



The Machine with a Human Face: From Artificial Intelligence to Artificial Sentience

Sylvain Lavelle^{1,2}(✉)

¹ ICAM Paris, Centre for Ethics, Technology and Society,
34 Points de Vue, 77127 Lieusaint, France
sylvain.lavelle@icam.fr

² Ecole des Hautes Etudes en Sciences Sociales,
54 boulevard Raspail, 75006 Paris, France

Abstract. The main challenge of technology is to facilitate the tasks and to transfer the functions that are usually performed by the humans to the non-humans. However, the pervasion of machines in everyday life requires that the non-humans are increasingly closer in their abilities to the ordinary thought, action and behaviour of the humans. This view merges the idea of the *Humaniter*, a longstanding myth in the history of technology: an artificial creature that thinks, acts and feels like a human to the point that one cannot make the difference between the two. In the wake of the opposition of Strong AI and Weak AI, this challenge can be expressed in terms of a shift from the performance of intelligence (reason, reasoning, cognition, judgment) to that of sentience (experience, sensation, emotion, consciousness). In other words, the challenge of technology if this possible shift is taken seriously is to move from the paradigm of Artificial Intelligence (AI) to that of Artificial Sentience (AS). But for the *Humaniter* not to be regarded as a mere myth, any intelligent or sentient machine must pass through a Test of Humanity that refers to or that differs from the Turing Test. One can suggest several options for this kind of test and also point out some conditions and limits to the very idea of the *Humaniter* as an artificial human.

Keywords: Artificial intelligence · Artificial sentience · *Humaniter*

1 The Myth of the *Humaniter*

The possible substitution of the machine for the human as regards the mind functions has been the heart of the research program on Artificial Intelligence (AI) since the outset. The idea is to produce an identical or similar performance in a machine for the logical and noetic abilities (from the Greek *logos*, reason, speech, and *noesis*, thought) said wrongly or rightly “superior”. There are undoubtedly many possible definitions of AI, and, failing to find a simple and unique one, which covers the whole spectrum of reasoning and cognition, it is nevertheless possible to indicate its objective: “The objective of artificial intelligence is, in the long term, to have everything that man can do in terms of reasoning done by a computer system (Ladrière) [1]”. The fact remains that intelligence, especially if it is artificial, is still the subject of many debates, some

going so far as to declare that “Artificial intelligence doesn’t exist” [2]. They mean that the substitution of the machine for the human is not so easy and that, rather than an artificial intelligence, it would probably be better to speak of an “enhanced intelligence” [2].

The problem is even more acute when it comes to substituting the machine for the human as regards other functions, sometimes called “inferior” when compared to those of intelligence. More precisely, these are abilities that can be broadly described as empirical and aesthetical (from the Greek *emperia*, experience, and *aisthesis*, sensation). These multiple aptitudes of experience, namely, sensation, perception, emotion and sentiment, to which consciousness can be added, turn attention to the sentience. In this other dimension, perhaps even more complex, of human life lies the origin of a research program which supplements that of artificial intelligence. One could name it, with all the precautions of use, the *Artificial Sentience* (AS), that is to say the exploration and the transfer of the functions and abilities of human experience and senses to a machine. However, it would be legitimate to ask, critically, if it goes for artificial sentience as for artificial intelligence, so that one could state as well that “Artificial sentience doesn’t exist” [3]¹.

Sentience is a challenge for artificial intelligence, but it can also be presented as its new frontier, and some, like Husain, speak without hesitation of the “sentient machine” [4]. The difficulty of a shift from AI to AS comes from the fact that all human functions and abilities, from the most intellectual to the most sensory, can be transferred to machines. The idea of a total substitution, which concerns both the mind and the body, corresponds to an ancient myth of technology, of the art of production of artefacts and possibly, of artificial creatures [5]. I propose to call it the *Myth of the Humaniter*, that is, the fictitious or imaginary idea of a humanoid artificial creature which allows a total and perfect substitution between human and machine. The Humaniter is an artificial creature which combines and articulates a whole set of human functions and abilities beyond those, more classic, of the Actor or the Producer: the *Reasoner*, for reasoning and intelligence, the *Cogniter*, for knowledge and belief, the *Voliter*, for will and desire, and finally, the *Experiencer*, for experience, sensation, emotion and possibly consciousness. The Humaniter as an artificial creature that replaces the human is held to be a myth, but this does not prevent many researchers from trying to realise it.

The thesis that I would like to defend is the following one: I do not think for a second that the Humaniter can exist, and in particular, that there can be an artificial creature which not only is intelligent, but moreover, can also be sentient. In other words, I think that the Humaniter is indeed a myth and that, just as there is no artificial intelligence, there is no artificial sentience. The major argument in support of this thesis is that, if an artificial creature such as the Humaniter existed, it would be actually a human being. It would be endowed with the same general physical and psychic

¹ As Matson stated, “very little, if any, distinctively human or animal behaviour can be duplicated or possibly simulated by existing machines. No existing machine is sentient, and nothing that any of them could do would go the slightest way toward indicating sentience. But these facts are no interest. We want to talk of possible, conceivable machines”, p. 78 .

constitution, and only its creation would be artificial, that is, non-sexual². Nevertheless, I am convinced that, in a certain type of interaction between the human and the machine, it is possible that the machine passes the test of humanity and can be confused with a human.

It is this point that I propose to explore in the second part of this paper, after having presented in the first part what the Humaniter is, considered in its different dimensions, from intelligence to sentience. In this short study, which is only a kind of brief inventory of some aspects of the problem, I propose after formulating the principles to identify certain conditions and limits of each of the options.

2 From Intelligence to Sentience

The Computer as an information and communication machine is at the heart of the research on AI and of the comparison of human and machine for the operations of reasoning and cognition. If one takes the broader paradigm of the Humaniter, it appears that the AS can be viewed as the other side of the transfer of functions and abilities from the human to the machine. However, some questions need to be raised concerning AI and AS, in particular on the difference between the strong version and the weak version and on the question of consciousness.

2.1 From Intelligence to Sentience, Through Consciousness

There are actually various reasons for moving from a research program on intelligence to a research program on sentience. From an operational point of view, it is justified by the need to have machines which, when interacting with humans, behave in a manner that shows their ability to express sensations, emotions and even consciousness. It is a way of making machines more “human”, especially if they are called upon to play an increasingly important role in our daily lives. From a reflexive point of view, this shift from intelligence to sentience is justified by the need to better understand the abilities of machines to process information, so that they come as close as possible to the humans processing. This makes it possible to measure the possible difference between the ability of the machine and that of the human and to learn therefore about the singularity of the human, which in turn questions the singularity of the machine [6].

The fact remains that this shift from AI to AS can be interpreted according to the Humaniter’s paradigm as an attempt to go through with a substitution of human functions and abilities by the machine. From this point of view, it takes sides, in the debate between strong AI and weak AI, if one takes up the cleavage proposed by Searle, in favour of a strong program, even if it can be shown that only the weak one can be achieved. Strong AI refers to the Artificial Intelligence program which envisions the intelligent machine as endowed with ability not only of reasoning, but also of consciousness. In comparison, weak AI bounds to an intelligent machine deprived from

² This is an argument which partly links up with the Frankenstein’s axiom: “An exact physical replica, although it has been produced, of a sentient being would itself be a sentient being”.

this full scope of abilities and that can then function as a device intended to better understand human intelligence, the only one provided with consciousness. One could say that the parallel works for artificial sentience, if we consider that strong AS designates the program of Artificial Sentience in which the sentient machine is endowed with an ability of sensation and emotion and to some extent, consciousness. In comparison, weak AS relates to a sentient machine deprived from this full scope of abilities and that can function as a device intended to better understand human sentience and the role of consciousness.

Now, one can come back to the two principles, the Principle of Artificial Intelligence and the Principle of Artificial Sentience, in order to clarify their meaning, but also, to show their conditions and limits.

2.2 Principle of Artificial Intelligence (AI)

The principle of AI can be formulated as follows:

- (1) *Principle of Artificial Intelligence (PAI):* A machine can think like a human.

From there, you can specify two versions, one strong and the other weak:

- (1.a) *Principle of strong AI:* A machine can think *exactly* like a human, both from the point of view of reasoning and of consciousness.
 (1.b) *Principle of weak AI:* A machine can think *approximately* like a human, both from the point of view of reasoning and of consciousness.

When it is said about a machine that it can think *like* a human, it does not mean the same thing as a machine that thinks *as* a human. Because if it was so, then it would mean that the machine *is* a human who, among other things, is able to say “I” - as in the expression “I think”. In this respect, it is illusory to think that when a machine says “I, Robot”, it has comparatively the same kind of meaning as in the sentence of a human who says “I, Human”... In addition, it can be noted that the difference between identity (“exactly”) and similarity (“approximately”) says nothing about the power of reasoning given to the machine by calculation, if only we compare it to that of the human. Thus, a machine can follow a reasoning which approximates that of the human, whereas it is of a power which is clearly higher than that of the human (as in the chess game with Deep Blue, or the Go game with Alpha Go).

Now, it goes without saying that the “thought” has a fairly diverse set of meanings, as suggested by Descartes who uses this term to designate reason, doubt and imagination as well as knowledge, will and sentiment [7]³. In another interpretation, a philosopher such as Wittgenstein would say that the concept “to think” requires to consider the multiple uses that one makes of it in everyday life, the relationship with language and with the behaviour of the body [8]: “194. We must never forget that ‘thinking’ is a word of everyday language ... We should not expect him to have a unified use, quite the contrary. 220. ‘Thinking’ is a largely ramified concept. A concept

³ “Je suis une chose qui pense, c’est-à-dire qui doute, qui affirme, qui nie, qui connaît peu de choses, qui en ignore beaucoup, qui veut, qui ne veut pas, qui imagine aussi, et qui sent”, III.

that contains many manifestations of life. The *phenomena* of thought are dispersed. 223. The expression on the face of someone who thinks, and that of an idiot. The brow that creases in reflection, in attention.” This last point raises the whole question of the bodily expression of the mental life, if assumed alike Wittgenstein that it cannot be only “internal”. It is, so to speak, its “external” face that draws attention, or more exactly, the coupling of thought with the words of language as well as the behaviour of the body.

A rational approach to the thought suggests that it merges with the different types of reasoning that have been identified in logic. It is customary in this field to differentiate between the reasoning which proceeds by deduction, by induction and by abduction: in other words, going from the general to the particular (syllogism) or from the particular to the general, with in the abduction a return to the particular. However, this implies that certain rules of reasoning stemming from logic are *norms* provided with a binding force for an ordinary reasoning which deviates quite often from them. In contrast, a broader approach to thinking includes, in addition to reasoning, a variety of forms of intelligence, with sometimes an emphasis put on creation and invention, that is, the production of novelty. This is how in a psychological version we distinguish a variety of forms of intelligence, grouped under the expression “multiple intelligence” [9]: the logical-mathematical form, which corresponds to the rational approach; but also, the linguistic, the spatial, the bodily (implementation in dance or sport), the intra-personal (ability to understand oneself) and the inter-personal (ability to understand others), the musical; to these forms can be added the naturalist (ability to recognize animals, plants) and the existential.

However, the term intelligence is taken with caution in psychology, and that of performance is often preferred to it for various fields: execution, adaptation, control, speed of processing, abilities of working memory, of reasoning. Thus, we can see that, even in the research program on artificial intelligence, there is still a long way to go to replace the human with machine. One wonders if the same kind of difficulties occurs in this part of the Humaniter paradigm which makes artificial sentience the other horizon of the research program.

2.3 Principle of Artificial Sentience (AS)

The principle of Artificial Sentience can be formulated as follows:

- (2) *Principle of Artificial Sentience* (PAS): A machine can feel like a human.

From there, you can specify two versions, one strong, and the other weak:

- (2.a) *Principle of strong AS*: A machine can feel *exactly* like a human, from the point of view of sensation and emotion and of consciousness.
 (2.b) *Principle of weak AS*: A machine can feel *approximately* like a human, from the point of view of sensation and emotion and of consciousness.

Again, when it is said that the machine can feel like a human, this view does not imply that it feels *as* a human, as that would imply that it is a human. The significant point is that a human able to say that he or she is a sentient being is a being who has the ability to speak, in particular to say “I”, which is linked to an ability to think.

Sentience belongs to the vast realm of “experience” which covers a wide range of meanings, from sensation to sentiment. The nuance as far as the thought is concerned is that experience heard in the sense of phenomenal experience is said in first person, at the risk of fall into Descartes’ solipsism. This is suggested by the experiential approach which focuses on phenomenal experience, supposedly distinct from one person to another, since no one can have access to the mental states of others. So, in the subjective approach to experience, I have no guarantee that the object as I see it (for example, a squirrel) is seen in the same way by someone else. However, this is a point disputed by Wittgenstein who notes that, by making my experience too much a strictly private matter, it can no longer be identified with that of the others. Yet even the most personal experiences, such as those of sensation and emotion, suppose a public language and disqualify the idea of a private one used by myself alone in order to express my personal sensation or emotion [8]: “275. Look at the blue of the sky and say to yourself: “How blue the sky is!” If you do it spontaneously - without philosophical intentions - the idea will never occur to you that this impression of colour belongs only to you. And you will have no hesitation in making this exclamation to someone else... 303. Just try to doubt - in a real case - the fear or pain of others!”. What the philosopher suggests is such behaviour of doubt, in a real situation, would be for those who suffer perfectly inappropriate and could arouse their greatest anger. This does not detract from the relativity of perception, in the “seeing like” mode, when, in a figure such as the Duck-Rabbit (Jastrow), some see a duck, where others see a rabbit [10].

The field of sentience covers a set of aspects of experience which incite to make a difference between sensation, perception, emotion, sentiment, and in a broad sense, consciousness [11]. Literally, sentience (from the Latin *sentientem*, in the nominative *sentiens*) describes the ability to feel, whereas the gerundive of the verb *sentire* refers to the concept of consciousness, taken in the sense of being aware of what surrounds us. It is common to express the difference in experience between sensation and perception by taking into account what the judgment adds, however basic it may be. This is about how I see an object, for example, a red cube: I have the sensation of the colour red, but I have the perception of a red body as a unit that occupies a certain volume in space. Sensation is distinguished from perception in the sense that my judgment, even at a primitive level, tells me that this object is not just a flat surface. It would then be necessary to add a concept, in this case the concept of “cube”, in order to differentiate between a cube and a square, or between a cube and a sphere. Emotion (from the Latin *emovere*, to set in motion) is another dimension of sentience that is not easy to define, perhaps because it covers actually a multitude of aspects. Emotions do have one thing in common, however: they are the subject’s overt responses to meaningful events and can trigger distinctive bodily changes and behaviours [12].

2.4 The Question of Consciousness

Consciousness is one of the most difficult problems to solve in any study of both intelligence and sentience, all the more if they are artificial. Numerous studies have attempted to define what consciousness is and to determine whether it is possible that a machine could be endowed with this ability characteristic of humans or certain non-humans (animals). As Chalmers puts it, it is customary to differentiate between “easy

problems” and “difficult problems” of consciousness. Easy problems are those which relate to the explanation of cognitive and behavioral functions: for instance, the ability to discriminate, categorize, and react to environmental stimuli; the integration of information by a cognitive system; the reportability of mental states; the ability of a system to access its own internal states; the focus of attention; the deliberate control of behavior; the difference between awakesness and sleep. The difficult problems are those relating to the explanation of conscious experience, whether you call it phenomenal or qualitative if what is insisted on is the *qualia*.

The difficulty of conscious experience is well explained by Chalmers [13]: “the easy problems are easy precisely because they concern the explanation of cognitive *abilities* and *functions*. When it comes to conscious experience, this sort of explanation fails. What makes the hard problem hard and almost unique is that it goes beyond problems about the performance of functions. To see this, note that even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience - perceptual discrimination, categorization, internal access, verbal report - there may still remain a further unanswered question: Why is the performance of these functions accompanied by experience? A simple explanation of the functions leaves this question open... This is not to say that experience has no function. Perhaps it will turn out to play an important cognitive role. But for any role it might play, there will be more to the explanation of experience than a simple explanation of the function ... The same criticism applies to any purely physical account of consciousness. For any physical process we specify there will be an unanswered question: Why should this process give rise to experience?”.

The question of consciousness, and more specifically, the possibility of an artificial consciousness is as much on the side of artificial intelligence as that of the artificial sentience. But when we take a closer look at the variety of abilities that are grouped under the term “consciousness”, it is not surprising that it is found on both sides, as this list shows [14]: attention, creativity, dynamism, emotion, imagination, intelligence, intentionality, language, quality (*qualia*), perception, self, volition. It follows that the dividing line between intelligence and sentience is not so obvious, especially when you consider the role that consciousness plays in distinguishing between human and machine. This question of consciousness refers more fundamentally to the kind of tests that have been designed to characterize a machine as capable of being endowed with intelligence and sentience.

3 Tests of Humanity

The Humaniter is a machine that, if it is meant to take the place of humans for all or a part of their functions or abilities, must be able to pass a test of humanity. The test of the machine intelligence is now quite well established, even if it continues to generate a lot of debate, but it remains to be clarified what is or what can be a test for the machine sentience. We favour here an approach that insists on the behaviour and the interpretation made of it by a judge, while another approach would rather emphasize the

architecture of a system [14]⁴. In addition, we favour an approach which, in the wake of the original Turing test, develops a test modality that is not limited to a computer, but results in the action of a machine (a robot), like in the Total Turing Test [15]. It can be suggested that, for intelligence as for sentience, a test of humanity calls upon the criterion of the action of an agent, but without thereby reducing to it the whole of the “human” behaviour of a machine.

3.1 Tests for the Intelligence of a Machine

The Turing Test for the machine intelligence was originally an imitation game which consisted in developing a machine that cannot be distinguished from a human [16]. Turing suggests that a judge J exchanges typed messages with a human being H on the one hand, and a machine M on the other, messages that can relate to all kinds of subjects. Judge J does not know which of his two interlocutors (whom he knows under the names A and B) is the machine M and the human H. In the rules of the game, it is stipulated that, after a series of exchanges, the Judge must guess who between the two of them is respectively the human being and the machine. Correct identification consists for Judge in producing a misidentification rate of 50% identical to what a random answer would give. Turing believes that if a machine one day does not allow correct identification by a human, it is an intelligent machine or, if you put it in other words, a “thinking machine”.

The Turing test procedure has given rise to many interesting discussions and is the source of a series of concrete IT achievements, despite the difficulty of a program to pass the Turing test. We can put ourselves in the place of a judge J who dialogues through a computer terminal with the two interlocutors A and B and, from there, identify some possible options. A first option, to recognize the machine, is to ask a question such as “what is the value of 327 at the 4 th power?” If A answers 11433811041 after a second and B refuses to answer, or waits a few minutes to propose a result, there is no doubt that A is the machine and B the human. However, the specialists who design the programs for taking the Turing test are not stupid and they foresee this coarse trick. Their program is able to lead the calculation of 327 to the 4 th power without difficulty in no time, but it will refuse to answer or ask ten minutes before providing a result, or even offer an incorrect answer. Everything that a human does not fully succeed and that a computer succeeds without difficulty is treated in the same way, by a computer that “pretends” to fail, as if failure was the hallmark of the human, not the machine.

In order to identify the human, the judge’s method must be based on tasks that humans can easily handle and that computers stumble upon. One of those aspects is humour - like for example, in this short joke: “Have you heard of the new restaurant called Karma? It has no menu, you get what you deserve”: the judge tells the funny story and asks A and B to explain where and why you should laugh. Another aspect is the news, a set of points that everyone is informed about and that can be used as a basis

⁴ Elamrani et Yampolskiy suggest that all the tests implies one human who interprets the interaction and favors an approach that is either oriented to the architecture, or to the behaviour.

for an identification attempt. One can also engage a computer in a conversation that deals with all kinds of subjects (for example, science, history, art, entertainment, music, etc.), which represents a high level of difficulty. Turing's idea seems correct in principle, because if you can fool a judge with a program, then your computer is operating in a mode that resembles intelligence. The fact remains that, even with the best programs, the result is still often only a simulated conversation that does not deceive a human judge for a long time.

3.2 Limits of the Intelligence Tests

Some philosophers, among them Searle, believe that even if you manage to complete a program that passes the Turing test, it does not prove that you have put intelligence into the computer [17]. This is his famous thought experiment of the Chinese Room in which an individual who does not speak a word of Chinese can nevertheless use a set of language rules and thus apply them to have a conversation in Chinese. Searle states that the non-Chinese speaker can provide correct answers from a syntactic point of view, when he or she understands nothing about the meaning of the symbols he or she manipulates from a semantic point of view. This is meant to show that computer programs are syntactic, they are only tools to manipulate symbols, while human thoughts have a semantic content, a meaning which is attached to words. The meaning of words in language is not reduced to syntax, because it comes from biophysical properties of our neurons that a program cannot possess. Hence this basic idea that programs cannot think: literally speaking, a computer does not speak Chinese, for it does not understand this language as humans can do... The machines will therefore never be intelligent, even if they pass the Turing Test, because the latter is not a sufficient condition for the intelligence of a machine⁵.

Another criticism of the Turing test was made by French, who argued that passing the test is not a necessary condition of intelligence [18]. He imagines the story of a people who would know only one species of bird, the seagull, and would face the problem of making a flying machine. In order to determine if they are successful in this project, these people would use the seagull test in which a machine is flying if it cannot be distinguished from a seagull whose behaviour is observed with help of a radar. The radar limits the precision of the request for imitation as the dialogue, by means of typed exchanges, limits the precision of the request for imitation in the implementation of the Turing test. In this test, planes, helicopters, hot air balloons and even other birds will not pass the seagull test and will therefore not be considered able to fly. Thus, the Turing test is perhaps a sufficient condition of intelligence, but of human intelligence, and it is linked to the language used for dialogues, which prevents it from being considered universal. It may well also be that many of our behaviours are highly dependent on the particular way in which our brain processes information at the deepest level, the one which of the subcognitive processes [18].

⁵ Now, Searle might admit that if one day the machines succeed in faithfully imitating humans, we would no doubt have to change our position, against a vision that by principle seems to reserve intelligence to humans.

3.3 Tests for the Sentience of a Machine

One can also imagine a test of humanity for the sentience of a machine in order to assess its ability to “feel” like a human in the interaction it can have with him or her. It is not certain that this kind of test should be based on that of Turing, but it can be fruitful to explore this track, as already envisaged by Campbell who speaks of a game of human imitation (*Imitation Man*) [19]. Attention is focused on the conversation between a judge J, a human H and a machine M, but the main difference in the sentience test is that the content of information and the mode of communication are not the only things that counts. It is also information and communication as it shows a set of aspects of a partner’s behaviour that testify the presence of a sentience in the language itself. It is therefore a linguistic evidence which grounds the Judge J’s judgment in the absence of access by the judge to physical data which could express the psychic state of A and B.

The test of humanity of a machine is moved towards the expression of sentience in verbal language, towards what reveals the sensation, the emotion, the sentiment or even the consciousness of A or B. This expression of sentience in language supposes not only an information content, but also a mode of communication. Now, if we stick to a classic Turing test, this implies that it can only be a verbal communication, to the exclusion of any non-verbal one which would pass through expressions of the body (gestures, facial expressions, etc.). It is therefore in the realm of the language used by the speaker that the sentience can reside, as when A or B says: “I have a toothache, it’s been a long time since it happened to me. It’s a pain that takes me to the bottom molar, on the right side, and goes down into the gums.” Or when A or B answers, to a question asked about its isolation and its awareness of it: “Yes, I feel a little lonely in this room, and I must say that, usually, I am rather someone who likes the presence of others.” If a machine M is able to express itself in this way, a judge J can conclude that it is a human H who communicates this message to him or her.

Now, one could also conceive that the sentience test takes into account, according to the access clause to physical information, and consequently, to psychic information, the non-verbal communication. The information and the communication of a machine M should then not only be taken for that of a human H by a judge J, but it should be also by virtue of a certain behaviour of the body. This supposes that, in its dynamic aspect, that of non-verbal communication, if however it is coherently coupled to verbal communication, a machine is identical to a human. It would be so if a machine M, of strictly human aspect, begins to get angry, with the blood rising to the cheeks, the sound of the voice raising, the arms waggling, while saying to children: “Listen, I have told you several times not to touch this Ming vase. Are you an idiot?” Now, we are entitled to wonder if verbal and non-verbal communication and the coherent coupling between the two do not requires something like a “grammar” of expressions. It is certainly this common grammar that judge J should share with machine M and human H in order to produce a correct interpretation of their behaviour.

3.4 Limits of the Sentience Tests

Searle is also famous for developing the pragmatics, the study of speech acts which supposes that the meaning of a statement depends on the speaker intentions and on the speech situation [20]. The speech act can succeed or fail depending on the situation, like when I say “I baptise this ship Queen Mary” and the conditions for success are or are not satisfied. For instance, I am achieving this baptism with a bottle of champagne, but the bottle does not break on the hull, which is interpreted as a failure of the speech act, since the ship is not properly baptised. Behind the intention of a speech act stands the intentionality, the relation of consciousness to an object, which can be extended by considering the background of this relation, the tendency or the disposition of belief and desire [21]. This notion of intentionality can be useful beyond intelligence for the sentience test as a necessary condition in order to give a meaning to an action. Now, one could argue that it is not a sufficient condition, in the sense that, if a machine passed the sentience test, it would not prove that it has intentionality. It would just prove that it behaves in action and communication as if it had an intentionality, but without experiencing sensation or emotion, neither desire and belief. In this respect, no more than syntax, the pragmatics of action would allow access to the meaning of words, if these depend on intentionality and are empty of any intention, desire or belief.

A more radical objection puts forward the idea of the “philosophical zombie”, as in Kirk’s thought experiment, according to which it is possible that a being has all the characteristics of a human, but is found lacking in sentience [22]. One can imagine as a logical possibility that an organism or even a machine cannot be distinguished from a normal human being, in its bodily and behavioural aspects, but that he feels nothing⁶. Thus, the being in question could for example be struck by an object, but not feel any pain, while giving the appearance in his reaction that he did feel pain. The argument is quite close in the spirit to that of the doctrine of physicalism and amounts to saying that consciousness is nothing more than a physical phenomenon. However, some researchers in artificial intelligence, like Minsky, try to show that the argument is circular: the possibility of something physically identical to a human, but deprived of subjective experience, assumes that the physical characters of humans are not what produce experience... which is contradictory to physicalism. It follows as far as artificial sentience is concerned that the relation between the physical aspects of the bodily behaviour is also narrowly coupled with the psychic activity of a mind endowed with a subjective experience - the “hard problem” of consciousness for Chalmers [13].

4 Human-Machine Interaction

A significant number of humanity tests give primacy to the interaction between the human and the machine, on the narrow mode if the interaction is bounded to conversation, on a broader mode if it includes the behaviour in action. This is all the

⁶ As stated by Kirk, «it is at least logically possible for there to be an organism indistinguishable from a normal human being in all anatomical, behavioral and other respects, yet insentient», p. 43.

difference between a partial Turing test (*Turing 1*) and a total Turing test (*Turing 2*), but one can wonder if the first kind, with its restrictive clause of conversation, is still relevant.

If it is considered obsolete, the move to the second kind of test, which widens it to action and behaviour, results in the production of machines that in their physical aspect can be confused with humans. This clause also assumes that the behaviour of these machines is like a door open on a psychic life which must have a certain form, or structure, and show some coherence. But perhaps this is too demanding a clause, for it suggests that humans themselves would be fully coherent beings, while simple ordinary experience proves the opposite. It is therefore towards the idea of a *personality* of machines, endowed with a certain character, coupled with a certain temper, that it would be wise to direct the reflection on the action of artificial devices. But the personality of a machine, if such a thing exists, cannot be kept aside from the codes of a society or a culture which owns a certain “grammar” of the person and for some of them only refers to the notion of *Ego*.

All in all, the Humaniter as an idea is a machine which not only *thinks* like a human, *feels* like a human, but also can coordinate these two kinds of abilities and link them to another one, action, so that the machine also *acts* like a human. One could thus suggest that the idea of a Humaniter, as a total machine, requires not only the principle of AI, the principle of AS, but also the principle of AA: Artificial Action. This AA Principle could be formulated as follows: *A machine can act like a human*. Thus, a Humaniter is a machine which articulates the functions and abilities of the Reasoner, the Cogniter, the Voliter, the Experiencer and, finally, the Actor. This is a point to keep in mind in the project of making a both intelligent and sentient machine, namely a machine that properly speaking is not and cannot be a Humaniter, but that nevertheless can be a machine with a human face.

References

1. Gâcogne, L.: Intelligence Artificielle. Ellipses, Paris (2015)
2. Julia, L.: L'intelligence artificielle n'existe pas. 1st Edn., Edi8, Paris, France (2019)
3. Matson, W.I.: Sentience. University of California Press, Berkeley, Los Angeles (1982)
4. Husain, A.: The Sentient Machine. Scribner, New York (2017)
5. Breton, P.: A l'image de l'homme Du Golem aux créatures virtuelles. Seuil, Paris (1995)
6. Kurzweil, R.: The Singularity is Near: How Humans Transcend Biology. Viking Penguin Group, New York (2005)
7. Descartes, R.: Méditations Métaphysiques. Vrin, Paris (2019)
8. Wittgenstein, L.: Remarks on the Philosophy of Psychology, 3rd edn. Wiley Blackwell, Malden, Oxford (1998)
9. Gardner, H.: Frames of Mind: The Theory of Multiple Intelligence. Basic Books, New York (1983)
10. Wittgenstein, L.: Philosophical Investigations, 4th edn. Wiley Blackwell, Malden, Oxford (2009)
11. Perkins, M.: Sensing the World. Hackett Publishing, Indianapolis (1983)
12. Scarantino, A.: Emotion. Stanford Encyclopedia of Philosophy (2018)
13. Chalmers, D.J.: The Character of Consciousness. Oxford University Press, Oxford (2010)

14. Elamrani, A., Yampolskiy, R.: Reviewing tests for machine consciousness. *J. Conscious. Stud.* **26**(5–6), 35–64 (2019)
15. Harnard, S.: Can machine be conscious? How? *J. Conscious. Stud.* **10**(4–5), 67–75 (2003)
16. Turing, A.: Computing machinery and intelligence. *Mind* **49**, 433–460 (1950)
17. Searle, J.: Minds, brains and programs. *Behav. Brain Sci.* **3**(3), 417–457 (1980)
18. French, R.M.: Subcognition and the limits of the turing test. *Mind* **99**, 53–65 (1990)
19. Campbell, K.: *Body and Mind*. University of Notre-Dame Press, Notre Dame (1984)
20. Searle, J.: *Speech Acts*. Cambridge University Press, Cambridge (1969)
21. Searle, J.: *Intentionality*. Cambridge University Press, Cambridge (1988)
22. Kirk, R.: Sentience and behaviour. *Mind* **83**(329), 43–60 (1974)