Studi di estetica, anno LII, IV serie, 1/2024
ISSN 0585-4733, ISSN digitale 1825-8646, DOI 10.7413/1825864661

Mario Verdicchio

Ekphrasis and prompt engineering.
A comparison in the era of generative AI

Abstract
In this work I try to show that there is a common thread between ekphrasis *and prompt engineering, in spite of the temporal distance of almost thirty centuries between their respective earliest known instances, and despite the blatant disparity in technological tools supporting them. An investigation of the characteristics of such thread can shed light on a number of aspects of our experience of textual and visual works, including the role of our imagination, that of our emotions, and what happens when human agency within these imaginative and emotive processes is substituted by the artificial agency of an AI system.*

Keywords
Ekphrasis*, Prompt engineering, Generative AI*

Received: 04/03/2024
Approved: 12/03/2024
Editing by: Alessandro Nannini

mario.verdicchio@unibg.it

1. *Introduction*

An attempt to compare *ekphrasis* and prompt engineering may appear to be odd, to say the least. On the one side, we have the rhetorical and literary exercise of describing an artwork with text, a practice that dates back at least to the eighth century B.C.E., when Homer described Achilles' shield in the epic *The Iliad*, the oldest among the surviving works of ancient Greek literature. On the other, we have the technique of organizing text to be fed to an Artificial Intelligence (AI) software to generate output in the form of more text, sounds, static and moving images, a practice that is becoming more and more widespread after the release of such software in the 2020s.

Do the two endeavors have anything in common? Does it even make sense to build a discourse that includes them both? Despite the temporal distance of almost thirty centuries between their earliest known instances, and despite the blatant disparity in technological tools supporting them, I will try to show that there is a common thread between *ekphrasis* and prompt engineering, and that an investigation of the characteristics of such thread can shed light on a number of aspects of our experience of textual and visual works, including the role of our imagination, that of our emotions, and what happens when human agency within these imaginative and emotive processes is substituted by the artificial agency of an AI system.

In the next section I will illustrate the main principles and technological components that constitute an AI system to start building a comparison with what happens when people get involved with *ekphrasis*.

2. *AI and generativity*

Artificial Intelligence is a subfield of Computer Science born in the 1950s, when a group of scholars proposed to give the status of a discipline to their efforts in modelling human learning and reasoning processes by means of operations in a computer (McCarthy *et al.* 2006). From a technological perspective, there was no significant distinction between the endeavors of the pioneers of AI and those of other computer scientists: they used the same kind of machines, in which their inserted finite sets of deterministic instructions, also known as programs, and from which, in turn, they obtained results. If there was a difference, that was in the goals for which computers were used: AI researchers were focusing on

programs that were meant to solve problems traditionally tackled by human beings. Indeed, in the two decades following the foundation of AI, great efforts were made in what we call automated reasoning, that is, the programming of a computer so that it can automatically perform the deductions from a set of hypotheses to the relevant logically entailed theses. This kind of computation is characterized by a top-down approach, from general principles to particular conclusions, governed by the rules of logic that computer scientists have programmed into a computer. Thus, there are no surprises in these operations, no unforeseen deviations from the rigid path marked by deductive principles.

In the same period in which automatic reasoning was being developed, other scientists started using computers in a radically different way. The beginning of Computer Art can be traced back to the 1960s, when three computer scientists began, independently of one another, to use digital computers to create geometric designs: Michael Noll at the Bell Labs in New Jersey, USA, George Nees at the Erlangen premises of the Siemens company, and Frieder Nake at the University of Stuttgart, both in Germany. Nake, in particular, is considered the first person ever to show the output of a computer in an exhibition: it happened in 1965 at the now defunct bookshop "Wendelin" in Stuttgart.

One of Nake's most famous works, titled "Random polygons", best embodies the paradigm that characterized the works of the early computer artists: it consists in an abstract print of a broken geometric line connecting various points scattered on the sheet, forming polygons. The positioning of the points is where the technique used by Nake departs from traditional programming: the x-y coordinates of the points are established *pseudo-randomly*. In computer science, we talk about pseudo-randomness when numerical values are established by means of a complex parametric function whose behavior a human programmer is able to set up and kick off, but not follow through because it is determined by a number of parameters (e.g. the milliseconds on the computer clock when the function's computation is launched) that are not visible nor immediately accessible. Once exposed to the output of such a function, not even the person who programmed that function into the computer is able to trace back the calculations that determined the resulting values, which thus appear to be random. This is not real randomness, since all the inner operations of a computer are deterministic by construction: this is just a seeming randomness that relies on an epistemic lack on behalf of the programmer.

Computer programmers have never been supposed to know everything: indeed, their ignorance of the solution to a problem is the very

reason why they resort to programming a computer in the first place. However, relying on pseudo-random number generating functions marks a departure from traditional programming because programmers relinquish to the computer a significantly larger part of their control over the operations. A computer artist like Nake, while programming with the aim to create his random polygons, had only a rough idea of what the output would look like, and would not have been able to know *a priori* exactly where the points and the lines would end up in the final composition. Indeed, he claimed authorship only of the general schema of his work, whereas he considered his computer the generator of the specific instance that was printed and shown in an exhibition (Nake 2012).

This creative approach is at the core of what is called Generative Art. There is no unique definition of this endeavor, but the existing ones can be roughly divided into two categories on the basis of their main focus: the above-mentioned epistemic lack on behalf of the programmer on the one side, and a form of autonomy attributed to the computer on the other.

An example of the first kind of definition is provided by philosopher Bringsjord and colleagues, who deem an artifact *A* designed by a human *H* as creative when, at end of a regular, error-free process, it outputs a result *r* such that *H* cannot explain how *A* produced *r* (Bringsjord *et al.* 2003). An example of the second kind is given by artist and art theoretician Galanter, who defines as generative a system that, under the guidance of an artist, operates with a certain degree of autonomy to produce or contribute to an artwork (Galanter 2003). Neither definition is fully satisfactory when it comes to computers.

The epistemic lack pointed at in the first definition can be eliminated by slowing down the computation by means of a step-by-step execution at a speed that is compatible with human perception and by making all the parameters involved in the computed function visible to the programmer with additional output instructions. In Nake's case, he could have predicted the exact coordinates of the points in his artwork, had he been able to observe the intermediate results of every computational step performed by his computer. Of course, such procedure would make the creative endeavor much more time-consuming, but it shows that this characteristic of generativity is more an issue of practicality than of definitory ontology.

Resorting to the concept of autonomy as in the second definition comes with another layer of conceptual issues built on top of the epistemic lack. Indeed, the two definitions are related: deeming an artifact autonomous is based on our lack of knowledge on its inner workings.

Typically, a highly automated artifact is designed to relieve humans of various tasks by executing them with minimal intervention thanks to the reliance on sensors to convey input from the external world into the artefact, internal processors for the computational elaboration of such input, and actuators to send the output back to the external world. If the required intervention is indeed minimal, then human users will not control the operation of these artifacts at all times, thus missing a comprehensive understanding of the environment in which the artifact functions. As a consequence, the behavior of such an artifact is often unpredictable in its details. If the artifact is properly designed and built and it brings the assigned task to completion in an unpredictable way, then its users will tend to attribute it autonomy. This analysis seems to boil an autonomous system down to a highly automated one that operates on behalf of epistemically lacking humans. Speaking of autonomy when it comes to artifacts may then seem inappropriate, since it appears to be much more similar to the automation we build in machines than the autonomy we attribute to humans (Johnson, Verdicchio 2017).

It is not my intention here to criticize the definitions of Generative Art. Despite their shortcomings, they are coherent in shedding light on a relinquishment of control by a human agent in the form of a delegation of execution, by means of properly written instructions, to a computing artifact. Differently from traditional logical programming, those instructions do not specify every detail: like in the case of pseudo-random numbers, some of them may trigger a pre-programmed function not fully known to the programmer, which ensures that the final output is characterized by some elements of surprise and unpredictability that, in turn, shroud the artifact with an illusory veil of originality and creativity.

This illusion has only gotten stronger in recent years, thanks to advancements in Machine Learning (ML). ML is considered a subfield of AI, but its paradigm is so radically different from that adopted by the AI pioneers in the 1950s that it may be considered an independent subfield of Computer Science on its own. If traditional AI stems from a logic-based approach to capture human reasoning in the form of computational operations, ML relies on a biologically inspired approach, according to which computers are to mimic the functionality of a human brain by running a complex network of simple mathematical functions, each simulating the workings of a single human neuron. The first relevant mathematical model, called artificial neuron, dates back to the 1940s (McCulloch, Pitts 1943) and eight decades later, thanks to the advent and evolution of

digital computers and telecommunication networks, Internet users can enjoy the fruits of this research.

In abstract hierarchical terms, ML is the exact opposite of traditional AI: AI goes top-down from universal principles to particular entailments, whereas ML works bottom-up from a large number of specific data values to general patterns. If AI is based on logical deduction, ML relies on statistical induction: data points from the outside world are examined in search for correlations, and the strongest correlations rise to the status of patterns that get stored in terms of numerical factors inside a network of artificial neurons. When the significant patterns from a set of data are detected and stored inside a neural network, we say in ML terms that the network has learned, and is now ready to process new batches of similar data. This is what happens, for instance, when a digital camera is able to detect faces in a scene and frames them on the display screen: the computer inside the camera runs a neural network that has been trained to detect patterns of pixels corresponding to facial features (e.g. eyes, nose, mouth, etc.), including those relative to the contour lines of a face, around which the computer adds the artificially generated pixels of a frame on top of the pixels showing the real-world scene captured by the camera's image sensor.

ML has been the key contributing factor to the latest developments in Generative Art, and since in spite of the above-mentioned difference in paradigms it is still considered AI, ML is the driving force of what is today known as Generative AI. Not all AI is about art, and there still are generative artistic endeavors that do not use AI; however, the most notable recent achievements in Generative Art, including those that lend themselves to a comparison with *ekphrasis*, are indeed AI-enhanced, so I will focus on generative AI.

The concepts used to define what constitutes a generative system, i.e. ignorance on behalf of humans and autonomy attributed to artifacts, are enhanced when AI is involved, because the relevant programming techniques significantly widen the context that lies outside the control of the user. In other words, compared to Nake exploiting the pseudo-random numbers in his computer, a user of an AI-based generative system today knows much less about what goes on inside the machine, and this is due to some characteristics that are intrinsic to this kind of AI.

Firstly, computation is not carried out by a single computer programmed by the person who is using it, but it is distributed among a number of computers connected through the Internet, containing neural networks trained on millions of data gathered online. This means that the

final output that the person in front of a computer receives depends only in small part on what is executed on that computer, whereas the majority of the computational work is carried out or, more precisely, has already been carried out on computers to which the person does not have access.

This previously completed work is the training of the neural networks prescribed by the principles of ML, and the data used for it must come from the context in which the AI system is supposed to work: like a computer in a digital camera has been trained on images of faces, so computers that are to work with music and sounds need to be trained on data relative to soundwaves, and computers that are meant for text composition and elaboration are going to be fed textual data.

These three examples represent the main media categories that are currently elaborated by computers (i.e. textual, sonic and visual) because past endeavors in physics and electronics have given us devices (i.e. keyboards, microphones, loudspeakers, cameras and monitors) allowing us to encode these media, that is, to describe them in numerical terms to be given to a computer as input, and to decode them, that is, to reconstruct them on the basis of the numerical output of a computer. Experiments with tactile and osmic experiences are still at a very early stage: that is why, for instance, we are not able to feel the smoothness of a fabric or smell the notes of a perfume by means of a computer online.

Numerical data from media that are amenable to encoding are employed to train neural networks in the sense that they are analyzed in search for statistically significant patterns. Depending on the media, such patterns correspond to specific shapes or visual styles (e.g. the abovementioned facial features or Van Gogh's signature brushstrokes), kinds of sounds (e.g. the soundwave a piano produces has a very specific shape), and sequences of words (e.g. in the English language a subject is very likely to be followed by a predicate). As already said, once these patterns are learned by a network, they can be used to process new data of the same sort. Such processing can be categorized into two main tasks: classification and composition. Classification means that new data are categorized on the basis of patterns learned from previously processed data. This type of task has been around for some time: other than face recognition on cameras we have image clustering on smartphones (Villalba *et al.* 2015), song recognition by means of apps (Wang 2006), and language detection in online translators and speech recognition (van Lieshout, Cardoso 2022), among others. Composition means that the learned patterns are used to synthesize new data. This is where the latest, most notable results in Generative AI research have been obtained. Given that the field

is rapidly evolving, and more and more players are joining, market dominance may quickly change sides; however, at the time of writing (February 2024), these outcomes revolve around one company, namely, OpenAI. OpenAI is a US-based AI research company founded in December 2015, which is generally acknowledged for the resurgence of ML-based AI research thanks to the 2022 release of ChatGPT, a chatbot (i.e. a conversational software) based on the company's neural networks trained on large natural language datasets gathered from the Internet (Marr 2023). For the purposes of this work, I will first focus on another software from the same company, called DALL-E, which exploits the same computational language model as ChatGPT, but for a different purpose: generating new synthetic images in correspondence to a caption provided by its user as input text (Marcus, Davis, Aaronson 2022). In what follows I will draw a comparison between what a DALL-E user does when they want to generate an image and *ekphrasis*. I will start with the subtle similarities between these two endeavors and then I will contrast them with the much more blatant differences, with the aim of shedding light on decades-long debates on agency, subjectivity and *ekphrasis* itself.

3. *Ekphrastic writings and generative prompts*

There is another reason why Generative AI users nowadays know even less about what happens inside their machine than the pioneers of Generative Art with computers. A software like DALL-E exploits computational language models at its interface with human users, to enable them to provide input in the form of sentences in their language. This important feature puts Generative AI software also at the forefront of another subfield of Computer Science, that is, natural language processing (NLP), aimed at the construction of computational artifacts that work with language in a way that is meaningful and contextually relevant to their human users. This means that if Nake operated his computer by means of a programming language that reflected the inner workings of the machine in terms of numerical values and operations, all a Generative AI user needs to make their system work is the language they use every day. No programming skills are required: even a lay person without a clue on how a computer works can operate such a system.

However, typing sentences into a computer does not allow for the same freedom that we have in conversing with another human being: despite a significant facilitation in interaction, we are still dealing with a

computing machine, after all. Thus, *prompt engineering* has recently emerged as a critical aspect of all endeavors in Generative AI: it deals with the crafting of precise and effective natural language instructions and queries given to an AI system to generate desired outputs (Bozkurt, Sharma 2023).

Prompt as a term has undergone a slight semantic change in this context. In the tradition of Computer Science, a prompt is a symbol, message, or command indicating that the system is ready to accept input, whereas in education, a prompt is a specific instruction or question given to students to guide their thinking or writing. The latter meaning has spread into Generative AI, and today the most common interpretation of prompt is not a symbol on the screen encouraging users, but the very input given by the users to the computer. Still, a common thread remains: a prompt as a symbol on a screen is a way for a computer to encourage a user to provide input to the computer, and an input for the machine is also, with a change of perspective, an output of the user's mind; similarly, a prompt as an instruction for a computer is how a human sets the artifact's internal mechanisms in motion for the production of a certain output. This overlap of meanings is where I look for grounds to build an analogy between prompt engineering and *ekphrasis*: they are both a textual mean to stimulate an agent to create.

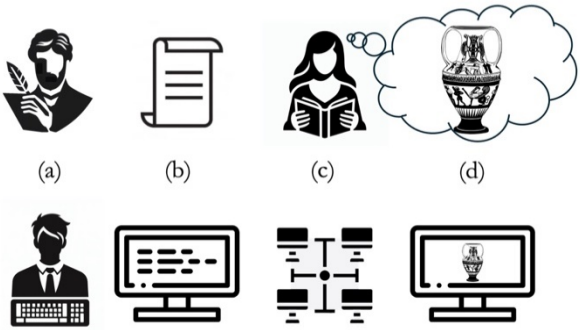

Figure 1: a comparison of elements in *ekphrasis* and in prompting a generative AI system

Figure 1 illustrates the main elements in the two endeavors, *ekphrasis* above and generative AI below, with a visual organization that aims at outlining this comparison.

(a) There is a human *composer* on both sides: a writer or a poet in the former case, a computer user or prompt engineer in the latter.

(b) The *composition* is textual in both cases: we have an *ekphrastic* passage in a novel or in a poem above, and a prompt in input to a computer below.

(c) There is a *generator*, a human who reads the *ekphrasis* on the one hand, and a network of computers running a generative AI system that receives the prompt on the other.

(d) Finally, the generated *outcome* is the goal for which the composition, a writing and a prompt, was created in the first place: a mental visualization of an artwork in the mind of the reader and a digital image in the computer that queried the generative AI, respectively.

Given the parallelisms in this functional framework, can writing prompts for a Generative AI system be considered a new type of *ekphrasis*? A number of objections come immediately to mind, especially in connection with the nature (human versus artificial) of the agents involved, but in the following I will try to show that an analysis of the counter-arguments is actually useful to support this apparently daring analogy.

First of all, there is a stark contrast between the two composers: we have none other than Homer on one side, and a common computer user on the other. Of course, although the passage describing Achilles' shield is considered "the founding instance of *ekphrasis* in Western literature" (Heffernan 2015: 38), the *Iliad* is not the only work of reference for *ekphrasis*. Still, widening the timeframe of observation does not seem to usher in any person without highly developed writing skills: from the Roman imperial times of Philostratus' *Imagines* (Philostratus 1931) to the Medieval times of Chaucer's *The book of the Duchess* (Chaucer 1982), to the modern ode *On a Grecian urn* (Keats 1819), and to contemporary poetry like *In Santa Maria del popolo* (Gunn 1958) or novels like *Kiss of the spider woman* (Puig 1976), these partly or fully ekphrastic efforts are all remarkable literary works.

How can a prompt engineer even compete? Indeed, the prompts one feeds to a Generative AI system (e.g. "Cityscape painting during a rainy day, focusing on reflections in puddles with a mix of soft and harsh brush strokes", or "Victorian-era painting of a masquerade ball with elaborate costumes, contrasting colors and soft lighting") cannot be considered works of art under any light. The striking disparity between the composers in the two contexts is reflected in the nature of their compositions. However, we gain a different perspective when we take the recipients of these compositions into account. The historical examples of *ekphrasis*

that I have mentioned are works intended to be read by human beings, and for the ekphrastic process to be complete, the minds of these readers must be stimulated by the words of the literary compositions so that their imagination allows them to form mental images of the artworks, or specific details of them, described in the text. None of this holds in the case of a Generative AI system.

Despite their affinity with certain characteristics of the human mind such as logical reasoning, symbol processing, and arithmetic operations, even the most technologically advanced AI devices remain inherently distinct from the mental dimension that underpins the human experience. In the same decades in which computer technology and AI were developing, significant debates took place in the context of philosophy of mind about the nature of the relations between these devices and the human faculties they were supposed to recreate or at least simulate. The arguments proposed by the most eager opponents of the idea of a *strong* AI, capable of generating essentially human faculties such as consciousness and subjectivity within a computational machine, still stand today: given how computers are constructed, in the form of digital circuits crunching numerical data, they offer no foothold for the belief that they are or will ever be able to entertain meanings in their circuits (Searle 1980) and therefore acquire experience and common sense (Dreyfus 1992). In simpler and more prosaic terms, writing for computers is very different from writing for human beings.

This does not mean that no talent is needed to write efficacious prompts. Computers may not participate in our semantic dimension, but the whole AI endeavor, and also some of its intersections with telecommunications and knowledge engineering like the Semantic Web effort, is devoted to compensate such lack in semantics with a sophisticated syntactic approach. In this technological context, semantics refers to a much more restricted, operational interpretation: the only meaning computers can entertain is expressed in terms of the effect of their operations in the state of the system in which they work. Here, semantics is synonymous with computer behavior (Goguen 1975). This is why it is called prompt *engineering*: writing a prompt is not just simply using your natural language, but a complex technique in which sentences must be formed in a way that reflects how the learned data and patterns are organized inside the Generative AI system so to stimulate the generation of results that best reflect what the user had in mind. Fine-tuning techniques, which typically involve iteratively adjusting the wording, structure and context of prompts based on the system's responses and performance, and

template-based approaches, based on predefined patterns or structures filled in with relevant information or contextually appropriate content are only some of the most representative techniques aimed at guiding the composition of prompts. *Mutatis mutandis*, learning prompt engineering is like attending a course in literature or creative writing.

The most evident contrast between what happens in *ekphrasis* and Generative AI lies in the generator, that is, the agent that is exposed to a textual composition and is thus stimulated to perform a visualization of sorts. On the one hand there is a person with their imagination, their experience, their taste; on the other, a system of computers with neural networks trained on all sorts of numerical data. It is very challenging to find commonalities under the light of the above-mentioned objections to strong AI. Moreover, an explanation of the subjective experiences that we have as beings with consciousness, a fundamental ingredient of our imaginative and creative faculties, is universally considered the hardest problem in neurosciences and philosophy of mind (Chalmers 2017). Without any clue on how to reduce the ontological divide between a human and a computer, the only perspective we can adopt to draw a comparison in the generation of a visual output from a textual description is the one provided by AI: if computers are excluded from experiences of any kind, the relevant syntactic descriptions will have to do. The activity of extracting patterns from data performed by neural networks serves this very purpose. For example, while a Generative AI system cannot visually perceive a nose in the same way a human does, through training its neural network with digital images containing faces and noses, it can gradually develop numerical representations of pixel configurations corresponding to nose features. Subsequently, the system can employ these representations to generate new images or modify existing ones. Despite the seemingly insurmountable differences in cognitive mechanisms between humans and AI, advancements in AI research enable a prompt such as "the painting of the Mona Lisa, only with a much larger nose" to elicit similar visual outcomes from both entities.

The disparity between a human generator and an artificial one is reflected also in the nature of the outcomes: in *ekphrasis* the result is a very personal mental visualization, whereas in a Generative AI system it is a digital image on a computer screen, ready to be printed or published online. We have a subjective experience in the former case, and an objective (also in the material sense of an object) production in the latter. If a human generator wanted to share with others the mental experience they acquired from the *ekphrasis*, they would have two possibilities: if

they have a certain artistic aptitude, they could create a sketch of their mental imagery, thus somehow manually mimicking the automated visual production of the Generative AI; if drawing is not an option, then they can only resort to describe their experience with words, thus reversing the ekphrastic process and going back to a textual form, of which we would then have two instances, the original composition and the newly minted generated outcome. Both options show that an attempt to share an ekphrastic experience does not fit the framework of traditional *ekphrasis*, which seems to be brought to completion in the subjective mind of the human generator. This, in turn, appears to undermine the analogy between *ekphrasis* and prompt engineering I am attempting to draw: how can they be related if one leads to a mental experience and the other to a digital image?

In the next section, I will shift the focus from the modalities to the purposes for which a poet or a writer may want to engage in *ekphrasis* in search for more commonalities with Generative AI.

4. *Purposes and caveats*

So far, I have been comparing the endeavor of *ekphrasis* with the latest, most advanced creative technologies only on the basis of what is performed on both sides and by whom or what. Apart from the hazard of involving AI in a topic that is usually tackled in literary studies, my endeavor may be arguably seen as a lighthearted attempt to follow the tradition of describing the formal components that define what *ekphrasis* is. Now, in an effort to avoid or, rather, set aside the obstacle of a mismatch in generated outcomes, I intend to follow the lead of Goldhill (2007) and steer my analysis towards the purposes for which *ekphrasis* exists as a literary device. My intent is the following: if I am able to show that *ekphrasis* and Generative AI may be seen as sharing a very significant goal, perhaps the role of the means by which such goal is pursued by these two endeavors will become less important. This, in turn, could mitigate the concerns about differences in the outcomes generated by each effort, thereby strengthening the analogy I am seeking to establish.

Among the many contributions in Goldhill's paper, the one that is most significant in the context of my work is the claim, backed by numerous references to scholars from the first century C.E. like the Alexandrian sophist Aelius Theon and the Roman educator Marcus Fabius Quintilianus, that a fundamental theoretical framework for comprehending *ekphrasis*

is provided by rhetoric. These rhetoricians proposed *ekphrasis* as a key skill to be taught to novice orators, as it hinges on a crucial rhetorical concept dating back to Aristotle known as *enargeia* (not to be mistaken for *energeia*), the ability to evoke vivid imagery (Aristotle 1932). The goal of the *enargeia*-laden *ekphrasis* is to immerse the audience so deeply in the narrative that they almost become viewers, with "almost" being a particularly critical keyword, since an exercise in rhetoric does not necessarily lead to factual observation. Indeed, the *enargeia* in ekphrastic speech is meant to penetrate the deepest emotions of the listener to the point of overpowering and subduing them. A powerful mental visualization triggered by an orator who has mastered *ekphrasis* can amaze, but when such astonishment is very intense, it may bypass critical intellect and, thus, enable the speaker to conceal facts and compel passive emotional experiences to the listeners. Rhetoric suggests a more profound understanding of visualization as a tool to manipulate the audience's emotions, something that Goldhill considers overlooked in contemporary discussions of *ekphrasis*.

This perspective could potentially put *ekphrasis* in conflict with fundamental principles that have long influenced aesthetics, such as Plato's critique of art's diversion from ultimate truth (Plato 2016) and Kant's cautions regarding art's manipulation for persuasive ends (Kant 1987). These exemplify an enduring link between aesthetics and truth, not merely from a logical or epistemic standpoint, but also as an ethical imperative. In a morally questionable twist, arguably with the posthumous blessing of ancient rhetoricians, *ekphrasis* and Generative AI converge towards the goal of truth-independent persuasion.

From the aspirational perspective of persuasion that goes straight to eliciting emotions circumventing rationality and intellect, using texts or images makes little difference, especially considering that what matters is the ultimate goal of affecting and manipulating a person. In this sense, there is a way to close the gap between *ekphrasis*, which involves a human when the textual composition is read, and Generative AI, in which no person is involved except for the prompt engineer: the generated visual outcome needs to be shown around. Indeed, if the purpose of a synthetic image is to support a rhetorical exercise, it does not make sense for it to be displayed only on the monitor of the user who initiated its generation. This is where AI and communication intersect, a crossroads that has witnessed a series of highly poignant events in recent years, revolving around the concept of truth, which is a key ingredient of the fundamental function of communicative processes: transmitting, from person to

person, a faithful depiction of reality. However, such adherence to reality is not always guaranteed, as exemplified by a rhetorical device like irony or, pushing boundaries to their limit, plain deception.

In this context, the election of the 45th President of the United States stands out as a pivotal moment, with expressions such as "fake news" and "alternative facts" storming into public discourse. While "fake news" already existed as a term, its frequent use to dismiss critical media reports has surely elevated its visibility and importance. Conversely, "alternative facts", coined by a communication consultant to justify the Press Secretary's exaggerated claims on the size of the crowd at the President's inauguration, exemplifies a more recent invention aimed at evading inconvenient truths. In such a climate, the audience is compelled to question the veracity of the media they consume, especially when Generative AI has risen to challenge perceptions of reality even further with the so-called *deepfakes*.

The depth in their name is not related to the gravity of the deception they enable but, more prosaically, to the number of structural levels of the neural networks used to generate them: these networks are deep in the sense that they need several layers where to learn and store all the patterns that numerically describe a person's physiognomy. Deepfakes may lend themselves as technological embodiment of the most powerful version of rhetorical exercise, and they are already considered instrumental in a new kind of conflict that has encompassed and surpassed the concept of cyber warfare: since an increasing number of people use social networks as their main source of news (Kim, Dennis 2019), we are now talking of a post-truth war (Enroth 2023). At the moment, deepfake technology is very sophisticated but not to the point of being undetectable by adequately trained neural networks: Generative AI is now in an arms race against itself, aiming for deception on the one side and detection on the other (Moy, Gradon 2023). These are indeed bizarre times, whose intricate entanglements of technology and power would require a truly masterful ekphrastic narrator to be described fully.

5. *Experiments and conclusions*

I want to conclude on a less distressing note. The descent into a post-truth hellscape of my attempt to show that *ekphrasis* and Generative AI are tightly related through prompt engineering may itself be deemed an exercise in rhetoric. Indeed, I used two tricks: I openly declared one, and

hid the other. I have already admitted that this potentially catastrophic identity between Generative AI and *ekphrasis* under the auspices of rhetoric is driven by a "goals over means" strategy taken to the extreme. This does not erase the difference in outcomes that I tried to ignore: *ekphrasis* emotes by means of text, Generative AI by means of images. This difference may turn out to be useful in a different discourse that is connected to the post-truth scenario and the rhetorical trick I have kept hidden so far. I lied by omission, not questioning at all the lasting power of *enargeia* even in a world engulfed in deepfakes. After the release of two Generative AI systems that are even more powerful than DALL-E, because they create videos and not just images from textual prompts, one by Google (Google Research 2024) and one by OpenAI (OpenAI 2024), many reactions on social networks were in line with the following, by a user called "da5ha": "With lumier [*sic*] and sora just released, I no longer trust any videos tbh [to be honest]. The ai revolution is going to break us". This comment paves the way to a scenario that seems to be more reasonable than the one where people keep getting emotionally manipulated by fake images and videos endlessly. Lies work only when they occur rarely: a stream of digital images with no connection to reality will eventually lose any rhetorical power, because it will become meaningless in a world where audiences have stopped paying attention. If anything will be broken by the AI revolution it won't be us, but the connection between images and truth that was consolidated by the advent of photography. Freed from the commitment to reality (or to a fake version of it), digital image generators may return to the artistic endeavors of their origins. Would there be a connection with *ekphrasis* still? I believe so, and here follow some experiments to shed light on what such a connection may be like.

First of all, it seems reasonable to foresee that Generative AI's main focus is always going to be images rather than texts, at least judging by the current state of a text-based tool like ChatGPT. I tested ChatGPT's knowledge around *ekphrasis* by asking the following question: "Do you know what *ekphrasis* is?" to which the AI replied: "Yes, *ekphrasis* is a literary device where a description of a visual work of art is used in literature. It's a vivid and often elaborate verbal description, aiming to create a mental image or evoke the same feeling as the original artwork".

This is a very satisfying reply, which we can imagine is also rather easy to implement, by means of a search and retrieval of some definition published online. I went on and tried to test ChatGPT's ability to apply such definition to a brand-new situation. For that, I made a rather daring request: "Generate an ekphrastic poem, describing the wonders of a city

built by composer Wolfgang Amadeus Mozart". The answer was a rather trite bunch of rhyming lines, relating to very generic ideas of music without expressing any particular characteristic of Mozart's work: "In Mozart's city, the streets sing, / With melodies that dance and ring. / Each cobblestone, a note so sweet, / In harmony, the city's heartbeat. (…) / So let us wander, hand in hand, / Through streets of music, where dreams expand. / In Wolfgang's city, where music reigns, / A place of wonder, free from chains".

To double-check, I made the same kind of request, this time with Gershwin instead of Mozart, to see whether there would be any difference guided by the distance between the two composers. What a disappointment. "In Gershwin's city, the rhythm's sway, / Where melodies paint the light of day. / Each street a stanza, each building a song, / In harmony, the city throngs. (…) / So let us wander, hand in hand, / Through streets of music, where dreams expand. / In George's city, where music reigns, / A place of wonder, free from chains".

The nadir was, without a doubt, the last stanza that was identical to Mozart's version, with just a swap in first names. These outcomes clearly show the parametric nature of the algorithm governing the generation of ekphrastic poems in ChatGPT. Why such subpar execution by a Generative AI that has taken the contexts of text composition, translation, and summarization by storm? When it comes to AI systems based on neural networks, the most likely explanation is lack of training: *ekphrasis* has not left a significant enough digital trace on the Internet to enable a proper training of ChatGPT. There are not enough data points for the syntax to capture the patterns of the semantic acrobatics required by an ekphrastic poem.

If sticking to the context of textual data will not ensure an interesting interaction with AI around *ekphrasis*, perhaps we need to return to image generation and look for ways to exploit this technology without steering towards the dangers of a deepfake-infused rhetoric. Interestingly, it seems like to keep the system away from harmful fakes, we need to maintain a safe distance from any commitment to reality. Indeed, *ekphrasis* has always dealt with a good kind of fakes, in terms of fiction depicting fictional artworks. Rosand calls such textual visualizations "twice-fictive" images, and deems giving them life a task for poets. Such work is not unilateral, because historically the favor has been returned and a reciprocal exchange between poets and painters has been established in a cycle of texts and pictures that the scholar assumes as a perspective from which to analyze Western art (Rosand 1990). If Generative AI is to assume a role in this cycle, it should be that of painters since my experiments with automated ekphrastic poetry were not promising at all. This was the

premise of my last experiment: going back to DALL-E and trying to use it as a painter of poetry. Since the rather rigid prescriptions of prompt engineering would impose too many constraints, I threw all precautions away, and inserted as prompt the dada poem *Alas!* by Picabia (1918) as it is, without any modification to make it more compatible with the traditional optimization strategies, except for eliminating the line breaks to have the whole poem in one prompt. The input was as follows: "Women men Affairs A country ambitious For sovereignty I love it when someone folds the eyes Of troubles Especially in the sea of the thorax But I'm telling disinterested lies It's almost the same thing The soul's truth Is the great cowardice of academic arrogance Looking into your eyes I'm content In my forgotten solitude".

As usual, DALL-E provided four different generated outcomes, which are shown in figure 2.

Figure 2: DALL-E's generated outcome with dada poem "Alas!" as prompt

I will leave to the reader the task of trying to map some of the words in the poem onto the visual elements in these synthetic images. I found the exercise very stimulating: it prompted me to read the poem with more attention and it enabled me to discover or, rather, interpret new, deeper meanings that I had not entertained the first time I approached the text.

In conclusion, is prompt engineering a new kind of *ekphrasis*? Perhaps not, since we obtained the most interesting results out of a Generative AI

system when we deliberately ignored its rules. However, it is the endeavor of prompt engineering, with its focus on natural languages, that brought Generative AI closer to *ekphrasis*, creating a common ground where the text-image intermedial cycle can keep going, with the chance of encompassing the latest, cutting-edge technological endeavors. Whether this is an opportunity or a risk is, despite the highly automated context, up to us. In this sense, number-crunching syntactic machines are not supportive in our standing search for semantically meaningful experiences.

Bibliography

Aristotle, *Rhetoric*, Engl. transl. L. Cooper, New York, Appleton - Century - Crofts, 1932.

Bozkurt, A., Sharma, R.C., *Generative AI and prompt engineering: the art of whispering to let the genie out of the algorithmic world*, "Asian journal of distance education", 18/2 (2023), pp. I-VII.

Bringsjord, S., Bello, P., Ferrucci, D., *Creativity, the Turing test, and the (better) Lovelace test*, in J.H. Moor (ed.), *The Turing test: the elusive standard of artificial intelligence*, Dordrecht, Springer, 2003, pp. 215-39.

Chalmers, D., *The hard problem of consciousness*, in M. Velmans, S. Schneider, (eds.), *The Blackwell companion to consciousness*, Malden, Blackwell Publishing, 2017, pp. 32-42.

Chaucer, G., *The Book of the Duchess*, ed. H. Phillips, Durham, Durham and St. Andrews Medieval Texts, 1982.

Dreyfus, H.L., *What computers still can't do: a critique of artificial reason*, London, MIT press, 1992.

Enroth, H., *Crisis of authority: the truth of post-truth*, "International journal of politics, culture, and society", 36/2 (2023), pp. 179-95.

Galanter, P., *What is generative art? Complexity theory as a context for art theory*, "Proceedings of generative art" (2003), http://www.philip-galanter.com/downloads/ga2003_paper.pdf (last visit: February 2024).

Goldhill, S., *What is ekphrasis for?*, "Classical philology", 102/1 (2007), pp. 1-19.

Google Research, *LUMIERE A space-time diffusion model for video generation* (2024), https://lumiere-video.github.io (last visit: February 2024).

Gunn, T., *In Santa Maria del popolo*, "Poetry", 93/1 (1958), p. 1.

Heffernan, J. A. W., *Ekphrasis: theory*, in G. Rippl (ed.), *Handbook of intermediality – Literature – Image – Sound – Music*, Berlin, De Gruyter, 2015, pp. 35-49.

Johnson, D.G., Verdicchio, M., *Reframing AI discourse*, "Minds and machines", 27 (2017), pp. 575-90.

Kant, I., *Critique of judgment* (1790), Engl. transl. W.S. Pluhar, Indianapolis, Hackett Publishing, 1987.

Keats, J., *On a Grecian urn*, "Annals of the fine arts" (1819), pp. 638-9.

Kim, A., Dennis, A.R., *Says who? The effects of presentation format and source rating on fake news in social media*, "MIS quarterly", 43/3 (2019), pp. 1025-39.

Marcus, G., Davis, E., Aaronson, S., *A very preliminary analysis of DALL-E 2*, arXiv:2204.13807, Apr 25, 2022, https://arxiv.org/abs/2204.13807 (last visit: February 2024).

Marr, B., *A short history of ChatGPT: how we got to where we are today*, "Forbes", May 19, 2023, https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/ (last visit: February 2024).

McCarthy *et al.*, *A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955*, "AI Magazine", 27/4 (2006), pp. 12-4.

McCulloch, W.S., Pitts, W., *A logical calculus of the ideas immanent in nervous activity*, "The bulletin of mathematical biophysics", 5 (1943), pp. 115-33.

Moy, W.R., Gradon, K.T., *Artificial intelligence in hybrid and information warfare: a double-edged sword*, in F. Cristiano *et al.*, (eds.), *Artificial intelligence and international conflict in cyberspace*, London, Routledge, 2023, pp. 47-74.

Nake, F., *Construction and intuition: creativity in early computer art*, in J. McCormack, M. d'Inverno (eds.), *Computers and creativity*, London, Springer, 2012, pp. 61-94.

OpenAI, *Creating video from text* (2024), https://openai.com/sora (last visit: February 2024).

Picabia, F., *Alas!,* in Id., *Poèmes et dessins de la fille née sans mère*, Lausanne, Imprimeries Réunies, 1918.

Philostratus the Elder, *Imagines*, Philostratus the Younger, *Imagines*, Callistratus, *Descriptions*, Engl. transl. A. Fairbanks, Cambridge (MA), Harvard University Press, 1931.

Plato, *Republic*, Engl. transl. A. Bloom, New York, Basic Books, 2016.

Puig, M., *El beso de la mujer araña*, Barcelona, Seix Barral, 1976, Engl. transl. *Kiss of the spider woman*, London, Vintage Books, 1980.

Rosand, D., *Ekphrasis and the generation of images*, "Arion: a journal of humanities and the classics", 1/1 (1990), pp. 61-105.

Searle, J.R., *Minds, brains, and programs*, "Behavioral and brain sciences", 3/3 (1980), pp. 417-24.

Van Lieshout, C., Cardoso, W., *Google Translate as a tool for self-directed language learning*, "Language learning & technology", 26/1 (2022), pp. 1-19.

Villalba, L.J., Orozco, A.L., Corripio, J.R., *Smartphone image clustering*, "Expert systems with applications", 42/4 (2015), pp. 1927-40.

Wang, A., *The Shazam music recognition service*, "Communications of the ACM", 49/8 (2006), pp. 44-8.