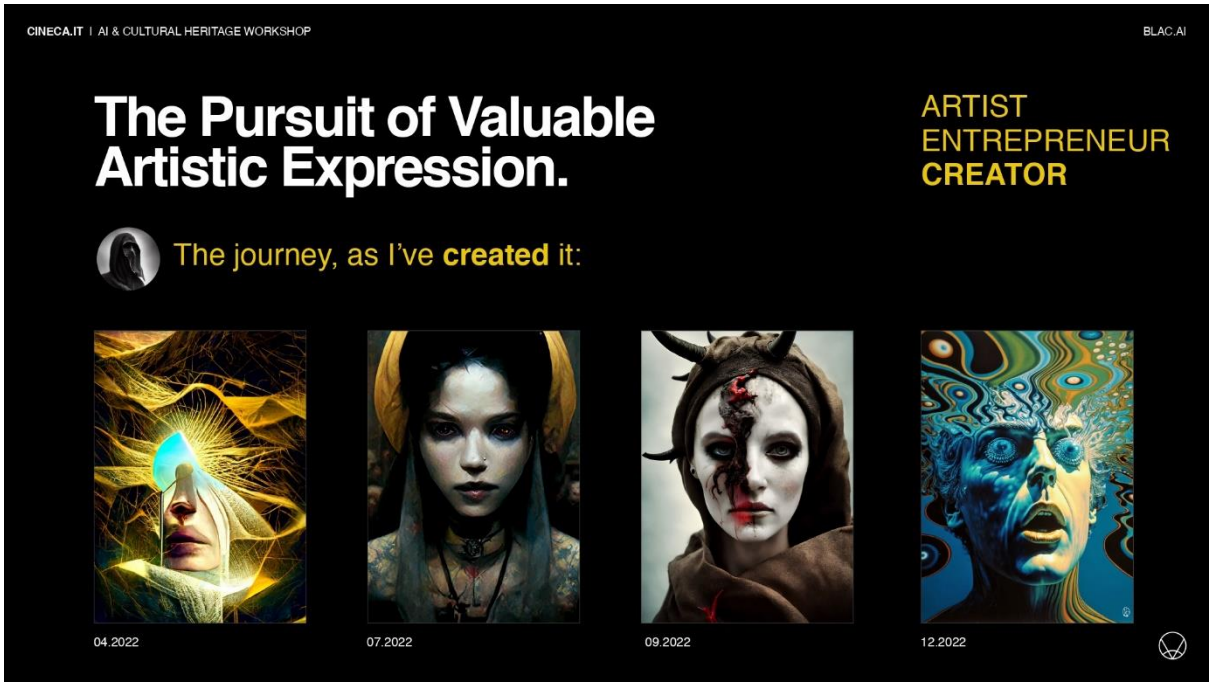


Fig. 2 The pursuit of valuable artistic expression

AI Art tools are allowing people to express themselves in a way that has never been done before. Real-time feedback with the creative process is possible, which is a massive shift in the way Art is produced and consumed. With these tools, Artists can flush out an entire concept for a series in mere weeks, whereas traditionally it could take months or years. AI is infusing itself into various programs, and almost every creative tool will have this type of technology embedded in it in the

next couple of years. Web3 and AI combined create a marker in birthing a new process for the way humans express themselves in a creative fashion, challenging the accepted norms of traditional Art forms. The purpose of the Renaissance was to change the way Art operated, and it changed everything, and right now, AI is disrupting the Art market by increasing the amount of volume and scalability of purchasable Art. It is important to note that when talking about expressive Art, it refers to Fine Art and Contemporary Art, not just a phone app that gives a new profile picture. Art is difficult to define, and there have been debates for centuries over whether AI is Art or not.

I define Art as: Art = Intentional Expression. Anything that is expressed with intention is Art, whether it be a painting that took a hundred hours to complete, an AI-generated piece done in five seconds, or even throwing a banana on the ground in social protest. As humans, we have the superpower to call anything we create with intentional expression "Art". Once you identify yourself as an artist, you have the power to create Art regardless of tools, medium, race, beliefs, purpose, or value. The work produced with intentional expression alongside AI by those who identify themselves as Artists is 100% Art. However, not everything created with AI is Art, and not everything produced in this new disruptive phase is Art. It is important to distinguish pure human creative expression from creating things just for the sake of seeing them. Not everything people create is Art, it is a delineation that is important to make. As we move through this next phase of human experience and the way we create things, it is crucial to remember the importance of intention behind the Art we produce.

We need to understand that just because it is easy to access the tools, it does not mean it is easy to master them. These tools require mastery and experience. Everyone has their own process, and we all have different ways of achieving our goals as artists (Fig 3; Fig. 4).

The real-time expressive feedback and scaling that we can achieve using these tools are something that we have not seen before. The feedback is so fast that it increases scalability and artistic expression, something that has not happened before. The expressive power has been enhanced, but this is on a completely different level. We are at a very interesting moment in Art in general, and this is what I believe to be the New Renaissance.

The why is more valuable than the how in creating Art. What is the reason behind creating what you are creating? When everyone has the ability to create perfect compositions and photorealism, and all these things without spending thousands of dollars on materials or tooling or programming– the WHY matters. The reason behind what we create becomes the value in my opinion. We need to know the artist's story, why they are creating what they are doing, and what

their narrative is (Fig. 5). The story behind the Art and why the Art affects the viewer is what establishes value to the Art.

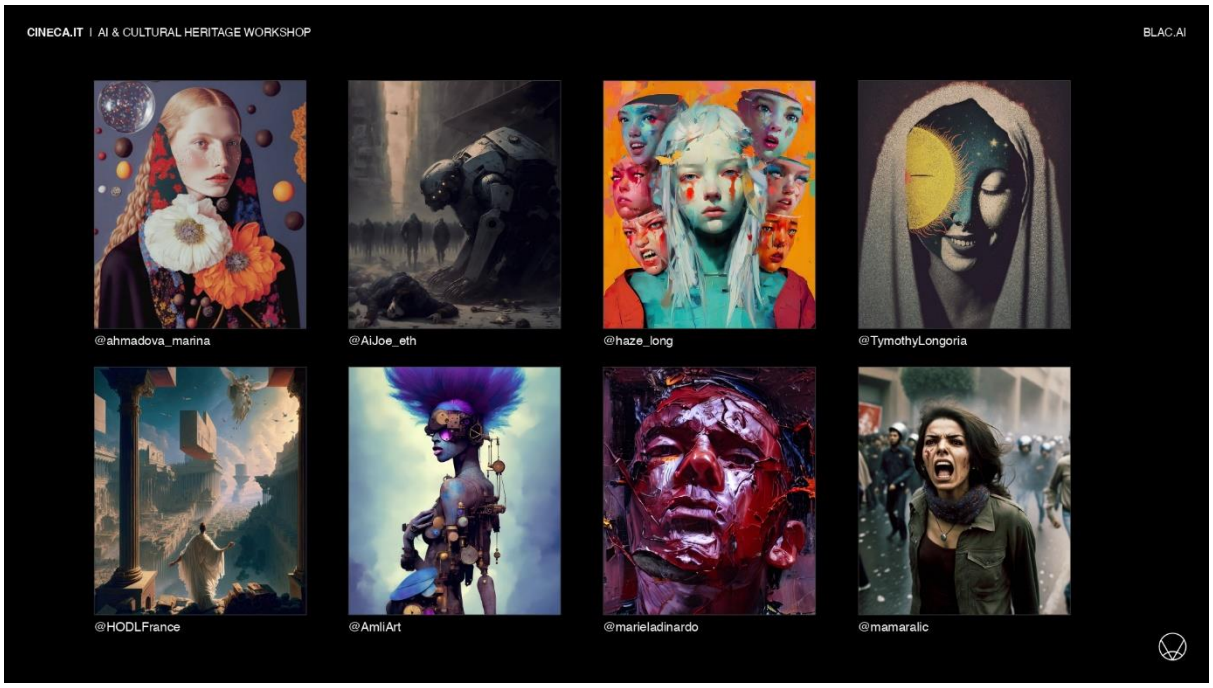


Fig. 3 Works by @ahmadova_marina, @AiJoe_eth, @haze_long, @TymothyLongoria, @HODLFrance, @AmliArt, @marieladinardo, @mamaralic

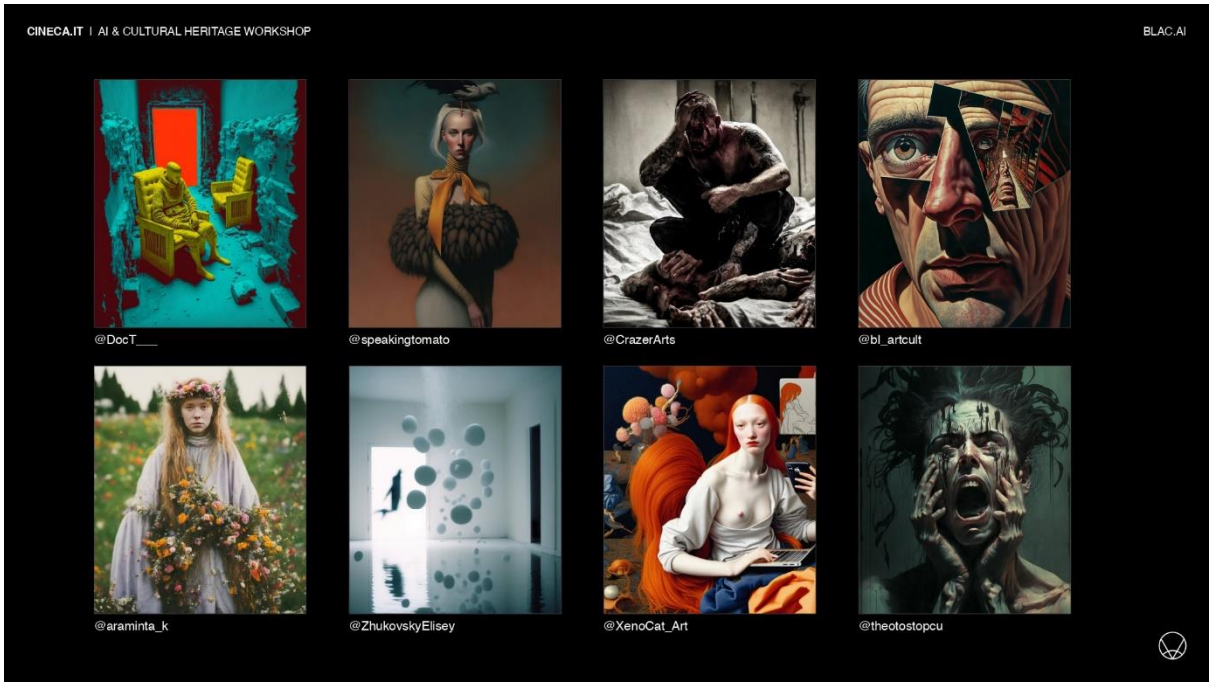


Fig. 4 Works by @DocT___, @speakingtomato, @CrazerArts, @bl_artcult, @araminta_k, @ZhukovskyElisey, @XenoCat_Art, @theotostopcu

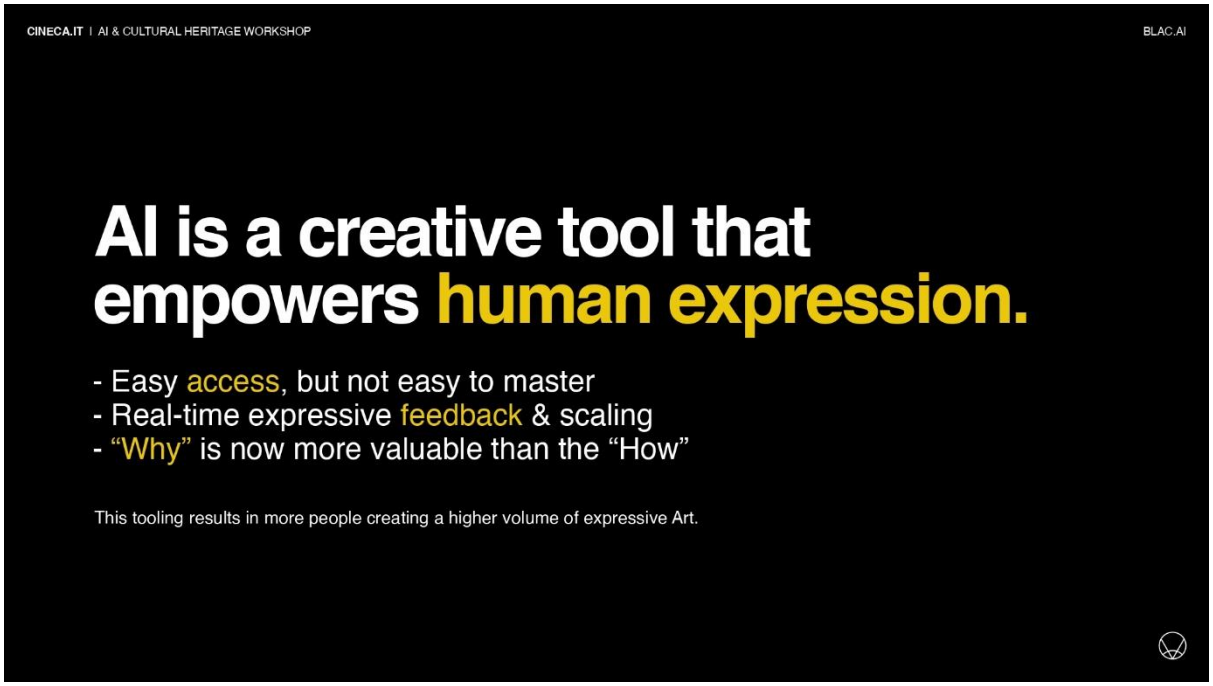


Fig. 5 AI is a creative tool that empowers human expression.

In conclusion, AI Art is indeed Art and is part of the New Renaissance that we are creating in real-time. The disruptive nature of AI Art is characteristic of previous Art movements that have paved the way for progressive artistic expression. AI Art tools are changing the way Art is produced and consumed, increasing volume and scalability, and disrupting the art market. The definition of Art is not limited to traditional mediums, and AI Art is an excellent example of how technology is shaping the way we express ourselves. The impact of AI Art on the Art industry is far-reaching, and it is a significant catalyst for the new wave of creative production.

ACKNOWLEDGEMENTS

The organization of the workshop and the publication of these proceedings would not have been possible without the help of Veronica Bortoluzzo, Simona Caraceni, Patrizia Coluccia, Luca Govoni, Margherita Montanari, Andrea Valiani, Paolo Vendruscolo, Elisabetta Zanette, Paolo Zuzolo.

ISBN - 9788886037457

DOI: 10.1388/WORKSHOP-AICH-01

CINECA