I castelli di Yale online. ANNALI DI FILOSOFIA
Vol. XII, n. 2, 2024, pp. 43-54 – ISSN 2282-5460

# What to Do with AI Images?
## Towards an Epistemology of Trust[1]

**Alessandra Scotti**

**Abstract.** *The aim of this article is to explore, within the field of visual culture, the ethical and social challenges raised by AI images, the reference imaginaries involved, and the expectations of meaning that arise. The paper will pursue the following line of argument: first, it will examine images that reflect forms of artificial intelligence according to the categories of operative image (i) and invisible image (ii), to investigate the agentive nature of AI images by asking "what do they want from us?" (Romele & Severo 2023) and relate these images to the critical literature that reflects on the concept of "image opératoire". To elucidate the ways in which technology and the imaginary mutually nourish each other, I will draw on Castoriadis and conceive AI imaginary as both instituted and instituting (Castoriadis 1975), so as to claim a "thoughtful" nature to AI images. This removes them from the stereotypical and numbing forms in which they are usually presented and allows their instituting as well as instituted nature to be imagined in a way that can enable trust in them. This process opens the door to an ethics of the visible that is deeply rooted in the aesthetic regime and situated on an affective level, so to speak.*

**Keywords.** AI, Images, Visual studies, Imaginary, Castoriadis.

**Alessandra Scotti** is Postdoctoral Research Fellow at the University of Turin. She was previously Postdoctoral Research Fellow at the University of Naples Federico II, Research Fellow at the Istituto Italiano per gli Studi Filosofici and at the Center for Advanced Studies - Southeast Europe of the University of Rijeka and Adjunct Professor of Bioethics. She is also a member of the editorial board of the journals «S&F_scienzaefilosofia.it» and «Chiasmi International». Her main research interests are contemporary French philosophy, with a special focus on Merleau-Ponty, philosophical anthropology, environmental and visual studies.
**EMAIL:** alessandra.scotti@unito.it

## 1. Introduction

The purpose of this article is to investigate, within the framework of visual studies, the ethical and social challenges posed by AI-generated images,[2] the imaginaries they conjure, and the expectations they fulfill. As Hal Foster (1988) has already noted, the visual does not only concern visual data, but also its *discursive nature*. In other words, what we see,

[1] I would like to thank Prof. Romele and the Irméceen (Institut de Recherche, Médias, Cultures, Communication et Numérique) for the warm welcome I received during my stay at the Sorbonne Nouvelle University as well as Prof. Lingua and the entire Erato (Ethical Research Center on Anthropology and Technology) team for the discussions and the trust they placed in me. Finally, I would like to thank the anonymous referees for carefully reading my manuscript and for their constructive comments, which contributed significantly to improving the quality of the paper.

[2] In this article, I will only deal with static images, as I have considered databases such as GettyImages and Shutterstock. I plan to analyse platforms such as HeyGen, Synthesia, and DeepBrain AI – which allow individuals without video editing skills to effortlessly create generative AI videos based solely on a script – in a future, as I believe that the inclusion of dynamic visual media could enrich the analysis and reflect the diversity of AI's visual output.

what we are able to see or what we are permitted to see each time requires a redefinition of the dialectic between the visible and the invisible inherent in every scopic regime and relates to both the politics of the image and the imaginary of a given society. This rooting of an image in gaze, and in the media that define vision, calls upon a historical dimension of the image: each era does not only have different styles, arts, and images, but also perceives them differently, which, in turn, feeds and constructs new imaginaries. So, the question is: how should we represent artificial intelligence, and what philosophical categories should we use to outline the technological imaginary associated with it?

Although the ethical debate on AI is vast and inherently interdisciplinary – involving philosophers of law, bioethicists, neuroscientists, educators, and "pure" philosophers – and although the tasks required of an AI ethicist are often numerous and unrealistic (Cocchiaro et al. 2024), the scholarly debate on AI images is virtually nonexistent. There is a significant gap in AI literacy regarding the issue of AI images; Romele (2022) rightly speaks of a blind spot.

One might ask why the image itself poses a problem. But this question naturally evolves into how to build public trust in AI and how to construct an epistemology of trust and belief through its visual representation, especially when an image, as the saying goes, is worth a thousand words. In this article, I will explore why the issue of what images we use to represent AI raises numerous philosophical questions, ranging from aesthetics to ethics, from philosophy of technology to politics. The structure of this paper is as follows: after a brief examination of the common features of the images we use to represent AI, the section following this introduction will attempt to decode the current socio-technological imaginary – particularly its homogenized and stereotypical features – by reflecting on the concept of the image itself, highlighting its agentic function on the one hand and its invisible nature on the other. In the second section, the agentive nature of images will be emphasized and a tradition that refers in particular to anthropology and art history will be referenced, so as to highlight the typical function of the image to «distribute the sensible» (Rancière 2013). This function will prove not to be limited to the vector of representation. Ultimately, this will also allow us to emphasize that images are never "aseptic" in their relationship to us, but always the historical product of underlying power relations. In the subsequent section, in view of their invisible nature, AI images will be considered in the context of an eye-to-machine process, in which the image becomes increasingly detached from its sensory nature. As a result, the reference imaginaries become self-reinforcing, triggering a process of progressive exclusion of the human gaze.

AI images' invisibility allows the discourse to move from a purely reductionist level – where the image is supposed to represent the thing – to a much more complex one, anchored in the affective sphere and related to beliefs and expectations of meaning, as well as to a trust-building effort, which attempts to go beyond a simple and sterile appeal to transparency. In other words, we do not trust a technology simply because it is presented to us transparently; rather, this trust emerges from an epistemological process of "public construction of trust", which also involves visual education. In this context, the theoretical reference to Castoriadis and his theory on the instituent as well instituted character of imaginaries in the last part of the paper is instrumental in highlighting their ability to identify something in the reference imaginary that unifies and sustains society, while acknowledging that this imaginary must remain dynamic to retain its meaning and avoid becoming sclerotized.

A very brief overview of images of AI – for instance, through a simple Google image search – reveals their markedly stereotypical nature, raising several theoretical questions:

1. What common characteristics define the images typically used to represent AI?

2. To what extent do these images distance us from AI, or even hinder our understanding of it? Consequently, how can we develop a visual education that fosters understanding rather than reinforcing this distance?

First of all, we can see that the majority of these images is anthropomorphic. This should not come as a surprise. As with non-human animals – anthropology teaches us that the earliest images depicted animals – they represent the other, the removed, the unsettling, and serve a specific function in the structuring of a society. For this reason, if every social, religious, political, and cultural identity in the act of its foundation provides for the exclusion of an otherness, then the establishment of human primacy corresponds to the disparagement of animals. This forgetting of the beastly goes hand in hand with the primal fear of animal vengeance, a recurring leitmotif that resonates throughout the obscure social and cultural rules encapsulated by the terms totem and taboo. This theme extends beyond primitive forms of society and even reaches the contemporary era, in certain forms.

The highly anthropomorphized images of AI are probably a somewhat clumsy attempt to make sense of them by attributing human traits to them. After all, we anthropomorphize anything to which we ascribe intelligence (machines, gods, aliens, etc.).

Let us consider a concrete example to substantiate this claim. The news section of the European Parliament's website features an entire article discussing the EU's AI Act as the first significant piece of legislation on artificial intelligence. The AI-generated image accompanying the article is entitled Limitless Visions and comes from Adobe Stock. It is a profile shot of two men facing each other against a blue background onto which blurred lines of source code are projected. The man on the left – who should represent AI – resembles the man on the right in every way. He is wearing a shirt, just like his human counterpart, and is depicted with an open hand, which in gestural language signifies helpfulness, loyalty, and authenticity. The AI is completely analogous to its human counterpart, except for the face, which is transparent, drawn by the contours of bright neural networks.

One consequence of the process highlighted in this example is that images which resemble humans reflect the distortions, prejudices, and cultural habits embedded in our society. Several studies (Cave et al. 2020; Aiello 2022) highlight how the more humanized images used to represent AI are, the more they carry gender stereotypes. For example, AI is often depicted as a female when portrayed as a nurse. Consider Image 1, which comes from one of the large image databases. Enter "nurse" as the search term and "AI-generated content" as the filter. The description of the image is particularly interesting: «A modern, AI-powered nurse robot is wearing a *white* lab coat and headset. She is sitting at a desk, looking at a computer screen with a patient's medical records. Her face is *calm* and *reassuring* as she listens intently to the patient's concerns» (italics mine). When we look at this picture, we immediately notice a few things: the nurse has a clear gender connotation – she is depicted with clearly visible breasts – she is highly anthropomorphized (as we saw earlier) and her human appearance is only given away by a machine-like neck, a dense tangle of dark-coloured wires, which contrast with the whiteness[3] of her uniform. This whiteness is also associated with the supposed calm and reassurance she is meant to instill in the viewer (or the hypothetical patient in front of her). Shutterstock's AI, which is created based on the images it has access to, acts as a kind of algorithmic food. If a culture has a long history of relegating women to subordinate roles, the developers of AI systems will likely ascribe a female gender to those roles.

[3] About the whiteness of AI see Cave & Dihal (2020).

Professional artists and communicators are well aware of the power of images (Freedberg 1989). Images can inform, inspire, attract attention, or simply entertain. Inaccurate or misleading images wield the same power, but their effects can potentially be harmful: they can misinform, limit imagination, or divert attention from divisive issues. In other words, images are never neutral, and the aesthetics they delineate are equally consistent with an ethics rooted in the regime of the senses.

One of the most controversial outcomes of using "bad" images of AI (where "bad" means stereotypical, numbing, replicating the same graphic patterns) is that they flatten the imaginary (if we can put it that way) by reducing it to existing imagery. Consider the history of advertising: when we look at an advertising campaign from many decades ago, we immediately feel a sense of detachment and incredulity because the image no longer translates an imaginary coherent with the current social reality. How can we imagine better images of AI[4] in this sense? Which philosophical categories can we use to deconstruct the reference imaginary of AI, while utilizing the imaginative, subversive, instituent, and non-instituted potential of images of AI?

## 2. What kind of image? Two useful categories of images of AI

In my opinion, it may be useful to categorize images representing artificial intelligence into two types:

1. Operational Image (i)
2. Invisible Image (ii)

First, it is essential to investigate the agentive nature of AI images – images as something that *does something* to us – and to relate these images to critical literature, particularly in the field of art history, which reflects on the concept of «operational image» (Freedberg 1989; Gell 1998; Mitchell 2005; Bredekamp 2010). It is then necessary to highlight the complex and, in some ways, unrepresentable or invisible nature of images of AI. They involve a series of operations that are not necessarily visible to the human eye, creating a machine-to-machine visual culture (Paglen 2016) and triggering a semantic shift whereby images are no longer sensory entities, but increasingly produced by machines for other machines. This fuels the process of machine learning training, which that should be analyzed from the perspective of dataset archaeology, given that no image is ever politically neutral and that images generated by GANs are not free of bias.

To begin with, I would argue that the question "What is an image?" requires an anthropological approach, since, as we shall see, the answer is primarily determined at the cultural level (Belting 2011). Secondly, I would like to recall Stiegler's words: «[t]here have never existed physical images (images-object) without the participation of mental images, since an image by definition is one that is seen (is in fact one when it is seen)» (Derrida & Stiegler 2002, 145) – by someone or something.

In the opening of *The Power of Images*, Freedberg writes:

[4] *Better Images of AI* is also a non-profit organization dedicated to researching, creating, curating, and providing better images of artificial intelligence. It commissions artists to produce free stock images that avoid perpetuating harmful myths about AI. Learn more at <https://betterimagesofai.org/about>.

> People are sexually aroused by pictures and sculptures; they break pictures and sculptures; they mutilate them, kiss them, cry before them, and go on journeys to them; they are calmed by them, stirred by them, and incited to revolt. They give thanks by means of them, expect to be elevated by them, and are moved to the highest levels of empathy and fear. They have always responded in these ways; they still do. They do so in societies we call primitive and in modern societies; in East and West, in Africa, America, Asia and Europe. (Freedberg 1989, 1).

This power that Freedberg ascribes to works of art should be extended to all images: the ability to generate emotions (repulsion or admiration, nostalgia, or even empathy) does not depend, we might say, on the medium by which it is embodied. It does not exclusively belong to painting or sculpture, but can also arise when looking at a photo in our iPhone album. In this sense, we can affirm that the image has its own agency, but image theory today actually engages with at least two significantly different interpretations of the concept of agency.

### 2.1 A double concept of agency

When we talk about image agency, I think we need to distinguish between at least two types. The first interpretation is primarily based on the theory of art formulated by British anthropologist Alfred Gell in his work *Art and Agency*. Broadly speaking, it falls within a theory of agency that emphasizes the attribution of subjective intentionality to images and objects, «which are seen as initiating causal sequences of a particular type» – which is actually Kant's definition of freedom: to initiate a new causal series – «that is, events caused by acts of mind or will or intention, rather than the mere concatenation of physical events. An agent is one who "causes events to happen" in their vicinity» (Gell 1998, 16).

This first interpretation of agency is complemented by another – which we can call Agency 2 – that does not require the postulation of subjectivity. Instead, it is transindividual and more concerned with the norms and social structures that Agency 1 overlooks – norms that it endures, shifts, or negotiates. The Agency 2 perspective reminds us that images and the ordinary relationships we have with them are the product of historical power relations that remain operative. We will try to navigate between these two interpretations of agency, while exploring the intermediate positions that lay the foundation to potentially articulate.

For example, the "racial" construction of our visual field required the creation and implementation of the visual norms and categories that underpinned the ideology of colonialism. Another good example of AI images' operational dimension is their use in medical imaging: AI offers the crucial ability to analyze vast volumes of medical images, enabling the identification of disease features that are unrecognizable to the human eye. In this sense, it holds the same epistemic value as the great optical technologies – from telescopes to microscopes – that made it possible to improve and expand the spectrum of vision. The same could be said of aerial surveillance imagery (see Dorian & Pousin 2023), which has always been linked to the development of military technologies and the new kind of surveillance gaze that drones deploy. In recent years, the application of these technologies for civilian use – such as satellite imagery – has redrawn the boundaries between private and public, personal and collective space. The widespread use of surveillance cameras that monitor and record ever-larger areas of public and private space, along with land mapping systems such as Google Street View, produces an immeasurable mass of images that no viewer will most likely ever see. This also leads to the colonization of previously invisible visual territories and gives rise to a huge field of research called surveillance studies (see Lyon 2007).

In Gell's view, art objects' importance lies not in beauty, but rather in fascination: he argues that what we call art objects, as well as many other things that we would not nec-

essarily think of classifying under this term, possess a force of fascination. In other words, art should be anchored in a theory of knowledge or, better yet, the aim should be a philosophy of the image – or even an assimilation of aesthetics and ontology, a sensory ontology, so to speak. Gell states: «I place all the emphasis on agency, intention, causation, result, and transformation. I view art as a system of action, intended to change the world rather than encode symbolic propositions about it. The "action"-centred approach to art is inherently more anthropological than the alternative semiotic approach because it is preoccupied with the practical mediatory role of art objects in the social process, rather than with the interpretation of objects "as if" they were texts» (Gell 1998, 6).

Speaking of an *image opératoire* also means recognizing that the question of the image is far from limited to representation. According to the seminal studies by Jean-Pierre Vernant (Vernant 1991), the reduction of the image to the semantic field of representation only took place in Greece in the 4th century BCE: Plato placed every image under the lens of *mimesis* and defined mimetic activity as *eidolon demiourgia*, or the «making of simulacra». However, *eidolon* was not always synonymous with mimetic artifice. Indeed, in the archaic period – between the 8$^{th}$ and 5$^{th}$ centuries BCE – an idol was not understood as an image confined to a relationship of resemblance to a model. Rather, it was associated with a certain ability to make visible, to portray, and to bring forth a supernatural power, even to realize it in concrete form. The image was hidden and governed by a specific dialectic of concealment and revelation.

It is no coincidence that when Vernant thought of the *eidolon*, he had in mind the ghostly images that populate Hades in Homer's masterpieces: the specter of Patroclus in Achilles' dream is not a copy of the flesh-and-blood man, nor is it a mere semblance, but an apparition, a double in the sense of a substitute, a marker, one who stands in for someone else – and in this sense it functions as an *evocative device*. The idol makes present and at the same time measures the abyss that separates the world of humans from that of the gods, the realm of the dead from that of the living: the archaic *eidolon*, characterized by this dialectic of presence and absence marked by magical element, manages to manifest something that is, at the same time, stamped with the seal of absence. The image gives us something to see, and at the same time it affects us.

So far, I have argued that images do not only have a purely cosmetic value but also an ethical-political dimension, albeit implicit: they convey expectations, prejudices, and desires of a particular society. Recognizing the agentic character of images is the first step toward broadening the discourse on images and giving it an ethical dimension. In doing so, I have identified two nuances of the meaning of agency: the first attributes a form of subjective intentionality to images, while the second does not concern images alone, but more broadly the imaginaries associated with the narratives that a given society makes of itself. The latter are never neutral, but mediate power relations that remain at least latent, if not explicit. Finally, recognizing the acting character of AI imaginaries means that the value of images is not limited to the representation vector.

Having clarifies this theoretical framework, I would like to move on to the other category at play in AI images, which is that of the *invisible image* (which, in a certain sense, also allows us to pass from Agency 1 to Agency 2) and to consider the more ethical and ethico-political implications of the unrepresentability of AI images.

## 3. Invisibility in AI images

At the end of the previous section, we established the importance of looking at AI images through the interpretative lens of the operative image, and not taking for granted the neutrality of the images that flood contemporary visual communication, both scien-

tific and popular. It is now a matter of turning to the other philosophical category at play: that of invisibility.

Even the most commonplace myths suggest that the idea of invisibility remains provocative and fascinating across centuries and cultures (the cloak of invisibility, the magic ring, etc.), fueling fairy tales and popular legends. We can say that each iteration of the relationship between the visible and the invisible raises questions not only of aesthetics calibre, but also ethical and political: how is it that some things remain invisible while others manifest themselves? What does the fundamental relationship between the visible and the invisible mean, especially – yet not exclusively – for visual culture? Our way of seeing and our acts of vision are necessarily indicative of our ethics, of how we envisage human relationships. Consider, for example, the complexity of the dynamics of being seen and not being seen – the so-called «invisibles», *those that have no part,* as Rancière (2004) described them – and how this dialectic continues to animate the discourse on political identities.

Moreover, the advent of mechanical vision introduces not only new visual regimes, but also new paradigms of visibility, in which – to use Paglen's words – «images are overwhelmingly invisible, even as they're ubiquitous and sculpting physical reality in ever more dramatic ways» (Paglen 2014, 3). Consider the images contained in databases such as ImageNet or Laion-5B, which currently hold almost 6 billion images and captions. These databases are used (among other things) to train Imagen or Stable Diffusion, which function similarly to the perhaps better-known Dall-E 2. They provide the staple diet for the production of novel AI-generated images. And yet, as Melanie Mitchell notes, «seeing – both seeing and looking – turns out to be one of the hardest of the "easy" things» (Mitchell 2019, 59). Unless they are blind from birth, humans have the ability to observe – that is, to look at and recognize an image in a fraction of a second. This is a process of which we are barely aware, almost an automatism. But that is not all: in visual processing, we are also able to link an image to a reference context by following a kind of visual pragmatics. We immediately know what to keep in an image and what to leave out. So, we could ask: What is the semio-pragmatic dimension of an image database? Who decides what to keep and what to leave out and how? The basic idea is that every «field of images is organized» (Rose 2010, 61), embedded in complex, specific and diverse «social relations, practices and institutions» (Rose 2010, 62).

Authors Kate Crawford and Trevor Paglen aim precisely to uncover «the construction of these training sets and the "underlying structures" that often betray "unquestioned assumptions"» (Crawford & Paglen 2021, 1006). The result is that «every layer of the architecture of a given training set is infused with politics» (Crawford & Paglen 2021, 1007). Data sets are never a collection of raw material to feed algorithms, instead they underlie policy operations. Consequently, to argue for the plausibility and necessity of an ethics of images, and AI images in particular, means to unravel, at least partially, these supposedly neutral political mechanisms: to unlearn how humans see, to learn to see the invisible. These invisible images invite us to rethink our concepts of visibility and invisibility. The latter is, from the outset, a matter of relationality: the invisible «is the secret counterpart of the visible», to use Merleau-Ponty's words (Merleau-Ponty 1969, 215).

As Øyvind Vågnes notes, «the subject of invisibility might be approached through the prisms of aesthetics, representation, technology and politics» (Vågnes 2019, 3). How should we invoke the concept of the invisible when referring to images of AI, though? First of all, these are images that struggle to activate the vector of representation. In a sense, they are empty containers of meaning. We might say that they owe their banality to the fact that meaning transcends the iconic signifier. We struggle to process, circulate, and consume good AI images because perhaps it does not make much sense to talk about AI in general terms. As a result, AI seems like an object that is too fragmented, hidden,

and invisible to come up with an adequate image of it. There was a moment in the history of philosophy when a similar difficulty arose: when the new fields of psychology and phrenology tried, then too by resorting to iconic representations, to make sense of that obscure thing called the brain. Just as a photograph of a brain or a measurement of a skull does not do justice to the complexity of the human mind. Similarly, using the image of an algorithm to represent AI is equally reductive. Furthermore, the issue of invisibility leads us to think about transparency, a topic of great interest in the branch of philosophy of technology that reflects on intelligent machines.

Technology is the (invisible) condition of visibility.[5] «Phenomenological transparency», as Michael Wheeler has called it (Wheeler 2022, 16), is the impossibility to focus one's attention on both the action and its (technological) means, and, for systems, to simultaneously see and be seen. Yet, this is not the first time in the history of science and technology that an object divides the world between those who understand how it works and those who do not.

People use AI without understanding it (although we are not dealing with strict categories here, but rather varying degrees of understanding – and acceptance). Most of us agree to taking planes, paying taxes, getting vaccinated, and undergoing surgery: we do not need to understand how these things work in order to use them. So why are there different requirements for algorithms today?

Of course, it is said that there is a moral risk, which is why it is necessary to clarify not the technological mechanics but the organizational mechanics of these digital tools. An explanatory effort must be made so that an epistemology of trust can be built. «[I]t is often impossible for an explanation to be completely faithful unless it is the complete description of the model itself» (Ribeiro et al. 2016, 3). Software, then, does not truly aim for explicability, but for interpretability. Transparency is not necessarily correlated with trust, quite the opposite: disclosure of the mechanism is not enough to restore any trust that may have been compromised. If, by definition, trust precisely concerns relationships in which one does not have complete knowledge, but in which trust is granted despite factual ignorance, it is quite natural that one cannot return to a state of full trust by means of a catalogue of empirical facts. Hence, certain affective dynamics – our experience of sensitivity and affect, that is, the emotional tone with which we inhabit the world – should be reintegrated into the discourse on the digital. In other words, it is a matter of engaging with the realm of the aesthetics of communication, where aesthetics does not refer to – or at least not primarily – art or the production of artifacts, but something inherent to perception: in this sense, ICTs can open or conceal the regime of the visible and provide us with access to the world – the world of the environment, the world of others and the world of the self. And that is why it makes sense to turn to the imaginary as a place where meaning is created.

## 4. AI imaginaries

So how can we create a symbolic space, and new imaginaries related to AI? First of all, it must be said that symbolic space is not the illusory complement of physical space, but rather the system of collective meanings that, while implicit, is able to guide the expectations, desires, and actual behaviors of subjects inhabiting the physical space. A symbolic

[5] As stated by Noë: «we never wonder what a door handle is when we encounter one. The very fact that the question arises means that the usefulness of the handle has been compromised. If we notice a handle, it is potentially the result of poor design» (Noë 2015, 101).

space exists where the creative power of a *shared social imaginary* resides. Through this concept, which is also widely used in social sciences, we refer to a shared *habitus* on which individuals build their expectations, their specific ways of being in the world and of interacting with each other. Imaginaries therefore operate primarily at image level, on the basis of an associative grammar made up of resemblances and references, rather than formalized concepts and explicitly structured intellectual schemes.

As stated by Taylor:

> By social imaginary, I mean something much broader and deeper than the intellectual schemes people may entertain when they think about social reality in a disengaged mode. I am thinking, rather, of the ways people imagine their social existence, how they fit together with others, how things go on between them and their fellows, the expectations that are normally met, and the deeper normative notions and images that underlie these expectations. (Taylor 2004, 23).

Sheila Jasanoff has also provided a good definition of sociotechnical imaginaries: «We redefine sociotechnical imaginaries as collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology» (Jasanoff et al. 2015, 4).

The concept of the imaginary thus acquires two functions: the first is to explain the existence of a society (the imaginary is what *holds everything together* in a society), and the second is to explain the existence of historical change, i.e. the fact that there is historical change and not endless repetition. For individuals to live, the existence of reality is not enough: there must be non-reality, the ability to distance oneself from what is. In order to do so, one must be able to ignore reality, or rather, to let possibility exist.

Long before Jasanoff, Kim, Taylor and Anderson, Castoriadis – who notably influenced their theories – enables us to think of socio-technical imaginaries as a hinge between macro and micro levels. This allows us to refer to different levels of discourse simultaneously and through the figure of the instituting/instituted imaginary. For Castoriadis, the social imaginary is a primary structuring (Castoriadis 1975), in which distinctions are articulated about what matters to each individual and to communities, through a series of investments of intellectual, practical and affective nature. Castoriadis' relevance to this discussion lies in the fact that he sheds light on the generative matrix of imaginaries. «We encounter in history the imaginary as a continued origin, always current foundation, a central component where what holds society together and what produces historical change are generated» (Castoriadis 2007, 145 – my translation).

A concrete example of an instituted and instituent imaginary is law: in Castoriadis' words, «one finds a dawning recognition of the fact that the source of the law is society itself, that we make our own laws – whence results the opening up of the possibility of challenging and putting into question the existing institutions of society, which now is no longer sacred, or in any case not sacred in the same way as before» (Castoriadis 1993, 105). Law is the only discursive practice that produces the world it designates. It is everywhere and nowhere, and this omnipresence is given by the very relations that it institutes.

A human being, according to Castoriadis, is essentially an imaginative being, who produces, consumes, and loves images. In the same vein, human being is not seen at all a rational being, as tradition would have it, but rather as a being who imagines and fantasizes, capable of delirium and irrationality, and in this sense opposed to animals. As we can read in *L'imaginaire comme tel*: «An animal that lacks food looks for its food; if it does not find it, it weakens and eventually dies. A man who lacks food also weakens; but before dying,

he looks for something else, makes a stick, invents a trap, wages a war, or tells himself a story» (Castoriadis 2007, 150 – my translation).

According to Castoriadis, institutions exist only in the symbolic sphere, providing a form of stabilization, and it is in this sense that the instituted form of the state emerges. At the same time, however, the symbolic dimension is constantly in flux, which is why Castoriadis introduces the idea of the instituting alongside the concept of institution. The instituting refers to the perpetuation of the other in the radical imaginary and describes the moment when instituting society bursts into instituted society and creates itself as a new (instituted) society. Castoriadis seeks to demonstrate the inventive nature of the social being: from this perspective, radical imagination, on both the individual and collective level, appears as a constitutive, or rather, institutive force of reality. It is significant that, when taking stock of the May 1968 events, Castoriadis writes: «all this is evidently inconceivable without a reversal not only of the existing institutions, but even of what we mean by institution» (Castoriadis et al. 1968, 140 – my translation).

As Merleau-Ponty said when he outlined his philosophy of institution (during his lectures at the Collège de France, 1954-55), its aim is to understand the dynamics through which all of human existence unfolds. This does not only include collective history, but also the affections, dreams, and desires that incarnate themselves in an «instituted life» (Merleau-Ponty 2010, 30). Institutions which do not want to become rigid, must always operate in the following manner: they must allow the flow to express itself without dissipating, settling instead into partial representations. These, far from being separate, are seen as a collective act, manifested in a series, but rather schemes, figures, and symbols.

The theoretical reference to Castoriadis is particularly fruitful, in my opinion, because it helps us to distinguish imagination from the imaginary: the imaginary is not a subsequent addition to the natural world and its pre-formed structures. In this perspective, it is seen neither as a subjective faculty nor as a cultural superstructure to be dispensed with. In Castoriadis' view, imagination is original and instituting. Therefore, it brings things into existence. However, doing so as an independent form means giving value to objects in a network of meanings suitable for use. All of this reminds us of the importance of narrative as a constitutive part of any socio-technical imaginary. Keeping in mind the creative generative matrix of thought (understood as "creative rupture") can help us to view the dominant contemporary imaginaries of AI (often disconnected from this technology's true state of the art) with skepticism. The way we envision AI is perhaps an aesthetic issue, but it also raises questions about the ethical and socio-political challenges we face.

The imaginary that we feed, shape, and nourish with images of AI is not something given once and for all, nor something that overwhelms us. If the image, as Foucault reminds us, is the oldest sign, the archaic sign, the *first sign* (Foucault 2008), it also represents *primitiveness* in the production of meaning and emotion. This is where we need to look in order to build a new epistemology of trust in AI.

## 5. Conclusions and possible directions for future research

The conclusions reached in the course of this discussion can be summarized as follows:

a) Ethical-Political stakes of AI images: AI imagery is not a merely aesthetic issue; it also embodies ontological and ethical-political dimensions that demand scrutiny.

b) Non-Representability of AI: Ontologically, AI-generated images expose the inherent difficulty – or impossibility – in fully representing the complexity of AI systems.

c) Analyzing AI images through the key categories of agency and visual regime (the dialectic between the visible and the invisible) can help to incorporate critical viewpoints into education and public discussion, so as to foster a more inclusive and thoughtful conversation about AI and its impact on society.

Some further research directions: human interaction is based on trust. Reframing AI imagery as an ethical-political challenge would foster this trust. I plan to explore on the practical implications of these theoretical findings in a future paper, by attempting to outline some guidelines for ethical AI imagery and actionable recommendations to help policymakers, educators, and AI developers build a less sclerotic and stereotypical imagery.

## Bibliography

Aiello, G. (2022). *Communication, Espace, Image*. Paris: Les Presses du Réel.

Belting, H. (2011). *An Anthropology of Images: Picture, Medium, Body*. Princeton, NJ: Princeton University Press.

Bredekamp, H. (2010). *Theorie des Bildakts*. Frankfurt a. M.: Suhrkamp.

Castoriadis, C. (1975). *The Imaginary Institution of Society*. Engl. Trans. K. Blarney. Cambridge, MA: MIT Press.

Castoriadis, C. et al. (1968). *Mai 68: La brèche. Premières réflexions sur les événements*. Paris: Fayard.

Castoriadis, C. (1993). The Greek and the Modern Political Imaginary. *Salmagundi*, 100, 102-129.

Castoriadis, C. (2007). *L'immaginaire comme tel*. Paris: Hermann.

Cave, S. et al. (2020). *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press.

Cave, S. & Dihal, K. (2020). The Whiteness of AI, *Philosophy & Technology,* 33, 685-703 (https://doi.org/10.1007/s13347-020-00415-6).

Cocchiaro, M. Z. et al. (2024). Who Is an AI Ethicist? An Empirical Study of Expertise, Skills, and Profiles to Build a Competency Framework. *SSRN* (https://dx.doi.org/10.2139/ssrn.4891907).

Crawford, K. & Paglen, T. (2021). Excavating AI: The Politics of Images in Machine Learning Training Sets. *AI & SOCIETY*, 36, 1105-1116 (https://doi.org/10.1007/s00146-021-01162-8).

Derrida, J. & Stiegler, B. (2002). *Echographies of Television: Filmed Interviews*. Engl. Trans. J. Bajorek. Cambridge: Polity Press.

Dorian, M. & Pousin, F. (eds). (2013). *Seeing from Above. The Aerial View in Visual Culture*. London-New York: I. B. Tauris.

Foster, H. (1988). *Vision and Visuality*. Seattle, WA: Bay Press.

Foucault, M. (2008). *This Is Not a Pipe*. Engl. Trans. J. Harkness. Berkeley, CA: University of California Press.

Freedberg, D. (1989). *The Power of Images: Studies in the History and Theory of Response*. Chicago: University of Chicago Press.

Gell, A. (1998). *Art and Agency: An Anthropological Theory*. Oxford: Oxford University Press.

Grønstad, A. & Vågnes, Ø. (eds). (2019). *Invisibility in Visual and Material Culture*. London: Palgrave Macmillan.

Jasanoff, S. & Kim, S-H. (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press.

Lyon D. (2007). *Surveillance Studies: An Overview.* Cambridge: Polity Press.

Merleau-Ponty, M. (1969). *The Visible and the Invisible*. Engl. Trans. A. Lingis. Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (2010). *Institution and Passivity: Course Notes from the Collège de France (1954-1955)*. Engl. Trans. L. Lawlor & H. Massey. Evanston, IL: Northwestern University Press.

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. New York: FSG (Farrar, Straus and Giroux).

Mitchell, W. J. T. (2005). *What Do Pictures Want? The Lives and Loves of Images*. Chicago: University of Chicago Press.

Noë, A. (2015). *Strange tools: Art and Human Nature*. New York: Hill and Wang.

Paglen, T. (2014). Operational Images. *e-flux Journal*, 59; November 2014 (https://www.e-flux.com/journal/59/61130/operational-images).

Rancière, J. (2004). *Disagreement: Politics and Philosophy*. Engl. Trans. J. Rose. Minneapolis, MN: University of Minnesota Press.

Rancière, J. (2013). *The Politics of Aesthetics: The Distribution of the Sensible*. Engl. Trans. G. Rockhill. London: Bloomsbury.

Ribeiro, M.T. et al. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *KDD '16 (Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining)*, 1135-1144 (https://doi.org/10.1145/2939672.2939778).

Romele, A. (2022). Images of Artificial Intelligence: A Blind Spot in AI Ethics. *Philosophy & Technology*, 35 (4), 1-19 (https://doi.org/10.1007/s13347-022-00498-3).

Romele, A. & Severo, M. (2023). Que veulent les images de l'IA? Une exploration de la communication scientifique visuelle de l'intelligence artificielle. Sociétés & Représentations, 55 (1), 179-201 (https://doi.org/10.3917/sr.055.0179).

Rose, G. (2010). *Visual Methodologies: An Introduction to Researching with Visual Methodologies*. Thousand Oaks, CA: Sage.

Taylor, C. (2004). *Modern Social Imaginaries*. Durham: Duke University Press.

Vernant, J.-P. (1991). *Mortals and Immortals: Collected Essays*. Ed. F. I. Zeitlin. Princeton, NJ: Princeton University Press.

Wheeler, M. (2022). Entre la transparence et l'intrusion des machines intelligentes. In: T. Reigeluth & S. Benlaksira (eds.), *Intelligence artificielle. Que faire de la transparence technique? (Revue de philosophie contemporaine, 2)* (13-29). Paris: Vrin.