palgrave communications
HUMANITIES | SOCIAL SCIENCES | BUSINESS

ARTICLE

https://doi.org/10.1057/s41599-020-0445-0 OPEN

Algorithmic unconscious: why psychoanalysis helps in understanding AI

Luca M. Possati[1]✉

ABSTRACT The central hypothesis of this paper is that the concepts and methods of psychoanalysis can be applied to the study of AI and human/AI interaction. The paper connects three research fields: machine behavior approach, psychoanalysis and anthropology of science. In the "Machine behavior: research perspectives" section, I argue that the behavior of AI systems cannot be studied only in a logical-mathematical or engineering perspective. We need to study AI systems not merely as engineering artifacts, but as a class of social actors with particular behavioral patterns and ecology. Hence, AI behavior cannot be fully understood without human and social sciences. In the "Why an unconscious for AI? What this paper is about" section, I give some clarifications about the aims of the paper. In the "Unconscious and technology. Lacan and Latour" section, I introduce the central thesis. I propose a re-interpretation of Lacan's psychoanalysis through Latour's anthropology of sciences. The aim of this re-interpretation is to show that the concept of unconscious is not so far from technique and technology. In the "The difficulty of being an AI" section, I argue that AI is a new stage in the human identification process, namely, a new development of the unconscious identification. After the imaginary and symbolic registers, AI is the third register of identification. Therefore, AI extends the movement that is at work in the Lacanian interpretation of the mirror stage and Oedipus complex and which Latour's reading helps us to clarify. From this point of view, I describe an AI system as a set of three contrasting forces: the human desire for identification, logic and machinery. In the "Miscomputation and information" section, I show how this interpretative model improves our understanding of AI.

[1] University of Porto, Porto, Portugal. ✉email: lupossati@gmail.com

Introduction

The central hypothesis of this paper is that the concepts and methods of psychoanalysis can be applied to the study of AI (artificial intelligence) and human/AI interaction. The paper connects three research fields: machine behavior approach, psychoanalysis and anthropology of science.

I intend to pose three questions:

- Why do humans need intelligent machines?
- What is AI's unconscious?
- How does this notion enrich our understanding of AI?

In the "Machine behavior: research perspectives" section, I argue that the behavior of big AI systems cannot be studied only from a logical-mathematical or engineering perspective. We need to study these systems that today regulate most of the aspects of our life as social agents in constant interaction with humans. Applying the concepts of psychoanalysis to AI means expanding the so-called "machine behavior approach".

In the "Why an unconscious for AI? What this paper is about" section, I give some clarifications on what this paper is about. In the "Unconscious and technology. Lacan and Latour" section, I propose a re-interpretation of Lacan's psychoanalysis and Latour's actor-network theory. I re-interpret Lacan's thesis on the mirror stage and oedipal complex through Latour's methodology. The aim of this re-interpretation is to show that the concept of unconscious is not so far from technology. The unconscious is the effect of a technical mediation. From this point of view, I apply the notion of unconscious to AI. Latour is the mediation between AI and psychoanalysis.

In the "The difficulty of being an AI" section, I apply this re-interpretation of Lacanian psychoanalysis to AI. I argue that AI is a new stage in the process of human identification, that is, a new development of the unconsciousthat I call the "algorithmic unconscious". In AI, the machine responds to a human desire for identification. Hence, AI extends the movement that is at work in the Lacanian interpretation of the mirror stage and Oedipus complex and which Latour's reading helps us to clarify. In particular, I describe an AI system as a set of three contrasting forces: the human desire for identification, logic and machinery, that is, the embodiment of logic.

In the "Miscomputation and information" section, I show that this interpretative model improves our understanding of AI in two ways. It gives us a new interpretation of two essential notions: miscomputation and information.

The result is a coherent and original model of interpretation of AI, which is able to explain the originality of AI compared to any other form of technology.

Machine behavior: research perspectives

[…] *we describe the emergence of an interdisciplinary field of scientific study*. This field is concerned with the scientific study of intelligent machines, not as engineering artifacts, but as a class of actors with particular behavioral patterns and ecology. This field overlaps with, but is distinct from, computer science and robotics. It treats machine behavior empirically. This is akin to how ethology and behavioral ecology study animal behavior by integrating physiology and biochemistry—intrinsic properties—with the study of ecology and evolution—properties shaped by the environment. Animal and human behaviors cannot be fully understood without the study of the contexts in which behaviors occur. Machine behavior similarly cannot be fully understood without the integrated study of algorithms and the social environments in which algorithms operate. […] Commentators and scholars from diverse fields—including, but not limited to, cognitive systems engineering, human computer interaction, human factors, science, technology and society, and safety engineering—are raising the alarm about the broad, unintended consequences of AI agents that can exhibit behaviors and produce downstream societal effects—both positive and negative—that are unanticipated by their creators (Rahwan et al. 2019, p. 477; emphasis added).

The behavior of AI systems is often studied in a strict technical engineering and instrumental manner. Many scholars are interested only in what the machine does and what results it achieves. However, another, broader and richer approach is possible, which takes into account not only the purposes for which the machines are created and their performance, but also their "life", that is, their behavior as agents that interact with the surrounding environment (human and non-human). This approach is called "machine behavior", i.e., the study of AI behavior, "especially the behavior of black box algorithms in real-world settings" (Rahwan et al. 2019, p. 477), through the conceptual schemes and methods of social sciences that are used to analyze the behavior of humans, animals and biological agents.

As various scholars claim, algorithms can be perfectly capable of adapting to new situations, even creating new forms of behavior. Cully et al. (2015) demonstrate that it is possible to construct an intelligent trial-and-error algorithm that "allows robots to adapt to damage in less than two minutes in large search spaces without requiring self-diagnosis or pre-specified contingency plans" (503). When the robot is damaged, it uses the prior knowledge "to guide a trial-and-error learning algorithm that conducts intelligent experiments to rapidly discover a behavior that compensates for the damage" (503), and this makes the robot able to adapt itself to many different possible situations, just like animals. "This new algorithm will enable more robust, effective, autonomous robots, and may shed light on the principles that animals use to adapt to injury" (503). Lipson (2019) has obtained the same results with another robotic experiment about autonomy and robots. AI systems are capable of creating a completely new form of behavior by adapting themselves to new contexts. This is an artificial creativity.

The machine behavior approach intends to examine the AI adaptability not from a strictly mathematical point of view, but from the interaction between these machines and the environment. "Rather than using metrics in the service of optimization against benchmarks, scholars of machine behavior *are interested in a broader set of indicators, much as social scientists explore a wide range of human behaviors in the realm of social, political or economic interactions*" (Rahwan et al. 2019, p. 479; emphasis added).

Studying the adaptation of AI to the environment also means studying the lack of this adaptation, i.e., the pathological behaviors that AI can develop. According to O'Neil (2016), an uncritical use of algorithms and AI systems can result in very dangerous consequences for society. Studying AI systems that process big data only from a mathematical and statistical point of view significantly undermines our understanding of the complexity of their functioning, hindering us from grasping the real issues that they imply. These AI systems can produce injustices, inequalities, and misunderstandings, feed prejudices and forms of discrimination, aggravate critical situations, or even create new ones. Furthermore, these systems are "black boxes", i.e., they are opaque. There are two explanations for this: (a) for legal and political reasons, their functioning is often not made accessible by the companies that create and use them; (b) the computation speed makes it impossible to understand not only the overall dynamics of the calculation but also the decisions that the systems make. Engineers struggle to explain why a certain algorithm has taken that action or how it will behave in another situation, e.g.,

in contact with other kinds of data (Voosen, 2017, p. 22). These algorithms are complex and ubiquitous, and it is very difficult to predict what they will do in the most diverse contexts.

O'Neil points out that these systems are based on mathematical models. Mathematical models are not just neutral set of formulas. These models "are opinions embedded in math" (O'Neil, 2016, p. 50), i.e., they are based on the evaluations and prejudices of those who design them, and therefore they reflect certain values and priorities—a world view. Now, mathematical models are the tool used by AI systems to analyze data, extract patterns, make forecasts and then make decisions. One of the risks—what O'Neil calls "feedback loop"—is that the systems based on mathematical models reproduce existing situations.

The typical example is that of college loans (O'Neil, 2016, p. 81). The system identifies the poorest segment of the population and bombard them with ads on university loans spreading the message—at least theoretically right—that better education can lead to a better employment and a better income. As a result, many poor people decide to take out a debt. Unfortunately, due to a period of economic recession, people who have contracted a debt and received training cannot find a job or lose what they already have. This is a situation that the model had not foreseen. People cannot repay the loan. The final result is that the poor become poorer—the starting situation is amplified. O'Neil demonstrates this loop through several examples. Some AI systems confirm also the prejudice that the poor are those who commit more crimes (O'Neil, 2016, pp. 90–94).

Therefore, according to O'Neil, the *Weapons of Math Destruction* have four main features:

- They control and change our behavior by directing our choices and our tendencies. They do this on the basis of decisions that have very large-scale effects.
- They can amplify wrong attitudes, that are contrary to the law or to the elementary rules of social coexistence, for example racial prejudices.
- They can cause irreparable damage, such as losing job, home and savings, or destroying social ties, for reasons that have no basis in reality, or that are based on partial visions of reality.
- They can be manipulated.

If we want to minimize the collateral effects of AI behavior, we must then look beyond their logical-mathematical structure and statistics. We must choose a model of analysis that is not based on the rigid distinction between humans and machines, but that focuses on the mutual interaction of them.

Studying machine behavior is not easy at all. AI behavior can be analyzed from at least six different perspectives: (a) the behavior of a single AI system, (b) the behavior of several AI systems that interact (without considering humans), (c) the interaction between AI systems and humans. Today most interactions on planet Earth is of the type b. Moreover, according to Rahwan et al. (2019), when we talk of interactions between AI systems and humans, we mean three different things: c.1) how AI systems influence human behavior, c.2) how humans influence AI systems behavior, c.3) how humans and AI systems are connected within complex hybrid systems, and hence can collaborate, compete or coordinate. All these layers are intertwined and influence each other, as the Fig. 1 shows.

The final goal of this paper is to understand if and how the methods and concepts of psychoanalysis can be applied to the interaction between machines and their environment on all these levels. Therefore, I want to widen the machine behavior approach. But, first of all, two questions need to be tackled: In what sense do I use the terms "conscious" and "unconscious"? How can I apply them to AI?

Why an unconscious for AI? What this paper is about

A clarification on the objectives of this paper is needed. In which sense do we talk of "algorithmic unconscious"? Can an algorithm be conscious or unconscious? Applying psychoanalysis to the study of AI can entail the question about the "consciousness of machines". The philosophical debate is very broad and hard to summarize in a single paper (Churchland and Smith, 1990; Larrey, 2019). Nevertheless, this is not the scope of this paper. I choose another type of problems and another perspective.

From my point of view, talking of "algorithmic unconscious" does not mean attributing consciousness, feelings or moods to machines. This paper will not treat the question: Can a sufficiently advanced AI become conscious or unconscious in the human sense? Machines and humans are completely different beings. When I talk about "artificial intelligence", I talk about the interaction between machines and human beings. A machine can be called "intelligent" only by the interaction with humans, i.e., bythe ability that the machine has to collaborate with humans in a useful way. AI is, therefore, the name of an intermediate field between machines and humans. It is a multifaceted concept, which takes on different meanings in relation to different contexts.

Two essential clarification must be made. The first: Freud's unconscious is not the preconscious, or the perceptive unconscious. It is not simply what does not come to consciousness. Freud's unconscious is the repressed, that is what *must not* be conscious, even if it remains connected to consciousness. Freud's unconscious is the result of a *resistance*, and this resistance causes effects on consciousness. Consciousness can understand these effects only through a specific technique, the psychoanalysis. In other words, for Freud the distinction between conscious and unconscious is the result of a repression, of a resistance, not the opposite. In his analysis of slips or dreams, Freud emphasizes the importance of unconscious censorship, of repression (Ellenberger, 1970). Exactly like political censorship, psychic censorship allows the manifestation of unconscious desires to reach the consciousness but only in a disguised form making them unrecognizable. Memories, desires, emotions, names and images are then assembled by unconscious into apparently absurd constructs. We need a specific technique in order to understand them. According to Freud, unconscious thoughts evade psychic censorship through specific mechanisms: condensation (*Verdichtung*) and displacement (*Verschiebung*) (Lacan, 1953–54; Ricoeur, 1965, part 2). Current research in cognitive science confirm the basic Freudian insights: a) certain cognitive processes are not only hidden but cannot be brought into consciousness; and b) the self is fundamentally non-unified, "and because of its fragmentation, self-awareness represents only a small segment of cognitive processes" (Tauber, 2013, p. 233).

In this paper, when I talk of "algorithmic unconscious", I am referring to the Freudian conception of unconscious. Lacan introduces another element: the language. Language is the great Other, which grounds the subject and roots it in a social context. And yet, precisely by doing this, language splits the subject, separating it from the most authentic part of itself. Language is *Spaltung* (splitting, repression). "It is in the splitting of the subject that the unconscious is articulated" (Lacan, 1966a, p. 24; translation is mine). The subject as social individual is a product of language, i.e., of repression. The child is *infans* because it is not yet defined by language; it is not subject to the symbolic repression. The child becomes a subject and "enter" the language only thanks to the Oedipus complex and the Name of the Father (Rifflet-Lemaire, 1970, pp. 130–131). For Lacan, the subject is always alienated.

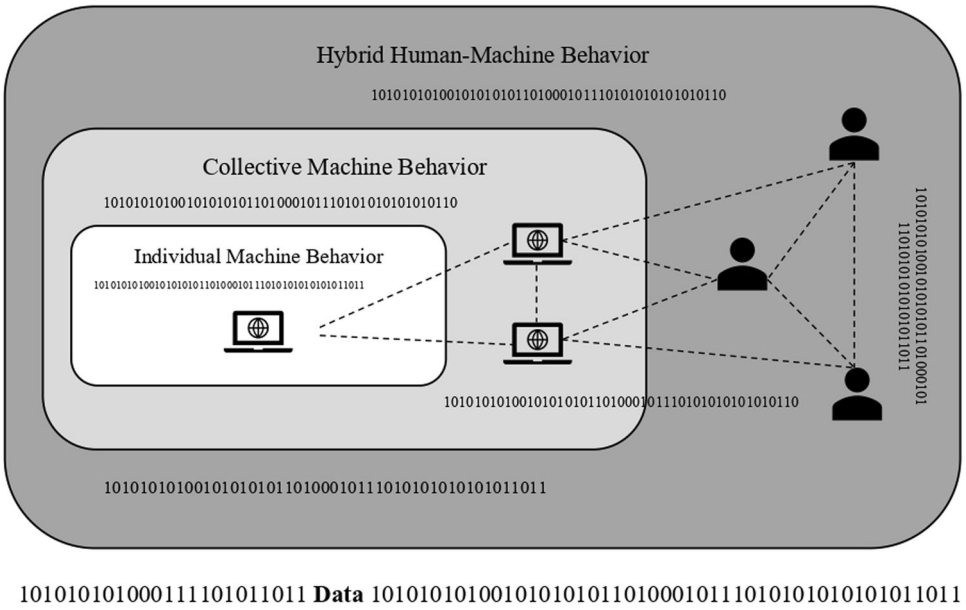

Fig. 1 The three main contexts of machine behavior that can be analyzed by using the methods of social sciences. The sphere of the behavior of the individual AI systems is always connected with that of the interaction between different AI systems and finally with that of the interaction between humans and AI.

The second clarification: In this paper I analyze the interaction between human unconscious (in the Freudian sense) and AI. I focus on the interaction, not on the isolated poles. I try to understand how technology plays an essential role in the formation of human unconscious, how humans project their own unconscious on machines and how these latter transform human unconscious. My goal is to analyze the relationship between the Freudian unconscious—in Lacan's interpretation—and AI as a network of actors.

Unconscious and technology. Lacan and Latour

In this section I propose a re-interpretation of some concepts of Lacanian psychoanalysis through Latour's actor-network theory. My goal is to show that this re-interpretation gives us new tools to understand AI and the relationship between humans and AI.

Before proceeding with the re-interpretation of Lacanian concepts, I want to answer three questions:

- Why do I choose as model Latour's anthropology of sciences?
- How do I interpret Latour?
- What is my critical approach to Latour?

My reading of Latour is above all philosophical.[1] What is the central point of Latour's philosophy? I want to summarize it using two formulas: symmetrical ontology and realism of resistance. These are the two aspects that interest me most in Latour's work. The first aspect—which is emphasized by object-oriented ontology (Harman, 2019)—is expressed by the first proposition of *Irréductions*: "Nothing is, by itself, either reducible or irreducible to anything else" (Latour, 2011, p. 158). This means that all entities are on exactly the same ontological footing. However, the "entities" are not static substances, but centers of force. "There are only trials of strength, of weakness. Or more simply, there are only trials. This is my point of departure: a verb, 'to try'" (Latour, 2011, p. 158).[2] Reality is a set of forces, trials and resistance: "Whatever resists trials is real", and "The real is not one thing among others but rather gradients of resistance" (Latour, 2011, p. 158). The "form" is the stabilization of forces, a dynamic equilibrium. "A form is the front line of a trial of strength that deforms, trans-forms, in-forms or per-forms it. Of course, once a form is stable, it no longer appears to be a trial of strength" (1.1.6). Here I see a profound analogy between Latour's ontology and Simondon's theory of individuation.

I do not want to analyze all the philosophical implications of these propositions. I want only to emphasize that the "principle of irreductibility" leads Latour to an approach that equates humans and non-humans, giving priority to their dynamic interactions, the "translations" of one another. It is a "flat ontology" (Harman, 2019, p. 54), i.e., an ontology that treats all objects in the same way, without any previous taxonomy, classification or hierarchy. Humans and non-humans are all "actors" in the same way within a network of associations (Latour, 2005), where translation processes, that is, exchanges of properties and skills, take place constantly. Latour calls this kind of network *collectif*, or "nature-culture". This position avoids any kind of dualism: "Such has always been the strategy of the French sociologist and anthropologist: to show the matter of spirit and the spirit of the matter, the culture in nature and the nature in culture" (Romele, 2019).

There are also other reasons that lead me to choose Latour. Latour's anthropology and sociology avoid reducing scientific facts to simple social phenomena, as well as some forms of Kantian constructivism; it also calls into question the classical epistemological approach in interpreting scientific facts, and this by considering the science "in action" (Woolgar and Latour, 1979). The result is a form of relativistic materialism (in the best sense of these expressions) which constitutes—in my opinion—a very mature form of realism.

Nevertheless, my reading of Latour is also critical. To my mind, Latour does not adequately consider AI and digital technology. I share the opinion according to which Latour "is for sure not a digital sociologist" and his views about the digital are contradictory with his main approach (Romele, 2019). Latour does not fully develop the principle of "irreductibility", that is the root of his symmetrical ontology. If he had done so, he should have tackled AI from the beginning. This is a paradoxical situation: on the one hand, the Latourian concept of *collectif* is very useful in explaining machine behavior and human-non-human contexts; on the other, in Latour a complete and rigorous theory of AI is absent. AI is the field in which the difference between humans and non-humans, or rather, the difference between living and

non-living beings, is radically challenged. In other words, the category of non-humans cannot be treated monolithically, as if all non-humans and the artifacts were the same. There is not only one technology, but many different technologies.

Does Latour really get rid of the modern compromise? In Latour humans always have the leading role in the constitution of the world: "the Pasteur network" creates the microbe, "the Joliot network" creates the nuclear chain reaction, the geologists in the Amazon create the forest, etc. AI overturns this scheme: there are new forms of human-non human hybrids in which it is the non-human part that has the control of the situation and "creates the fact". Hence, AI would be the real big test for the actor-network theory. In a nutshell, my criticism is that, in Latour, there is no a real AI theory. Nevertheless, precisely because of its philosophical assumptions, his method can be applied to AI with advantageous results. In other words, Latour lacks cybernetics, i.e., a theory of intelligent machines in the sense of von Neumann (1958) and Günther (1963).

This is my critical interpretation of Latour. On this basis, I suggest carrying out a twofold extension of Latour's approach, namely, in the direction of AI and psychoanalysis. I will show that it is possible to re-interpret some fundamental concepts of psychoanalysis by using Latour's anthropology of science. The remarkable feature that emerges from this re-interpretation is that the unconscious is essentially technical; the unconscious presupposes and produces non-humans, i.e. artifacts.

My strategy is clear. I use Latour's anthropology of sciences as the mediation between psychanalysis and AI. From this point of view, I will show that the concept of "algorithmic unconscious" is plausible and can be a useful tool in analyzing algorithm behavior.

The mirror stage: from technology to the unconscious. I propose here a new interpretation of Lacan's mirror stage. My argument will be organized in two phases. In the first phase I will present a short description of the mirror stage according to Lacan's texts. In the second phase I will develop an interpretation of the mirror stage in terms of Latour's anthropology of sciences.

The mirror stage is the starting point of Lacan's psychoanalysis, in a historical and theoretical sense. It is a complex dynamic of looks, postures, movements and sensations that concern the child between six and eighteen months of life (Lacan, 1966a; see also Roudinesco, 2009). The child does not speak and does not yet have complete control of his/her body: he/she can barely stand up, cannot walk and must be supported by the adult. The child is not autonomous. This is a fundamental anthropological fact: the human being is born with a body that does not immediately control, is late in development and needs adults for a long time (see Bolk and the concept of "fetalization").

When such a small child looks in the mirror, he/she first sees another child who reproduces his/her gestures and movements. He/she and the other in the mirror are synchronized: if one raises an arm, the other does the same; if one turns the mouth, the other does the same, etc. The child then tries to get to know the other child but encounters an unexpected resistance: the cold and homogeneous surface of the mirror, which is an impassable border. The child tries to get around the mirror, but this attempt fails. He/she cannot find the other child in the mirror. At this point, something happens that clearly changes the child's perception in that situation. The child turns his/her gaze in a different direction and sees other things that surround him/her and the adult who supports him/her reflected in the mirror. Thus, the child realizes that he/she sees duplicates because the mirror produces a doubling of the real. Everything is reflected in the mirror, and the reflection doubles everything. Everything is doubled, except for one thing: a face, the face of the other child. Everything is doubled, except for that childish face. The child cannot find the mirrored child in the surrounding environment. For this reason, in this unique element, the child recognizes his/her face and realizes the first fundamental distinction between himself/herself and the rest of the world. Thanks to the mirror, the child perceives his/her body as a unity, and this fact gives him/her satisfaction and pleasure. Even if he/she cannot speak, he/she enjoys.

In Lacan's view, the mirror stage reveals the tragedy of the human quest for identity. Human being can grasp his/her identity only by passing through the other, something external, that is, an artefact: the mirror. The subject is originally alienated. The mirror image is not only external to the subject but is also false. The mirror distorts things. It gives us a reversed image of ourselves and the things. This means that the child is separated from his/her identity; the identification of the ego (Je) always must pass through the other, the imago in the mirror (moi), the "ideal ego". Here is the paradox: in order to have an identity, the subject must alienate himself/herself: the imago remains unreachable and risks turning into a paranoid delusion. The Lacanian subject is born divided, split, alienated from the beginning.

With the mirror stage, Lacan offers us a materialistic theory of subjectivity. The ego (moi) is the image reflected in the mirror. Thanks to the mirror, the child uses his/her imagination to give himself/herself an identity. This first imago will then be enriched through the comparison with others. The child defines himself/herself through imagination and tends to present himself/herself to others through the imaginative construction that he/she created. The stage of the mirror is the initial moment of every intersubjective relationship through which the subject identifies itself. For Lacan, the psychic development of the human subject passes from the identification with the imago in the mirror to the imaginary identification with other persons. After the mirror phase, the child passes through a series of identifications: first with the mother, then with the other members of the family, especially brothers or sisters. However, this process is a source of pain and instability: the "other" self is the source and the threat of the identification at the same time because it is internal and external at the same time.

Here, I see the connection with the other "primordial scene" of psychoanalysis: the Oedipus complex. In the Oedipus complex, the subject identifies himself/herself with the father, who imposes the prohibition of having sexual relations with the mother. However, by identifying with the father, the child no longer identifies with a mirrored image or with a similar one like his/her mother or brother. Through the father, the child identifies with language and culture. Like Lévi Strauss, Lacan thinks that the prohibition of incest is the condition of society and culture. The father has a symbolic function, not an imaginary one. He ends the cycle of paranoid identifications. The identification with the father is symbolic, i.e., an identification with the symbol, with language (Tarizzo, 2003). However, the symbol is also addressing the other; it is a request for recognition and the beginning of a new experience of desire. The patient is a person who has remained a prisoner of his/her identifications. The healing isthe passage from the un-symbolized imaginary to the symbolized imaginary, i.e., a limited imaginary that can accept the social law of the prohibition of incest and be part of a family and a social contest (Rifflet-Lemaire, 1970, p. 138).

Latour's re-interpretation of the mirror stage. Let us now try to re-interpret the primitive Lacan's scene in terms of Latour's concept of *collectif*. Let us read this passage from *The Pasteurization of France*:

There are not only "social" relations, relations between human and human. Society is not made up just of humans, for everywhere microbes intervene and act. We are in the presence not just of an Eskimo and an anthropologist, a father and his child, a midwife and her client, a prostitute and her client, a pilgrim and his God, not forgetting Mohammed his prophet. In all these relations, these one-on-one confrontations, these duels, these contracts, other agents are present, acting, exchanging their contracts, imposing their aims, and redefining the social bond in a different way. Cholera is no respecter of Mecca, but it enters the intestine of the hadji; the gas bacillus has nothing against the woman in childbirth, but it requires that she die. In the midst of so-called "social" relations, they both form alliances that complicate those relations in a terrible way (Latour, 1986, p. 35).

We can say the same thing for psychoanalysis. Psychoanalysis is a technique—this is a central aspect in Freud and Lacan. Psychoanalysis is a technique, precisely in two ways: (a) artifacts mediate the relationship between the analyst and the patient (the setting, for instance); (b) the technique acts in the formation of the unconscious—there is a technical mediation of unconscious.

In the mirror stage scene, the *actors* are three: (a) the child, (b) the mirror, (c) the objects (humans and non-humans) that surround the child and are reflected in the mirror. These actors are all on the same footing: they are neither reducible nor irreducible to each other (*Irreductions*, 1.1.1). The actors define each other, giving each other strategies, qualities and wills.

The child is reflected in the mirror and, with this gesture, creates a network, an association. Properties and qualities begin to circulate between the actors. The child is an *entelechy* (1.3.1). It is a force that wants to be stronger than others, and therefore enrolls other forces (1.3.2), that is, the mirror and the objects that surround it, among which there is also the adult who supports the child. In this relationship (1.3.4), each actor acts and undergoes trials and resistance.

In this perspective, we can no longer say that the imago constitutes the identity of the child. It is the connection between humans and non-humans that constitutes this identity. This means that the imago is not a mere visual or auditory perception, but a complex series of mediations between humans and non-humans. The Lacanian imago (moi) is the product of a technology, or the effect of a technical object: the mirror. The unconscious is the effect of technical and material mediations.[3] Technology produces the imaginary ego (moi) that must be repressed by language. We are overtaken by what we manufacture. The mirror is not a simple tool that acts as a link between the subject and the imago. The mirror is an actor like others. The same thing can be said about Freud: in *Beyond the Pleasure Principle*, the child—Freud's nephew—learns to cope with the absence of the mother through the use of a spool, that is a technical object, an artifact, which he launches and draws to himself. In this game Freud captures the phenomenon of repetition compulsion.

Let us analyze the *collectif* [child + mirror + surrounding objects] through the categories that Latour describes in the sixth chapter of *L'espoir de Pandore*. In this chapter Latour distinguishes four levels of technical mediation: translation, composition, articulation, and Black-Box. Let us try to reconstruct the mirror stage through these four categories.

Each actor has an action program—a set of actions, intentions, goals or functions—that clashes with that of other actors in the network. When this happens, there are two possibilities: (a) the cancellation of one of the forces or (b) the merging of the forces and the creation of a new action program. The condition of (a) and (b) is what Latour calls "translation" (Latour, 2007, p. 188; translation is mine), i.e., a process of mediation and transformation of action programs with "the creation of a bond that did not exist before and that, with more or less intensity, modifies the two original terms" (Latour, 2007, p. 188). In the translation each actor maintains its action program and its objectives, but a connection is built. An essential phenomenon occurs in this process: the passage of qualities and capacities from one actor to another one. The child is no longer only a child, but a child-in-front-of-the-mirror who receives from the mirror certain qualities and abilities—first of all the ability to recognize the duplicates of the surrounding things and himself/herself. In contact with the child, the mirror is transformed: it is no longer a simple object, but the place where the child looks for and finds his/her identity and therefore the enjoyment. It is also the place of a privileged relationship between the child and the adult who holds him/her.

The mirror is no longer the mirror-resting-on-the-table but becomes the mirror-in-front-of-the-child and therefore the mirror-instrument-of-identification. Humans and non-humans have no fixed essences: in the *collectif* every actor undergoes a transformation of its qualities and abilities. The association [child + mirror] is a human-non human hybrid. This hybrid expands later, including other actors, or even other humans-non humans hybrids, i.e., the surrounding objects. These actors play a very important role in Lacan's *collectif* because it is thanks to their presence and reflection in the mirror that the child can identify himself/herself with the mirrored image. Hence, to read the mirror stage in Latourian terms means to overcome a rigid subject/object dualism and to understand all the complexity of the *collectif*, that is, to understand that subject and object are not innocent inhabitants of the metaphysical world but "polemical entities" (Latour, 2007, p. 314), i.e., dynamics of forces and resistances. The final result of the mirror stage is the identification with the ideal ego, but this identification is accomplished neither by the child nor by the mirror nor by the surrounding humans and non-humans, but by the associations of all them. "Action is not simply a "property of humans, but a property of an association of actors" (Latour, 2007, p. 192).

There are two forms of translation. Latour calls the first "composition". Every actor has an "action program": the child is reflected in the mirror and is pleased to see himself/herself, while the mirror produces images and the other human and non-human hybrids interact in several different ways (the adult holds up the child, talks to him, and this influences the child's experience, but the child can also be distracted by other objects reflected in the mirror, such as a toy, etc.). The process of translation and mutual adaptation between action programs goes on until the child recognizes himself/herself in the mirrored image. However, this is a precarious equilibrium. New identifications take place.

The second form of translation is called "articulation". By "articulation" Latour means that the sense of the actions within the network depends on humans-non humans relations. Non-humans can play an active role in these relations. The sense of child's actions is created by the mirror. The child is posed by the adult in front of the mirror and looks at it. The mirror produces the doubling that triggers the child's experience of auto-recognition and identification. By reflecting the image of other human-non human hybrids surrounding the child, the mirror leads the child to believe that the only image that is not a duplicate is his/her image, his/her duplicate. This process can be described as follows:

mirror → mirrored objects (humans and non-humans) → child's auto-recognition → child's identification → distinction between ideal ego/external world → *imago*

The relationship with artefactsprecedes and determines the subject's identification process. There is no sovereign subject that creates meaning. There is instead a technical object (the mirror, an artifact) that produces what Latour calls an *articulation*, a

connection between humans and non-humans that produces new meanings and identifications. Understanding the meaning of an action does not mean investigating the mind of the person who performed it. It means carefully analyzing the processes of translation, composition, and association between humans and non-humans in a given situation.

The last step of our scheme is the fracture between the ideal ego and the external world. The child becomes paranoid: he/she tends to identify himself/herself with an abstract imago and to separate himself/herself from the rest of reality. This fracture completely covers and eliminates the mediation between humans and non-humans that we have just described. Everything is reduced to the ideal ego and the subject/objects dualism. Now, I interpret the mirror stage outcome by using Latour's fourth category, la *mise en boîte noire*, the Black-Box, "an operation that makes the joint production of actors and artifacts totally opaque" (Latour, 2007, pp. 192–193).

This Latourian category is very important, and I will say more on it later. My thesis is that the unconscious is a *mise en boîte noire*, a Black-Box. Latour gives us—even if he does not intend to do that—the keys to a new phylogenesis of the unconscious and therefore the beginning of a new kind of psychoanalysis.

Latour's idea is simple: the technical artifact hides the set of practices that constitute it—the network of mediations. The final results (for example, the microbe in Pasteur) produce a sort of paradoxical feedback: they cover and hide the interactions, processes and dynamics that produced them. The last phase of the work hides the path which leads to it. "When a machine works effectively, when a state of affairs is established, we are only interested in the inputs and outputs, not on their internal complexity. This is how, paradoxically, science and technology know success. They become opaque and obscure" (Latour, 2007, p. 329). When a scientific fact or an artifact is established and "closed", the *collectif* disappears being crystallized in its outcome —it is "reduced to a single point", says Latour.

Now, there are two Black-Boxes in Lacan's mirror stage: (a) the first coincides with the imago itself, which hides the mirror and the rest of the surrounding world. The imago produces the auto-recognition and the identification that are abstractions from the technical and material conditions that constitute them. The child removes the mediation of non-humans like the mirror, and therefore he/she distinguishes himself/herself from them. In other words, the image that constitutes the child's identification is also what blinds the child and makes him/her incapable of grasping the imaginary nature of his/her identification. This first Black-Box is weak because it closes and reopens many times: the child goes through many different imaginary identifications. (b) The second Black-Box is much more stable and coincides with the transition from the imaginary to the symbolic, therefore with the Oedipus complex. The symbolic—the Name of the Father, more on it later—"closes" the mirror stage making it a Black-Box. In fact, according to Lacan, the symbolic interrupts the imaginary identifications. This interruption coincides with the *Spaltung*, the repression. The symbolic removes the imaginary making it a symbolized imaginary. Reduced to a Black-Box, the imaginary can be limited, removed. This operation is the origin both of the distinction between conscious and unconscious, and of a new form of unconscious. From now on, we distinguish the unconscious that speaks and the unconscious that does not speak.

The oedipal complex: from the unconscious to technology. The Oedipal complex is the fundamental structure of emotional and interpersonal relationships in which the human being is immersed. Freud (2005, 2011, 2012) has hypothesized that the Oedipus complex occurs when the child is three to five years old. This psycho-affective organization is based on the attraction toward the parent of the other sex and the jealousy and hostility toward the parent of same sex. The core of this organization is the prohibition of incest. According to Freud, the Oedipus complex is a fundamental element in the development of the human personality, and if it is not overcome, it constitutes the basic nucleus of all psychopathologies. The entire original phantasmal world of humans is related to the Oedipus complex. The formation of the super-ego is also seen as resulting from the introjection of the paternal prohibition of having sexual relations with parents, brothers, and family members in general.

I will proceed in the following way. I will give, as in previous section, a description of Lacan's interpretation of the Oedipus complex. This does not want to be an exhaustive analysis of the Oedipus complex, but only a short description of how Lacan interprets this primordial "scene". Secondly, I will re-interpret Lacan's Oedipus complex by using Latour's actor-network theory.

Following Lévi-Strauss (1955, 1962), Lacan claims that the prohibition of incest constitutes a universal law that differentiates the state of human civilization from the state of nature. In a series of articles, Lacan introduces the expression "Name of the Father" which defines the acceptance of the social law and marks the passage from the potentially psychotic pre-human condition, that of the mirror stage, to a real human condition. In fact, in his opinion, the psychotic has not internalized the "Name of the Father". The originality of Lacan's interpretation compared to Freud's is that the "Name of the Father" does not coincide with the actual father.

In Lacan's view, the mother and child live a symbiotic relationship that breaks with birth. Both have a kind of nostalgia for this original condition and want to recreate it. Weaning is the traumatic phase: the contiguity between the bodies, maintained by breastfeeding, is interrupted. Both the mother and child have a regressive desire: they want to return to the situation of original dependence. This is a desire for identification; the child identifies himself/herself with the mother (Lacan, 1966a, pp. 35–45).

The "Name of the Father" breaks this situation and prevents the regressive impulse. What characterizes Lacan's interpretation is that the paternal prohibition is considered in symbolic terms. This allows Lacan to apply the Oedipus complex both to males and females. The father represents the language, and therefore the social law of coexistence. Lacan (1998, 1966b) introduces the concept of the "Name of the Father" that is based on French homophony between *nom* (name) and *non* (no, negation) to highlight the legislative and prohibitive role of the society. The "Name of the Father" is the original repression; it diverts the immediate and original impulse of the mother and child. This suspension of the impulse opens the space of the sign, of the appearance of language. Shortly, Lacan re-interprets the Oedipus complex through linguistics and post-modernism. The Oedipus complex and the "Name of the Father" mark the passage from the imagi"nary of the mirror stage, and the series of identification that it produces, to the symbolic.

Let us clarify this thesis. What Lacan has in common with many other important interpreters of Freud "is the claim that Freud's most original and important innovations were obscured and compromised by his effort to embed psychoanalysis in biology and thereby to scientize his vision of the psyche" (Mitchell and Black, 1995, p. 195). Lacan re-interprets Freud's conception of unconscious through Saussure's linguistics and Lévi-Strauss's anthropology. He argues that the essential Freud's discovery is a new way of understanding language and its relation to experience and subjectivity. The famous statement "the unconscious is structured like a language" means that the unconscious is another language, another logic, independent of the subject, but that reveals the truth of the subject. The

condensation is therefore conceived as a metaphor (synchronic order), while the displacement as a metonymy (diachronic order). In the first case, the linguistic signifiers are superimposed, juxtaposed, synthesized, while in the second an exchange takes place, that is, a substitution of one signifier with another. Metaphor and metonymy are the two axes along which the Lacanian unconscious works. The slip should therefore be interpreted as a process of segmentation and re-structuring of the signifiers.

This thesis could not be understood, however, without mentioning another crucial aspect: the fracture between *signifier* and *signified*, which Lacan takes from Saussure. According to the Genevan linguist, the signifier is the phonological element of the sign; it is the *image acoustique* linked to a signified, the immaterial meaning. Re-interpreting Saussure, Lacan introduces the following formula:

$$\frac{S}{s}$$

In this formula, "S" indicates the signifier and "s" the signified. The formula affirms the primacy of the signifier over the signified, i.e., the primacy of the normative, mechanical, and material dimensions of the language. The signifier is a meaningless material element in a closed differential system, i.e., the structure. Thus, the signified (and so the meaning, the subject, the ego) is only a secondary effect of the combination and re-combination of signifiers. There is never a full, absolute signified. The signified is something that is always "pushed back" by the succession of signifiers and shaped by the "symbolic chain". Whereas Saussure places the signified over the signifier, "Lacan inverts the formula, putting the signified under the signifier, to which he ascribes primacy in the life of the psyche, subject and society" (Elliott' 2015, 106).

Lacan re-interprets the Saussurian distinction between signifier and signified in terms of *repression*. The unconscious is at the same time what guides the game of combination and re-combination of signifiers (the unconscious that speaks) and what is repressed and censored by signifiers (the unconscious that does not speak). The meanings produced by the symbolic chain tell us about a more fundamental reality: the enjoyment (*jouissance*), that is, the unconscious that does not speak. The symbolic chain is repression because it "saves" the subject from the enjoyment and allows him/her—through the analysis—to build a new relationship with the force of enjoyment, namely, with the drive.

This is the meaning of Lacan's interpretation of the Oedipus complex: The "Name of the Father" (the first essential signifier) breaks the imaginary union between the child and the mother—the enjoyment—and imposes the social law. Metaphorically speaking, the signifier is like a strongbox; there is a bomb in this strongbox, and this bomb is the enjoyment, the pure desire. The analytic process intends to open the strongbox and disarm the bomb. Only through the interpretation of the symbolic chain and the game of signifiers the patient can recognize his/her desire and enter in relation with the enjoyment in a good way—hence the symptom becomes *sinthome* (Di Ciaccia, 2013).

Let us summarize the result of our analysis. Language is the result of a deviation. Lacan re-interprets the Oedipus complex from a linguistic point of view. It is not the language that produces the deviation of the desire, but the reverse: the deviation of the primitive desire for identification is the cause of the emergence of the sign and language, that is, what Lacan calls the "signifying chain". What do we see here? An instrument, a technology, namely language, comes from an unconscious dynamic. The unconscious produces a technique and therefore also new cognitive abilities. The mirror and language show that the unconscious can be externalized in artifacts that strengthen or simulate our activities and emotions, etc. The Lacanian interpretation of the Oedipus complex is much less "mythical" and sex-focused than the Freudian one. Indeed, it helps to "demythologize" and criticize Freud's point of view.

The oedipal complex in Latourian terms. Here, I stop the reconstruction of Lacan's thought and start its re-interpretation in Latour's terms. As we said before, the child wants to reconstruct the symbiosis with his/her mother (and vice versa), but his/her desire is blocked by the father. A deviation of desire takes place. In Latour's terms, this deviation of the child's action program opens the door to the intervention of another actor, which is the language, i.e., a technology, an artifact. For Latour, the language is not only a set of symbols which are connected by deterministic rules, but an actor similar to all other human and non-human actors (Latour, 2007, p. 150). Language does not imply any abstraction from the world, but it is rooted in the world and has meaning thanks to the connection with other actors.

This view is also very close to (and very far from) Lacan. It is very close because Lacan also thinks that language surpasses humans and envelops them. It is very far because Latour does not rigidly interpret language as a game of differences based on rigid rules. For Latour, Lacan is still a victim of the "modern compromise".

Given this, we can describe the Oedipus complex as a *collectif* composed by four actors [child + father + mother + language]. The action programs of child and language connect each other. Therefore, a process of translation and mutual adaptation begins. From a Lacanian point of view, the purpose of the child is the reunion with the mother, while the purpose of language is the signifying chain. In the translation process, an exchange of qualities and abilities takes place, in both directions. Both actors transform themselves. This means that if the child becomes the language, the language becomes the child. The child is no longer *infans*: he/she enters into connection with the word and then he/she knows a new type of desire. Thanks to language, the child can dominate his own narcissistic impulses and to live with other human beings in a community. Thanks to language, the child abandons the narcissism of the imago and enters the social world. Hence, the child experiences finiteness and search for recognition. The intervention of the non-human (the language) resolves the contrast between humans. The situation can be described by the Fig. 2.

However, this re-interpretation still lacks an essential point. Following Latour, we must also go in the other direction: if the child becomes a language, *the language becomes the child*. A technical object—the language—acquires some of the child's characteristics. A translation takes place. What does it mean?

That "the language becomes a child" means two things: (a) that the child acquires certain characteristics of the language, so that

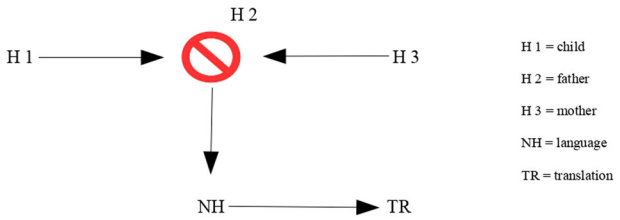

Fig. 2 The scheme of oedipal complex according to the re-interpretation of Lacan through Latour'anthropology of sciences. The Oedipus complex is a collectif composed by four actors [child + father + mother + language]. Language is an actor similar to all other human and non-human actors in the network.

his/her psychic dynamics is structured like language, as Freud proved and Lacan emphasized; (b) that language takes on human characteristics, that is, it assimilates humans way of being and becomes its expression.

The human being acquires the characteristics of language, and therefore the social law of the prohibition of incest, i.e., the "Name of the Father". In this way the child abandons the mirror stage and avoids psychosis. The unconscious *becomes* a language.

The language has a "double existence": on the one hand, it is a technology (Ong, 1982) that represents the social law and is structured as the significant chain; on the other, it assimilates the characteristics of the child. This is the crucial point of our re-interpretation. An artifact can assimilate a human strategy, where "assimilate" means "simulate" or "respond to a human need". While the child assimilates the repressive mechanism of the significant chain, language assimilates and transforms the search for identification of the child. Language is not just a a set of sounds or marks. The identification—following Lacan—becomes symbolic. However, this does not mean that language seeks identification for itself. The identification process always remains unidirectional: from humans to language. But language does not play a passive role in this process.

If we follow the symmetrical anthropology of Latour, we must admit this exchange of properties in both directions. This does not mean thinking of the language in magical terms, as if the language were becoming a human subject in all respects. Affirming that the language can assimilate human characteristics and simulate them means to abandon what Latour calls "modern compromise" (see Latour 1991) and to understand that language and human subject are two systems of forces in constant interaction and exchange. The *collectif* is not a metaphysical entity, but a dynamic theoretical model to explain associations of human and non-human actors. When Latour speaks of humans and non-humans, he does not want to designate subjects and/or objects. The concept of *collectif* arises from a profound critique of traditional metaphysics and its classic pairs of opposites: subject/object and fact/value. If the subject and the object are "closed" and opposing concepts, the human and the non-human are instead "open" concepts, in constant interaction and mutually defining each other. Only if we admit this interaction, we can have a theoretical model that explains the phylogenesis of AI and so justifies the originality of AI. My thesis, in fact, is that AI prolongs the movement of language.

I use Latour to highlight a crucial aspect: a technology (language) arises from an unconscious tension. By saying this, I am not trying to explain the general relationship between machines and their environment, but the relationship between machines and humans from the point of view of psychoanalysis. I absolutely do not want to explain the relationship between machines and the world through the Oedipus complex. Obviously, this relationship is a much larger and more complex issue. My goal is to try to clarify the relationship between a particular form of technology (AI systems) and humans through the dynamics of subjectivation described by Lacan in the mirror stage and in his re-interpretation of the oedipal complex. What does Latour add to Lacan? Latour's actor-network theory helps complete the Lacan's model of identification. Thanks to the Oedipus complex, the child becomes a social and speaking being. Then the language becomes a child, that is, it plays an active role in the identification process. A technical object is the expression of an unconscious dynamics, which structures it, that is, defines its meaning.

Stating that language can assimilate human characteristics is tantamount to deleting the distinction between *langue* and *parole*, i.e., the distinction between the structure and the act as introduced by Saussure (1949). This distinction does not exist at all. We have to think the relationship between *langue* and *parole* as a translation process that goes in both directions. *Langue* is translated into *parole*, and *parole* into *langue*. The language is a perennial negotiation between these two levels. The two actors shape each other.

Now, this reinterpretation of the Oedipus complex is entirely in line with Tauber's "cognitive unconscious" (Tauber, 2013; see also Tauber, 2010). Lacan's reinterpretation through Latour allows us to abandon an exclusively sexual conception of the oedipal complex (as the sexual repression by the father in Freud's terms) and to move towards a much more integrated model of psyche, not only in itself (relationship between conscious and unconscious activities), but also in relation to the technology. The Oedipus complex becomes a phase of the identification process. If we interpret the identification process as the first elementary relationship between the conscious and the unconscious (the one that defines both at the same time), then the Oedipus complex is nothing more than the moment when this relationship becomes *an active collaboration*—not just a repression—*that gives rise to new artifacts and new cognitive processes*. In other words, the tension between Id and ego produces a technology that helps human beings develop their cognitive abilities. The relationship between conscious and unconscious is dynamic and complex, and requires an integrated approach.

The difficulty of being an AI

In this section, I apply the results of the re-interpretation of Lacan to AI. I claim that AI is a new stage in the process of human identification, that is, a new development of the unconscious, the "algorithmic unconscious". In AI, the machine responds to a human desire for identification. AI extends the same movement that is at work in the Lacanian interpretation of the mirror stage and the Oedipus complex and which Latour's reading helps us to clarify. The unconscious is at the same time the effect of a technological mediation and the origin of a new form of technology.

A possible objector would say that our thesis is a simple metaphor. Metaphors are useful, but substantial mechanisms must be provided to make the underlying analogy theoretically or conceptually significant. The objector would say: If an identification process goes from humans to machines, then the same process goes in the opposite direction, that is, from machines to humans—the machines identify with humans. However, this conclusion produces absurd consequences: How can machines identify with humans? Does the machine identification process have the same structure as the human one? If human identification produces language and AI, what does machine identification produce? Our hypothesis is contradictory.

The only way to avoid these consequences is to claim that the identification process is unique and goes from humans to machines.

A useful line may be that of Mondal (2017). Mondal's methodological approach is that of cognitive sciences, although it has a lot in common with psychoanalysis. The fundamental connection between Mondal's approach and psychoanalysis is the study of natural language as a means of understanding and deciphering the "mentality", that is, the set of thoughts, ideas, emotions and feelings. However, Mondal goes further than psychoanalysis because he states that natural language and mind are closely connected, which means that the complex structures of natural language correspond to mental structures and conceptual relationships, what Mondal calls "forms of mind". Different human groups correspond to different natural languages and therefore to different "forms of minds". This means that mental structures are "interpreted structures", namely structures that can be revealed

through the analysis of the natural language. This does not mean —Mondal underlines—that mental structures are determined or caused by natural language. Nevertheless, mental structures can be described and investigated through natural language. Given these premises, Mondal investigates the possibility of interpreting non-human organic and non-organic "other minds", including AI, through the analysis of natural language.

According to Mondal, nothing prevents us from identifying, through the analysis of natural language, mental structures that can also be identified in machines. In other words, what makes a set of circuits and logic an AI system is the human interpretation of this set. What makes a function a computation is not the function itself, but the human interpretation of this function. We can attribute a "mind" to a machine without necessarily anthropomorphizing the machine. We can detect mental structures in machines without having to argue that these structures are a simple consequence of human interpretation. Latour's theory is completely in line with this way of thinking.

Now, I take into consideration the example of a neuronal network for the recognition of vocal language as described in Palo and Mohanty (2018). The desire for identification from humans to machines develops in four distinct phases:

1. Project: Humans design and build a machine (the neuronal network) that is as intelligent as they are.
2. Interpretation: Humans (designers, engineers or users) interpret the functioning of the machine; this means that they observe the behavior of the machine and evaluate (a) if they can recognize in this behavior patterns that are similar to theirs, (b) to what extent the machine is able to collaborate with them (in our case, to what extent the machine is able to recognize the voice and join a conversation). This evaluation is based on human and technical criteria (in our case, natural language, emotions, tone of voice, context, etc.).
3. Identification: If the evaluation is positive, humans attribute to the machine their quality and therefore a mind (in our case, the ability to speak and join a conversation); this does not mean that the machine becomes a human being seeking identification, but that humans interpret the behavior of machine in this way. If the evaluation is negative, humans go back to the project and re-start the interpretation.
4. Mirror effect: The machine is able to interpret humans, namely, to assimilate the human way of speaking (this is the machine learning process) and develop it autonomously, as occurs today in various neuronal network for voice recognition—but also in networks like *Google Translate*. AI is an interpreted and interpretive technology; this is the "mirror effect". Furthermore, the mirror effect is also subject to a human interpretation. This triggers a new cycle of interpretation and identification. For instance, humans can think of themselves as machines. Current cognitive science that encompasses AI, psychology, neuroscience, linguistics, philosophy and other related disciplines think of the human mind as a computing machine (see Boden, 2006). This means that humans have also delegated some of the qualities and capabilities of machines to themselves.

These four phases can be interpreted as a *collectif* in Latourian terms. In each of these phases conscious and unconscious processes interact and cooperate. The root of this process if the human projective identification. The human being is the main actor. The directionality of the process is unique. Even if the machine is able to reproduce and develop the process autonomously, the process depends on the human interpretation of machine behavior. This act of interpretation is not arbitrary, as if a machine could be intelligent for me and stupid for someone else. As I said, the interpretation is based on human experience and technical requirements. The interpretation must respect two major constraints: logic and machinery, i.e., the embodiment of the logical structure.

In the next sections, I will proceed as follows: (a) I will give a brief description of what a computational system is; (b) I will illustrate the two fundamental limits of the identification process, logic and machinery.

Computational systems. Computational systems have at least three fundamental dimensions: (a) formal computation, (b) physical computation, (c) experimental computation (see Primiero, 2020). These systems are abstract, physical and experimental artifacts at the same time. In other words, each computational system is a mathematical structure that is realized in a machine (the computer) by which we carry out scientific tasks using models and simulations (Turner, 2018; Piccinini, 2018). The algorithm always has three lives: an abstract one, a physical one and an experimental one. These are three necessarily connected dimensions.

I do not want to carry out a detailed analysis of every aspect of computing here. This is not the purpose of the present paper. There are already important studies on the subject (see Boolos et al. 2007). I want to limit myself to some historical considerations. Priestley (2011) has shown that, in the history of computer, the paths of engineering and logic have developed in parallel for a long time.

Babbage's machine was able to perform some mathematical operations: it was the embodiment of some algorithms, but it could not be programmed for more complex tasks. It was therefore not really a computer—despite the similarities. The purpose of the first real digital calculators (Mark I, ENIAC, the Bell Labs Relay Machines, etc.) was to automate scientific calculation in a new way. Scientists ask these machines to extend human computation ability, that is, to carry out longer and more complex computations. Along this path, mechanical computation and logic (the computability theory) have ignored each other for a long time. "The logical investigation of the concept of effective computability and the development of the machines were largely independent of each other" (Priestley, 2011, p. 122). Since 1950, computing engineering developed enormously thanks to the Second World War. Of particular importance was the development of the so-called stored-program design, especially with von Neumann's EDVAC. However, the awareness of the importance of unifying these engineering efforts with logical computation emerged only at an advanced stage of this development. Computers were not born as "logical machines". "Turing's 1950 paper was not specifically logical or technical, but rather a philosophical contribution to the discussion of the cybernetic question of whether machines could think" (Priestley, 2011, p. 155). Only with the development of the ALGOL programming language (three versions: 1958, 1960, 1968) logic begins to penetrate the world of computing machinery.

Therefore, the historical research reveals an essential aspect. The relationship between logic and computing machinery is not easy at all. Circuits are not analogous to mathematical formulas; they are ontologically different. There is a tensional relationship between logic and computing machinery at the core of any computational system. The analogy between coding and logic is the result of a long intellectual work. What we commonly understand as software (programming languages) are the result of this story. The analyses contained in Haigh (2019) confirm this point.

Logic and machinery: the limits of the identification. In a computational system, the human desire for identification has to

face two essential conditions: logic and machinery, i.e., the material apparatus that embodies the abstract logical structure. In addition, these two poles contrast with each other causing tensions. AI is a field of forces in which there are three actors (human desire, logic and machinery) in constant competition.

The machinery is not an inert object. The materiality of the machine is a text to be deciphered. The machine is an artifact, and therefore it is the embodiment of a set of different types of normativities (imaginative, technical, social, socio-technical, and behavioral), as Grosman and Reigeluth (2019) show. It is interesting that these normativities can compete, collaborate or coexist without interacting inside the machine. The human desire to identify with the machine must deal with the limits posed by the possibilities offered by the material structure of the system.

Even logic is not a peaceful field. The logical mathematical theory of computation was the result of a deep crisis in the foundation of mathematics at the end of the 19th century. This crisis became a "war" between different conceptions of truth, correctness and infinity in mathematics (Lindström et al. 2009; Adams, 1983; Copeland et al. 2013). The current manuals of logic and computability theory are Black-Boxes that hide this conflict, as well as the different strategies that have been used to solve it. In general, the logic behind our computers and AI is the logic of Frege and Russell, of Turing and von Neumann (Printz, 2018).

Logic and material possibilities establish what is computable and what is not computable. Human desire must constantly negotiate with these limits to run the machine and make it intelligent. The set of these complex negotiations, which can fail or succeed, is what I call "algorithmic unconscious". This is evident, in my opinion, if we look at the overall structure of an AI system and at all the levels that compose it: the quantum level (transistors and silicon, atoms and electrons), the physical architecture (circuits architecture and wiring, cache memory, etc), the microprocessor architecture and hardware programming, the hardware/software interfaces, up to the more abstract logical structure of the machine.

All of these negotiations are repressed. The complexity of the relationships between these levels does not come to light in the programming and use of systems. Firstly, it does not come to light in the programming because programming necessarily imposes the distinction between an internal language (the compiler) and an external language (the high-level language) (Printz, 2018, pp. 194–204). This means a1) that the programmer does not necessarily need to know the structure of the compiler and the machinery in order to program; a2) that a high-level language can be applied to multiple different systems; it is independent of the application context (the compiler is not). The same can be said for the customer/user of an AI system. The customer/user does not need to know the internal structure of the system and the engineering that defines it. This "Black-Box effect" hides the tensions between logic, machinery and human desire.

Therefore, I want to underline basically three points: (a) the desire for identification is at work in AI, as demonstrated also by the excess of imagination that is produced by films and novels on this theme and which often exceed the reality of AI; (b) from a psychoanalytic point of view, the desire for identification has a structure and a phylogeny that can be described in Lacan's terms re-interpreted through Latour; (c) if we accept this structure and this phylogeny, AI can be understood as a new stage in the history of the human search for identification. After the imaginary and symbolic registers, AI is the third register of identification.

Miscomputation and information

We said that the tensions between logic, machinery and human desire remain hidden by the system's Black-Box effect. This statement would remain a simple theoretical hypothesis if it did not give us new tools to understand two important phenomena: miscomputation and information noise. From a psychoanalytic point of view, these two phenomena, generally considered marginal, acquire decisive importance.

Miscomputation. The literature on miscomputation is broad. Piccinini lists many cases of miscomputation, including a failure of a hardware component, a faulty interaction between hardware and software, a mistake in computer design and a programming error, etc. (2018, pp. 523–524). Floridi et al. (2015) distinguishes two main types of malfunctioning: dysfunction and misfunction. A dysfunction "occurs when an artifact token *t* either does not (sometimes) or cannot (ever) do what it is supposed to do", whereas a malfunction "occurs when an artifact token *t* may do what it is supposed to do (possibly for all tokens *t* of a given type *T*), but it also yields some unintended and undesirable effect(s), at least occasionally" (Floridi et al. 2015, p. 4). Software, *understood as type*, may misfunction in some limited sense, "but that it cannot dysfunction". The reason for this is that "malfunction of types is always reducible to errors in the software design and, thus, in stricter terms, incorrectly-designed software cannot execute the function originally intended at the functional specification level" (Floridi et al. 2015, p. 4).

I think that these studies have the limit of providing a merely technical description of miscomputation. They build classifications, but do not explain what really miscomputation is. From my point of view, miscomputation is a fundamental phenomenon in order to understand the relationship between AI and its environment. Miscomputations must therefore be studied as if they were Freudian slips or failed acts. They express the tensions between human desire, logic and machinery, at different levels (design, implementation, hardware, testing, etc.), that cannot be controlled and repressed. As Vial (2013) points out, the tendency to have errors and bugs is an ontological feature of software and AI. There will always be in any systems an irreducible tendency to instability, to the deviation from the design parameters and requirements, and thus from the "normal" functionality. "A computer cannot live without bugs. Even if computer programs are written by humans, they are never entirely controllable a priori [by humans]" (Vial, 2013, p. 163; translation is mine). AI instability is another name of the algorithmic unconscious.

Information and noise. The analogy between the concept of Black-Box and that of repression could raise objections. Trying to clarify, I still propose to follow the well-known book *Laboratory Life*.[4]

According to Woolgar and Latour (1979)[5], behind the scientific fact there is always the laboratory, namely, a set of times, spaces, practices, groups of humans and non-humans, flows of money and information, negotiations and power relationships. However, the laboratory is invisible in the scientific fact. In the somatostatin there is not Guillemin's work, even though this latter was necessary in order to create somatostatin as scientific fact. Somatostatin is a Black-Box: nobody puts its existence into question. It is evident. Nevertheless, if some research results put this existence into question, the Black-Box would be reopened and the debate would restart. Woolgar and Latour (1979) claim that to "open" a fact means to continue discussing about it, whereas to "close" a fact means to stop discussing. The controversies between researchers are essential. This is a very complex dynamic: the more a fact is important and attracts the attention of researchers, the less it will become a stable Black-Box because it will be constantly "reopened" and discussed again. Instead, when a fact does not raise the interest of researchers, it is

"closed" very quickly and becomes a Black-Box. Among the facts, there is a relationship of "gravitational attraction"; a re-opened fact "attracts" other facts and forces the researchers either to reopen or to close them.

Can we compare this idea of Black-Box with the notion of repression in psychoanalysis? I think so, because what drives the researchers to close the debate and "crystallize" it in a Black-Box is the fear of disorder, i.e., the constant uncertainty deriving from the open debate. This aspect is evident in the concept of noise analyzed by Woolgar and Latour (1979, pp. 45–48), deriving from information theory. The concept of noise can be summarized in the idea that information is defined and measured in relation to a background of equally probable events. The more the noise—that is the confusion—decreases, the more information is clear and convincing. Information is the most probable event.

As Woolgar and Latour (1979, pp. 50–52) claim, the concept of noise indicates two types of factors: the first type is the set of equally-probable events, while the second is the set of factors—rhetorical, technical, psychological, competitive, etc.—that influence and determine the probability that an event has to become a scientific fact. The scientific fact is the most probable event among others, *but this probability is defined by the second type factors*. Information without noise is impossible because *information is the effect of noise*. Disorder is the rule, while order is the exception. Scientists produce order from disorder. The analogy between information and the game of Go is essential; the Go is a game that begins without a fixed pattern and becomes a rigid structure (Woolgar and Latour, 1979, p. 60). As a result, noise is at the same time what jeopardizes and what allows information (Woolgar and Latour, 1979, p. 49). The presence of noise jeopardizes also the translation of information in a *collectif*. This is another sense of the "algorithmic unconscious".

The birth of a Black-Box coincides with the separation of information from noise. An event is considered more probable than others. And yet, this process hides the fact that the noise is the condition of information. This oblivion is properly the unconscious, that is, the denial of disorder, not the disorder itself—the repression. A Black-Box is the denial of disorder. Woolgar and Latour (1979) use an economic language very similar to the psychoanalytic one: reopening a Black-Box requires an "investment" that is too large in psychological, economic and social terms.

Conclusions. A new research project

I think time has come to modify what is meant by "artificial intelligence". The goal of this paper was to open a new perspective on AI. Is this goal achieved? I think this is a realistic model, in the sense that it explains the *content* of AI, that is, the originality of AI—what distinguishes it from any other form of technology. The originality of AI lies in its profound connection with the human unconscious. This model traces an AI phylogeny. Furthermore, from its application to AI, psychoanalysis can gain a new field of research and analysis tools. This is a new field of investigation: the psychoanalysis of artifacts. This is an important epistemological result.

This paper is also the first step in a much broader research project concerning the relationship between AI and mental illness. The project aims to study mental illness as a *collectif* of humans and non-human actors. These are the central issues of the project: Can AI help treat mental illness? Can AI systems help doctors understand and diagnose mental illnesses? Do AI systems tend to assimilate these diseases and reproduce them? Can AI systems get sick? Can we apply the category of neurosis (understood as a lack of adaptation to the environment) to AI? Can we apply the category of psychosis (understood as an alteration of the relationship with reality, such as delirium or hallucination) to AI? Do new types of miscomputations emerge?

The project does not want to study simply how AI applies to the analysis of diseases, but also how mental diseases and AI mutually define and transform each other. Studies already exist that analyze how AI can be used by doctors in making diagnosis (Griffiths et al., 2017; Fiske et al., 2019). However, there are no methodologies involving the interaction between patients and AI.

The core of the project is the creation of groups composed by humans with psychiatric problems and AI systems in order to study (a) how AI systems react in relation to certain patient dynamics, and therefore to what extent they assimilate and interpret these dynamics; (b) how patients react to the interaction with AI systems.[6] The AI systems would be machine learning systems for studying natural language and body movements. These systems will analyze patients' language and movements. "For instance, sentences that do not follow a logical pattern can be a critical symptom in schizophrenia. Shifts in tone or peace can hint at mania or depression".[7] A fundamental aspect would be the interaction between patients and AI, for example through the use of tests or free conversations. These tests would be important to understand how AI evolves in contact with patients.

Data availability

All data generated or analyzed during this study are included in this published article.

Received: 19 December 2019; Accepted: 24 March 2020;
Published online: 24 April 2020

Notes

1 By saying this, I do not want to classify Latour, calling him a philosopher. I am just saying that *my point of view* on his work is philosophical. Latour remains an unclassifiable thinker.
2 This is the reason why there is a "priority of the controversies" in Latour's sociology (Latour, 2005, p. 25).
3 An objector could reply that the mirror is only an accidental element in Lacan's text. True identification occurs in the encounter between the child and the mother. However, the mother cannot be seen by the child as another human subject because—following Lacan—the child still lacks the linguistic relationship. This is the point. The mother is considered as an object among many others: a material object that helps the child to produce its identifying image.
4 I do not want to give a too unified image of Latour's work. Latour has in fact criticized the investigations in *Laboratory Life*: see Latour (2011, p. 121).
5 See also Latour (1989, pp. 319–321).
6 See: https://www.healtheuropa.eu/ai-in-psychiatry-detecting-mental-illness-with-artificial-intelligence/95028/.
7 https://www.healtheuropa.eu/ai-in-psychiatry-detecting-mental-illness-with-artificial-intelligence/95028/.

References

Adams R (1983) An early history of recursive functions and computability. From Gödel to Turing. Docent Press, Boston
Boden MA (2006) Mind as machine: a history of cognitive science. Clarendon Press, London
Boolos G, Burgess J, Richard C (2007) Computability and logic. Cambridge University Press, Cambridge (Ist edn. 1974)
Churchland P, Smith P (1990) Could a machine think? Sci Am 262:32–39
Copeland JB, Posy CJ, Shagrir O (eds) (2013) Computability. Turing, Gödel, Church, and Beyond. MIT Press, Cambridge
Cully A, Clune J, Tarapore D, Mouret J-B (2015) Robots that can adapt like animals. Nature 521:503–507
Di Ciaccia A (2013) Il godimento in Lacan. La Psicoanalisi. Studi internazionali del campo freudiano; http://www.lapsicoanalisi.it/psicoanalisi/index.php/per-voi/rubrica-di-antonio-di-ciaccia/132-il-godimento-in-lacan.html
Ellenberger H (1970) The discovery of unconscious: the history and evolution of dynamic psychiatry. Basic Books, New York
Elliott A (2015) Psychoanalytic Theory. Palgrave Macmillan, London-New York
Fiske A, Henningsen P, Buyx A (2019) Your robot therapist will see your now. J Med Internet Res 21:112–124

Floridi L, Fresco N, Primiero G (2015) On malfunctioning software. Synthèse 192 (4):1199–1220
Freud S (2005) The unconscious. Penguin, London
Freud S (2011) Three essays on the theory of sexuality. Martino Fine Books, Eastford (Ist edn. 1905)
Freud S (2012) A general introduction to psychoanalysis. Wordsworth, Hertfordshire (Ist edn. 1917)
Griffiths F, Bryce C, Cave J, Dritsaki M, Fraser J, Hamilton K, Huxley C, Ignatowicz A, Sung Wook K, Kimani P, Madan J, Slowther A-M, Sujan M, Sturt J (2017) Timely digital patient-clinician communication in specialist clinical services for young people. J Med Internet Res 19:112–124
Grosman J, Reigeluth T (2019) Perspectives on algorithmic normativities: engineers, objects, activities. Big Data Soc 1:1–6
Günther G (1963) Das Bewußtsein der Machinen. Eine Metaphysik der Kibernetik. Agis Verlag, Baden Baden
Haigh T (ed) (2019) Exploring the rarly digital. Springer, Berlin
Harman G (2019) Object-oriented ontology. Penguin, London
Lévi-Strauss C (1955) Tristes tropiques. Plon, Paris
Lévi-Strauss C (1962) La pensée sauvage. Plon, Paris
Lacan J (1953–54) La psychanalyse. PUF, Paris
Lacan J (1966a) Ecrits I. Seuil, Paris
Lacan J (1966b) Ecrits II. Seuil, Paris
Lacan J (1998) Le séminaire: les transformations de l'inconscient. Seuil, Paris
Larrey P (2019) Artificial humanity. An essay on the philosophy of artificial intelligence. IFPress, Roma
Latour B (1986) The Pasteurization of France. Harvard University Press
Latour B (1989) La science en action. Introduction à la sociologie des sciences. La Découverte, Paris (Ist edn. 1987)
Latour B (1991) Nous n'avons jamais été modernes. La Découverte, Paris
Latour B (2005) Reassembling the Social. Oxford University Press
Latour B (2007) L'espoir de Pandore. La Découverte, Paris, (Ist edn. 1999)
Latour B (2011) Pasteur: guerre et paix des microbes, suivi de Irréductions. La Découverte, Paris, (Ist edn. 1984)
Lindström S, Palmgren E, Segerberg K, Stoltenberg-Hansen V (eds) (2009) Logicism, intuitionism, and formalism. What has become of them? Springer, Berlin
Lipson H (2019) Robots on the run. Nature 568:174–175
Mitchell S, Black M (1995) Freud and beyond. a history of modern psychoanalytic thought. Basic Books, New York
Mondal P (2017) Natural language and possible minds: how language uncovers the cognitive landscape of nature. Brill, Leiden-Boston
O'Neill C (2016) Weapons of math destruction. Crown Books, Washington
Ong W (1982) Orality and literacy. Routledge, London-New York
Palo Kumar H, Mohanty MN (2018) Comparative analysis of neural network for speech emotion recognition. Int J Eng Technol 7:112–116
Piccinini G (2018) Physical computation. Oxford University Press
Priestley M (2011) A science of operations. Machines. Logic and the invention of programming. Springer, Berlin
Primiero G (2020) On the foundations of computing. Oxford University Press
Printz J (2018) Survivrons-nous à la technologie? Aux sources du cyberespace et des sciences de la complexité. Les acteurs du savoir, Paris
Rahwan I, Cebrian M, Obradovich O, Bongard J, Bonnefon J-F, Breazeal C, Crandall J, Christakis N, Couzin I, Jackson MO, Jennings N, Kamar E, Kloumann I, Larochelle H, Lazer D, McElreath R, Mislove A, Parkes D, Pentland A, Roberts M, Shariff A, Tenenbaum J, Wellman M (2019) Machine behavior. Nature 568:477–486
Ricoeur P (1965) De l'interprétation. Essai sur Freud. Seuil, Paris
Rifflet-Lemaire A (1970) Jacques Lacan. Dessart, Bruxelles
Romele A (2019) Digital hermeneutics. Philosophical investigations in new media and technologies. Routledge, London-New York
Roudinesco E (2009) L'histoire de la psychanalyse en France–Jacques Lacan. Hachette, Paris
Saussure F (1949) Cours de linguistique general. In: Bally C, Sechehaye A, Riedlinger A (eds) Payot, Paris
Tarizzo D (2003) Introduzione a Lacan. Laterza, Roma-Bari
Tauber AI (2010) Freud, the reluctant philosopher. Princeton University Press
Tauber AI (2013) Freud without oedipus: the cognitive unconscious. Philos, Psychiatry, Psychol 20(3):231–241
Turing A (1950) Computing machinery and intelligence. Mind LIX (236):433–460
Turner R (2018) Computational artifacts. Towards a philosophy of computer science. Springer, Berlin
Vial S (2013) L'être et l'ecran. Puf, Paris
Von Neumann J (1958) The computer and the brain. Yale University Press
Voosen P (2017) The AI detectives. As neural nets push into science, researchers probe packs. Science 357(6346):22–27
Woolgar S, Latour B (1979) Laboratory life. The social construction of scientific facts. Sage, Los Angeles

Acknowledgements

This publication is funded with National Funds through the FCT/MCTES - Fundação para a Ciência e a Tecnologia/ Ministério da Ciência, Tecnologia e Ensino Superior (Foundation for Science and Technology/Ministry for Science, Technology and Higher Education - Portugal), in the framework of the Project of the Institute of Philosophy with the reference UIDB/00502/2020.

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.M.P.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.