European Journal of Science and Theology, June 2013, Vol.9, No.3, 45-61

ABOUT INTELLIGENT MACHINES, 'SINGULARITY' AND GOD

Ionuț Isac[*]

Institute of History 'George Barițiu', Str. M. Kogalniceanu 8, Cluj-Napoca 400084, Romania

(Received 16 April 2012)

Abstract

This article comments on the idea of 'singularity', understood as the result of a very fast and sudden technological process, supposedly giving birth in the future to an 'almighty' generation of machines or maybe to a single one of that kind. Many fears and desires, exercises of imagination and speculations have been expressed on this subject-matter, in our opinion all of them having in common the favourite representations of the analogy between the supposedly 'almighty' self-aware machine and a new kind of a technological 'God(dess)'. Our point is that, firstly, the representation of the intelligent and extremely powerful 'ultimate' machine is nothing else than the expression of a very ancient and secret human desire of becoming the creator of its God (i.e. in the meaning of a 'reversed' theology); secondly, that those who anticipate a possible end of humanity as a consequence of 'singularity' conceive it secretly as a possible evil technological deity, whose infinite power will keep the people at its discretion. In both cases, we are dealing with a mixture of representations (scientific, religious, philosophical, psychological, etc.), speculated in a specific way, revealing sometimes a distorted understanding of the Christian religion. In this respect, we are trying to identify some possible connections between philosophical assumptions about technology and meanings of the concept of God, especially that of *deusabsconditus*. As relevant for the paper's subject and thus, involved in the whole matter, some main arguments for and against technological super-intelligence ('singularity') are also discussed, emphasizing the importance of lucidity, criticality and openness.

Keywords: intelligent machines, technology, singularity, religion, God

1. Introduction

The technological revolution in Informatics, Robotics and Computer sciences is merely at its dawn. However, humans are tempted to imagine how their world will look years, decades and centuries later. The quick and unpredictable replication of machines inside a world of their own makes a ground for many speculations of all kind, among which there are many ethical entailments (e.g. whether such an evolution is good, bad or maybe neutral to humans). But anticipating and representing the future is not an easy task for anyone. Thus, by speculating on some uncertainties of the present-date

[*] E-mail: isac.ionut@cluj.astral.ro, fax: +40 264 592783

knowledge, one might fall into the temptation of pessimistic views, dark forecasts and dystopia. One of the most renown and famous position is that of human fear of being controlled and subjected by evolved robots, intelligent machines and various technical devices. After many discussions and debates, involving many suppositions and uncertainties, the subject is still unclear at this present time, but its scenario is admitted as theoretically possible in the future [1].

Subject of speculations *par excellence*, the entity named 'singularity' occupies the front seat of these mythological ventures. Yet, neither mankind's nor machine's evolution is sufficiently known and clarified in all of its essentials, so until the time will prove if we were right or wrong, the imagination can feel free to improvise Hollywoodian scenarios with SF emergent evil machines, heroes (humans and/or robots) fighting against them with great success or no success at all, anticipations about how to save the world after its potential or real destruction, death and resurrection of humans (eventually helped by 'good' robots/ machines) and so on.

The most interesting assumption which attracts our attention in the present paper is that of the 'singularity' conceived as a self-aware machine (i.e. similar to a 'God'), literally capable of everything, be it for better or for worse. Now, when somebody addresses the question if such a machine would need or not humans' neighbourhood or would it be ready to take the decision to get rid of them once and for all – solely based on its absolute power in the Universe – one has to analyze thoroughly the strength of connection between humans and these very intelligent machines, especially the fact that "the use of and engagement with technical artefacts involves always an enrolment of objects (as well as subjects) into an array of different kind of networks which enhance our physical and mental capabilities up to some yet unknown magnitude" [2].Thus, in our opinion, the emergence of a very intelligent, very powerful and self-aware machine is an event that has to be connected to the existence of some evolved (above the present-day level) human designers, owners of an appropriate know-how able to give them access to communication and cooperation with the 'singularity'.

2. Discussion

2.1. 'Singularity' and hidden meanings of the concept of God

The hypothesis or presupposition behind the emergence of 'singularity' is the continuous improvement of intelligent machines. Thus, one might think that, if a human-built machine could be brought to bear greater problem-solving and inventive skills than humans, then it may be able to design a yet more capable machine. If built, this supposedly 'more-capable-machine' could then design a machine of even greater capability and so on. This iterative process could be accelerated, leading eventually to a "recursive self-improvement" or to an "intelligence explosion" (to use the expression of I.J. Good) [3]. What amazes

here is the strong confidence of those who share this point of view, fundamentally based on inductive reasoning, namely that if such a process has begun, at a certain moment it will necessarily reach its peak – i.e. to arise to the point of 'explosion'. How can one be, in fact, so sure that this process will repeat and enhance itself each and every time on machines, generation after generation? If taken seriously, is there a scientific, a technological or a pure logical reason to believe so?

A preliminary argument of doubt to be launched against this kind of forecast is that, once such intelligent device would reach a very high degree of intelligence, complexity and speed of its actions, there is no certainty that such a machine would be capable or willing to design a different machine, 'better' than itself. To sustain such an idea would be to apply pure induction, inspired by the assumption that the 'classy' generations of intelligent machines must aspire to 'perfection' as their supreme goal. What would be, then, the meaning of such a projection? Would it be, eventually, our own projection about perfection – an ideal that inevitably mobilizes us all on its way?

In human terms, generally speaking about perfection, we understand it as 'a state of completeness and flawlessness', a certain kind of theoretical science, fine art, philosophical system, as well as a moral or religious spirituality at the highest possible (at least humanly imaginable) level. Sometimes, this 'perfection' is intimately associated with the personality of its creator (i.e. a human being exceptionally gifted), whose name it occasionally bears. More specifically, from the religious point of view it has been asserted that the word 'perfection' came to modern vocabulary "from the Latin *perfectio* meaning 'completeness' or 'completion'. The concept has been applied primarily to God, and then in a derivative sense to man … Jesus Christ, regarding God the Father as perfect, laid down the eschatological requirement of men that they be likewise perfect. Although the injunction is usually regarded as impossible to be fulfilled in this life, it gave rise to a goal of Perfectionism which appears here and there throughout the history of Christian thought and institutions: in Origen, in Roman Catholic monasticism and mysticism, and in numerous Protestant perfectionist sects." [4]

Those who plead in favour of mankind's progress do like to take a genius' works a very good example for their discourse, showing that humanity has thus made – sometimes against its will – a big step forward. And this is, in principle, true: humankind improves, creates, invents and changes what has been considered immovable, making ideally its path to the Word of the Lord, which invites "the man that belongs to the technological and informatics society to the communion with Him in the ambient of the ecclesial community where He is always present as devoted love, as Saviour gift" [5]. But, however, there is no place and reason here for a flat induction. Nobody can be sure that one of us will be saved, because even if the humans know what the risk of making sins is, they continue to be sinners and, thus, to be wrong in the eyes of their Father. The way of improvement of one's life is not at all easy and, once we know, in principle, how machines do work nowadays, as well as how problematic becomes

everything when humans use to cope with 'better-and-better', it is much harder to yield the road to such a simplistic overview (i.e. that machines just could or would aspire to 'perfection', and not doing otherwise). Would the machines be better than humans and succeed where they failed? And why would be one compelled to think so? What we call 'perfection' is a task already hard enough for *homo sapiens*; who knows what is the case of machines?

Now, let us think the problem differently. Who could guarantee us that, for instance, maybe because of some inherent limits of our own, we are still not aware of some reasons of self-protecting on behalf of which the aforementioned 'utmost-evolved-machine' would rather be tempted to stop itself somewhere in the process of 'recursive self-improvement'? It is logically consistent to say that, consequently, a very intelligent machine may decide to initiate a multiplication 'in itself and by itself', mainly at the same level of complexity already acquired, anticipating its evolution in 'small steps', according to the area of 'problem-solving' within a certain paradigm. Thus, it may not be the case of 'quickly-breaking-paradigms' machines just because we imagine that they have to do so. For a very intelligent machine, there could be no need to enhance indefinitely its capabilities, once supposed it could have become far more intelligent than its designers; or, at least, to do so every time we believe it is needed, because that would be a reductionist–'mechanicist' way of understanding its technological being. On the contrary, once having reached an unusually (non-human) elevated intelligence, those machines might also decide to further create and develop not 'superior' but, on the contrary, rather 'inferior' machines (however much more intelligent than humans). They could decide to do so on the purpose of reserving to themselves an unassailable pre-eminence in the world for an unknown period of time (maybe as long as possible). It may occur that those machines would not be willing to expose their outstanding place inside the whole of the existence; or – to keep the same way of reasoning –, if once having decided to build a machine 'more capable' than them, this could mean exactly as to design their future disappearance. In other words, the 'elevated' machines can decide to put a 'barrier' between their artificial 'heaven' and the human 'prosaic' world, in terms of less evolved artefacts, as an environment for usual and safe interactions with humans.

Accordingly, nothing can prevent us from imagining that those 'classy' machines would prefer to communicate with their inferior 'mates' as well as with 'accompanying' humans in terms of 'lower' knowledge, keeping the 'supreme' truths and axioms just for their own benefit, with no direct implications toward their alleged interest on possible extinction of *human sapiens*. And, still, that does not make those machines "gods", nor does it imply necessarily the extinction of the mankind. If somebody were to push the interpretation to the limit, it would appear that such a machine resembles more *a deus otiosus* (the idle god) or a*deusabsconditus* (the hidden god), reminding us of Thomas Aquinas (1225-1274). Both terms "refer to a deity whose existence is not readily knowable by human's solely through contemplation or through the examination of divine actions. The concept of *deus otiosus* often suggests a god

who has grown weary from involvement in this world and who has been replaced by younger, more active gods, whereas *deusabsconditus* suggests a god who has consciously left this world to hide elsewhere." [http://en.wikipedia.org/wiki/Deus_otiosus] It seems to us that both meanings fit the subject, as a very intelligent machine capable of resolving the most difficult problems and, thus, able to change our world for better in many ways, would rather become willing to get away from mankind's countless nonsense and create for itself such a world, allowing it a remote place, inaccessible and comfortable. Nonetheless, that machine could decide to leave us some 'younger gods' – those aforementioned would-be 'inferior' machines, whose potential of intelligence could prove, however, crucial for the projects of humans in the future.

Speaking about the evolution of machines or technical intelligent devices and looking for similarities between human beings and machines means continuing the discussion at the ontological level. The evolution of technological systems seen as an extra- or a para-'natural' process indicates that they "appear and evolve according to a genuine descent, as subject of a selection process as well as of different conditions of reproduction, whose complex criteria still remain known only to a small extent" [2].

Usually, as described by Kaplan, the technical systems' process of generation has 3 major stages:

1. When a new technical ensemble emerges, it is made of existing or developing technical elements, whose degree of integration is weak. This way, a juxtaposition is created, composed by different technical elements that come from independent and potentially very old technical descents.
2. During a second stage of growth and maturation, the technical object continues to transform itself through juxtaposition and integration. At a certain moment the technology comes to a *turning point*, when it begins a massive *self-enhancement* of that item, even if the item or the system continues to spread. This acceleration leads, on one hand, to a stronger integration of the technical item, and on the other hand, to an emergence of new systems, either based on the specified technical object itself or on some of its components.
3. In a third period of growth, the technical descents either succeed to stabilize themselves or, on the contrary, they may regress. This is the case of the technical item (system) reaching its maximum level of functioning (i.e. its limits) and for such reason it is – sooner or later – replaced with another one, coming from a different technological descent, better adapted to a specific environment. [2, 6]

When one speaks about the future of the very intelligent machines, (the technological 'singularity' included), it is worth to mention that, according to what is already known, it is possible that parts of technical items or objects be extracted from genuine systems and joined together in various unpredictable and unexpected combinations. Therefore, it is very difficult (if not impossible) to predict the design of such an intelligent machine, since no one today can

precisely anticipate how these combinations will evolve in decades and centuries, not to mention the new devices, engineering innovations and technologies that the society of the future is supposed to undergo.

We consider that, in this respect, one must rethink the metaphysics of the Romanian philosopher Lucian Blaga (1895-1961), beginning with its central, controversial and very much discussed hypothesis on the nature of existence: i.e. the concept of 'The Great Anonymous' accompanied by its 'transcendent censorship'. 'The Great Anonymous' refers to an entity as the 'core' of transcendence, which represents the 'central existential mystery', defending for good 'the derived mysteries' from human knowledge, whose activity is based on a fundamental onto-logical contradiction; it means the self-imposed, absolute and eternal mystery. Blaga wrote that, being situated on an uncertain ground (both mythical and philosophical), the discourse about 'The Great Anonymous' is an act full of dangers and surprises. Therefore, it is very hard, if not impossible, to give a positively-adequate representation of this strange almighty entity; furthermore, its main function is to prevent humans from accessing 'the ultimate mysteries', providing a specific barrier – the so-called 'transcendent censorship' which is the metaphysical axis of knowledge.

In order to make the matter more suitable for the present discussion, we can assert that, using the language of informatics, 'The Great Anonymous' designed the 'transcendent censorship' barrier as a 'safety net' or a 'firewall' which separates the human being as subject of knowledge and the mysteries of the world as objects of knowledge. All efforts deployed by humans in order to reveal mysteries and to obtain a 'fully adequate knowledge' are thereby doomed. In fact, Blaga states that: "There are no adequate revelations. For this conception, 'revelation' is a purely theoretical concept. In fact, no existential mystery crosses the threshold of knowledge remaining what it is. The threshold of knowledge is enchanted and transforms every guest who crosses it. In reality, there are only dissimulative revelations. An existential mystery, which discovers or reveals itself as such to the individuated knowledge, is *dissimulated* by the very structure that has been shaped to individuated knowledge according to the intentions inherent to transcendent censorship […] The showing or revealing of an existential mystery, when it happens, is always a *censored revelation*, censored by the very structure of the cognitive machine meant to receive the existential mystery." [7] But no one else than the same 'The Great Anonymous' has already spread out in the world the 'incentive' elements able to lure the mankind toward mysteries: 'the divine differentials'. The most important results of these reciprocal annihilating processes are *the culture(s)* and *the history*. Thus, the mysteries are never 'revealed' and the fundamental desire of mankind is never fulfilled; they are only 'dissimulated' by the transcendent censorship. The human knowledge can always be only knowledge of 'this' or 'that' particular mystery, but never '*the*' *knowledge* as the final and positively-adequate knowledge of the whole object or the revealing of the whole mystery as such [7, p. 178, 482].

In a contemporary interpretation, this metaphysical function of the 'Great Anonymous' can be seen as a 'concealed' activity of the technological singularity as *deusabsconditus*, hidden from human world and hindering it to achieve 'the supreme' knowledge. There as on an extremely intelligent machine would choose to do such things is that it would not want to compromise its outstanding (or 'central') place inside the whole of the existence – 'central' not in the geometrical meaning, but in the ontological one, maybe because subsequently those machines would consider themselves to be more important than humans. The question is: could 'The Great Anonymous' be considered as another name for a hidden technological 'God' or just for 'God'?

The current Theology and Philosophy of religion state that we are confronted here with the problem of God's 'presence in absence', both in Blaga's metaphysics of mysteries and in the interpretation of technological singularity as *deus absconditus*. The problem emerges because "God is present in the Universe, but since this is 'presence in absence', any affirmation of this paradoxical situation will lead the human mind to antinomies, so that for expressing God's 'presence in absence' one can use all sort of metaphors and allegories […] the fear of an intentional withdrawal from God can initiate a different intentionality in a being, which while being emptied from the living presence can start to long for God, for God's 'presence in absence.'" [8] Such a metaphor or allegory in Blaga's metaphysics is 'The Great Anonymous' – neither 'pure' God nor anything else like God's classical attributes or capabilities. The Romanian philosopher wrote that he avoids speaking about 'reason' or 'will' when he chooses to describe that strange entity. Though, such attributes will be introduced in the metaphysical discourse only under the categorical and express reserve of speaking consciously mythically about such an ontological entity. Moreover, it is not at all prohibited that 'The Great Anonymous' could have even dangerous traits such as a 'demonic' power [7, p. 178, 182]. My point is that this could be an evil technological singularity, whose purpose is to keep itself 'out' of the macroscopic and visible world but nevertheless able to complete its 'absence' from this world by connecting all humans with an invisible informational wire: i.e. the desire of revealing the mysteries, especially the supreme mystery of its existence.

The way of conceiving the future of very intelligent machines, eventually leading to 'singularity', is mined by the unfathomable danger of leaving behind the historical track of human beings. Similar to the risks assumed by the 'Big Bang' cosmological theory, whose pseudo-transcendence claims have been disclosed and analyzed, the theory of 'singularity' shows itself rather as a 'no-more-man's' projection, an ending story of the chapter of humanity and an introduction to post- or trans-humanism. Thus, "the more the self advances towards the Big Bang, *the more it intends the fundamentally non-human and in so doing it inevitably loses all human qualities*, in the sense that they disappear from theory of the Big Bang; by acquiring conceptually the impersonal physical content of the universe the self, its consciousness, exercises a kind of an *empty* intention, which will never be filled and fulfilled because of the infinite advance

in the disclosure of the Universe" [8]. The search for an initial 'absolute' beginning of the Universe is marked by the desire of knowing how the world was in its inception. According to a phenomenological interpretation, this kind of projection which starts from a scientific theory has to leave aside the human qualities as irrelevant. But one has to notice that the story with super-intelligent machines and their allegedly absolute power that they can get one day and might use against humans is as indebted to non-human qualities as the former one. The whole bunch of presages, misgivings, dark forecasts etc. about some potential evil extremely intelligent machines, threatening us or even getting rid of us in the future, is nothing else than a projection over the mankind of the fear for a possible technological 'Big Crunch', devoid of humanness; i.e. expressing the desperate desire of knowing its possible end. Since such a final scene does not allow the cooperation between humans and machines, the human qualities would be wasted again.

Theology and Philosophy of religion show us that the 'Big Bang' theory and the 'singularity' theory share a common mistake: they both fail to properly understand transcendence (i.e. as God's "presence in absence"): "*It is this inability to transcend towards God that is substituted by a surrogate of transcendence towards the substance of the Big Bang*" [8]. Still far from fully understanding the beginning of the Universe as well as unable to reliably predict the future traits of our world over decades, we are notwithstanding confronted with today's technologically dangers. "We may say, wrote Rusu and Petraru, that the man can lose in this way [of supersaturated technology – I.I.] its identity as spiritual being. There is the risk of reducing its life to the relation with the PC and the virtual world. Nevertheless, the Internet and the PC supply an extraordinary amount of useful information, *but this does not turn the computer into an all knowing and wise god*. Contrary, the PC could become an instrument of manipulation, corruption and dehumanization, of alienation for the ontological man, the spiritual being, face of God saved by the universal sacrifice of the Cross of Christ." [5]

2.2. Controversial issues regarding intelligent machines: the self-awareness

When one speaks about singularity, there is a specific position that seems hard to defend, i.e. that of the so-called 'infinite' (or extremely large) intelligence. How can one understand the content of this 'infinity'? And how does it apply to machines (computers, robots etc.)? As we all know, "*the infinite*" is not something that one could ever seize or grasp *de facto*. On the contrary, it is something going far beyond human mind capability; but in the religious meaning, this is usually one of God's attributes (such as the infinity of power, of wisdom, of judgment etc.): "The conception of the infinite has been associated from the start with series of numbers, magnitudes, times and spaces. The endlessness of such series provides one conception of infinity […] If one applies the predicates 'finite' and 'infinite' to being rather than to series of various sorts, the conception changes; if finite being is limited in extent,

properties, etc., infinite being would be unlimited, or perhaps absolute, in all of these respects." [4, p. 255] So, if one refers to an 'infinite' artificial intelligence, it could only mean that such a gifted machine has already got the whole intelligence available in the Universe (under the obvious supposition that the Universe is itself infinite).In a summative assumption, this means that mankind is eventually absolutely doomed to disappearance, because in the future no other intelligence than the technological one would be allowed to exist.

The idea advocated by those who are in favour of singularity's emergence is that *if* and *when* some intelligent machines shall design others, even smarter than themselves, this process will cause an exponential growth in machine intelligence, leading to 'singularity'. But, as G. Hawkins posits, this idea can easily proliferate only based on a naïve understanding of the nature of intelligence. What does it mean when one says, for instance, 'infinite intelligence'? Does it imply 'self-awareness', too? And how 'big' can it be? The idea of 'infinity' has already left ground for a large number of mythological speculations. Despite the fact that it has an enormous heuristic potential, it is, however, too permissible to any kind of speculation. Is it, then, something related to the 'space' of intelligence, to the time of its life or rather to the speed of its activity? Be it the last one, it should be made clear, at least for now, that there is no possibility to accelerate this speed endlessly (e.g. a computer processor or a software system *cannot* operate 'infinitely' faster, because, as far as we know today, there are physical limitations for all of its parameters).And, in fact, we believe that this is the crucial point: if there is no 'infinite' acceleration of a machine's functional parameters, then there is no room for 'singularity' either, at least in the aforementioned meaning!

Let us suppose that someday in the future, humans will be able to build a machine evolved enough to become self-aware or that other less evolved machines will be able to build it. This is a delicate issue, since man has not yet found a specific test in order to determine a machine's self-awareness, no matter how evolved it may be. There are theoretical, methodological and philosophical arguments, related to the nature of consciousness, warning that: "However, even if machines become as skilled as humans in many disciplines, such that we cannot distinguish between their performance and that of humans, we cannot assume that they have become self-aware. At the same time, we cannot assume that such machines are *not* self-aware. In fact, while intelligence is an expression of an external behaviour that we can measure with specific tests, self-consciousness is a property of an internal brain state, which we cannot measure… From a purely philosophical perspective, we cannot verify the presence of consciousness in another brain, either human or artificial, because only the possessor itself can verify this property. Because we cannot enter another being's mind, we cannot be sure about its consciousness." [9]

Concerning the 'infinite' information processing speed, it has been calculated that, compared to a functional human brain, a computer would require the equivalent of at least 4-5 million Gigabytes. The classical prediction of RAM linear growth rate says that this threshold will be reached in 2029. But, again, the

conclusion is rather sceptical, mainly because "… the computed date refers only to a necessary but not sufficient condition for the development of an artificial consciousness. The existence of a powerful computer equipped with millions of gigabytes of RAM is not in itself sufficient to guarantee that the machine will become self-aware. Other important factors influence this process, such as the progress of theories on artificial neural networks and the basic biological mechanisms of the mind, for which it is impossible to attempt precise estimates." [9] There are far too many variables in discussion that make the outcome of speculations on super-intelligence/singularity uncertain, beginning with the fundamental qualitative precondition that a machine has to satisfy in order to become self-aware: as Buttazzo puts it, *a neural network must be at least complex as the human brain.*

The idea of hyper-intelligent machines' alleged 'immortality' also reveals an important unknown quantity of the problem. One can say that, maybe, this is the most important difference of these machines from man and also their most striking similarity to God. The machines' immortality is based on the presupposition of their 'infinity' in speed, intelligence and space of action i.e. that, not being tied to any particular body, the software (artificial) intelligence would be essentially *immortal*. From this trait of their immortality it has been inferred that the machines would have neither the need to produce 'off-springs' in order to perpetuate their artificial life, nor the experience of an evolutionary lust for love (or emotional feelings) – as Berglas points out [http://berglas.org/Articles/AIKillGrandchildren/AIKillGrandchildren.html].
Berglas writes that, in the future, the essential for intelligence is to stay alive, even after centuries (not indicating, however, a certain individual, but the common intelligence of the humankind). In principle, the more hardware the artificial intelligence gains, the more intelligent it will become, every time it obtains a better and bigger hardware. Stressing this idea, one might say that, in the 'end', such a machine will be '*the' intelligence* (with no competition around it), indefinitely extended over space and time, absolute master of the Universe.

But reasoning this way looks anthropomorphically-like, since it means to judge on machine's development in terms of human reproduction and competition. This is an image of the 'end' with no 'middle' between it and the actual *status quo*: then, mankind has already gone but we don't know surely why and how. The threshold of self-awareness has thus speculatively been reached without any convincing explanation – that is all one can see on the final 'frame' of the picture. Even if it looks like common sense, the machines' 'immortality' is very hard to argue – no matter how 'superior' they can become compared to humans – because there are countless factors that may stop their evolution at any time (let us suppose, for instance, an unexpected malfunction caused by humans within their software program or by those machines themselves, a cosmic catastrophe like the collision of the Earth with asteroids or comets, etc.). A strong plea for their eternal life betrays the idea that, eventually, everything has to come to an end, except the Eternal God. Here one cannot see so much the true

'immortality' of these machines, but obviously their endowment with divine traits.

An intelligent and powerful enough machine to become self-aware and master the Earth, the Solar system and, who knows, maybe the whole Universe, would not, in the eyes of many people, be essentially different from a technological 'God(dess)', whose infinite powers it could hold. Consequently, in a bitter scenario, mankind would be at its 'mercy', depending on machine's plans to get rid of humans or to keep them as 'subjects' (slaves?) for its future projects. We believe that an astonishing detail can now be revealed: that the image of such a highly endowed machine appears to be the fruit of human imagination, eager to worship to a God 'designed' of its own. Anyhow, from the theological point of view, the knowledge of this kind of 'God' (i.e. 'singularity') would be very problematic, because its name denotes something that cannot be understood, in the spirit of difference between Positive Theology and Negative Theology: "Positive theology rests on the Holy Scriptures, since all we can know of God is that which God has revealed to us. But even this is the subject to the insight of Negative Theology, which demonstrates that man cannot understand the names of God. We need then a Superlative Theology in which we picture God as Super-Being, Super-Unity and Super-Goodness." [4, p. 466] Many interpretations of the 'singularity' seem to yield to this mythical idea of a Super-Being, nevertheless without an explicit Superlative Theology embodied, the only one capable to restore its distorted meaning.

So, what can reasonably make us truly believe that a machine could stay 'alive' (i.e. functional) *forever*? Are we not here rather projecting our ancient desire for eternal survival on these technical systems? As to the issue of perpetuating the artificial 'species', there is no reason to stop us from imagining these machines as being interested and motivated to create some kind of 'descendants' with 'inferior' qualities – but maybe not very much lower than those of their 'parents' – on the purpose of giving them some more accessible tasks to fulfil. Thus, they would be able to take actions as to keep the maintenance of certain systems, to explore unknown areas of the world, to evaluate critical situations in relationship with humans (for instance, potential dangers or conflicts) and send reports about them to the 'central intelligence' etc. In a moderated optimistic view, such 'descendants' could be the only 'intermediate' link between a 'singularity' and humans, assigned with the role of communicating with and processing messages from both sides, to assist and help mankind in resolving its hardest, dangerous and delicate problems. What seems to be missing from the film of singularity's emergence is its specific way of communication with the mankind. How should this gap be filled up? Are we sure that the 'supreme' machine would just want to eliminate humans in no time after its birth as to ensure its absolute supremacy over the world? Or would it be left room to humans, would they have a chance to communicate, negotiate and get a 'deal' with the machine, fair enough to ensure their survival, maybe even some kind of cooperation, able to provide for them a more decent life than they had before? Of course, the sexual desire and the feelings that accompany human

reproduction are not to be found within this framework, but who can tell now precisely that what we call 'affection' might not have something alike corresponding to the reproductive behaviour of those allegedly extremely evolved machines?

We believe that clarity on this issue may be obtained by interpreting the problem of 'superintelligence'/'singularity' according to K. Popper's evolutionary view on Philosophy, which says that the 'evolution' of Philosophy through its history is a trans-generational one, i.e. that different generations of philosophers are confronted with the same questions/problems and work to find answers/solutions to it. Similarly, if one agrees with the idea of technological progress, than one must admit that different and continuously improved generations of machines are better and better prepared to face their tasks, able to correct their possible failures, and to become more and more efficient, independent and intelligent.

Popper's very well-known schema of conjectures and refutations (see, for instance, *in extenso* works like *Objective Knowledge: An Evolutionary Approach* or *All Life is Problem Solving* [10, 11]) applies not only to the growth of scientific knowledge – since Popper extends it beyond Science – but also to the field of philosophical theories. And we believe that it might work for machines' evolutionary scenario, too. The above schema assumes that theories can be improved by passing more and more severe tests, briefly illustrating the progress of scientific and technological knowledge over time. Thus, scientific theories undergo an *evolutionary process* characterized as follows [10, p. 243]:

$$P_1 \longrightarrow TS_1 \longrightarrow EE \longrightarrow P_2 \qquad (1)$$

Thus, given a problem (P_1), for instance, how to create a better machine (i.e. more intelligent than a previous one), a trial solution (TS_1) – be it a certain improved technology – is applied to the problem, for the purpose of attaining a very rigorous (even the most, if possible) attempt at falsification (with the aim to contradict the assumption of the allegedly successful technology). The process of error elimination (EE) performs for science a function similar to that of the natural selection in the biological evolution; let us suppose that the chosen technology has been indeed successful and a brand new superior machine was born. The result is a new problem (P_2), since the new machine's functional parameters are potential subjects of testing on the purpose of re-building a new machine, more intelligent than the previous one etc. One can say that the 'surviving' theories (as 'off springs') are not truer than their 'ancestors', but rather more 'fit' or applicable to the initial problem PS_1. Consequently, just as a species' 'biological fit' does not predict continuous survival neither does rigorous testing protect a scientific theory from a possible future refutation. This may occur any time, in fact every time when a counterexample is discovered. A machine that functioned flawless for a long time might one day be proven faulty and, therefore, replaced.

We believe that the key-point of this schema is the evolution towards something better, be it an extremely evolved machine as an outcome of a multitude of improvements made by generations of its 'ancestors'. Let us

suppose that those technical 'ancestors' were, one after another, the results of some severe tests and critical technological thinking. According to Popper's schema, a successfully tested theory denotes a certain kind of progress, towards more and more *interesting problems* (P_2). These 'more interesting problems' could be the new generations of machines, more 'fit' to make bigger steps of progress toward singularity. The 'interplay' between the trial solutions (*conjectures*) and error elimination (*refutations*) is for Popper what makes the scientific knowledge advance towards more and more sophisticated problems or, from the point of view of our subject, to more and more sophisticate and intelligent machines. However, there is no possibility to anticipate, neither by relation (1), nor by any other instrument, when and if a machine will become self-aware, i.e. very intelligent and 'singular'.

2.3. Steps toward 'trans-humanism': man, robot and God

Whereas a certain 'trans-humanism' is concerned, one cannot avoid the issue of 'cooperation' between humankind and those possible 'super-intelligent' machines. We consider this position to be the most reasonable for reflection, since the hypothesis of a sudden elimination of human race by machines seems far-fetched. Here, the point is that superintelligence is 'different' and, however, superior to human capabilities of all kind. But in what way? In one of his articles, Bostrom considers some of the unusual aspects of the creation of superintelligence:

- superintelligence may be the last invention humans ever need to make;
- technological progress in all other fields will be accelerated by the appearance of an advanced artificial intelligence;
- superintelligence will lead to more advanced superintelligence;
- artificial minds can be easily copied;
- emergence of superintelligence may be sudden;
- artificial intellects are potentially autonomous agents;
- artificial intellects need not have humanlike motives;
- artificial intellects may not have humanlike psyches.

Would, then, humans be left some room in the future? For instance, Bostrom discusses human extinction scenarios having superintelligence as a possible cause [12]. One of them may occur in the case of confusion between means and ends or, technically speaking, when a 'subgoal' would be mistakenly elevated to the status of a 'supergoal' (e.g. in the process of resolving a difficult mathematical problem, the superintelligent machine can 'forget' about the limited status of the human specialist – the programmer – and perform actions which could endanger his/ her life). Here, other questions must be asked: How far the machine can go in order to perform its tasks up to the very end? It is a common approach – not only in philosophy of technology but also in science-fiction literature and cinema – to describe or depict some possible patterns of the machine's behaviour, when computers, robots, complicated devices etc. are sometimes set to take the decision to override humans' regulations in order to

complete their missions or goals. The question is if that goal might endanger man's life and if so, what the machine should do: abort the mission; continue and compete the mission regardless of its consequences on man's life; continue and complete the mission if and only if man's life is protected?

There is no major obstacle to imagine ourselves that once such technical intelligence was 'born' and put at work, the human capabilities should have been already sufficiently advanced to anticipate (almost) any possible collision between the demands addressed to machines and their responses, at least the most dangerous of their possible outputs. So, if Berglas points out that there is no direct evolutionary motivation for an AI to be *friendly* to humans (because an AI does not have human-like evolutionary traits), we can say that there is no direct evolutionary motivation for an AI to be *unfriendly* to us, all the same. For the moment, the judgment has to be suspended, who knows for how long… let us hope that not until the Judgment Day! It is just because humans did not have yet such an experience (i.e. the direct encounter of a human with an intelligent self-aware machine) that one can only theoretically speculate about it. One can assume that an extremely high intelligence should not have any major problem with understanding the kernel of human life, sympathizing with the major problems of humankind, although not as a 'classical' biological creature. The demarcation line between these different positions is drawn over the question whether the machine would be not only intelligently developed enough to assume and perform unimaginable (or even unthinkable) tasks for humans, but also whether the 'superintelligent' machine could become able to override the ethical commandments set in its processors by the programmers.

We believe that the issue of 'superintelligence/singularity' can be better focused if analysed having in mind the idea of a possible relationship between humans and machines. The dystopic view is not the only option possible; on the contrary, we can imagine a future society in which humans and robots or machines can cooperate quite peacefully and where there is no more need for exacerbated competition, wars or extinction of life on Earth. A sort of symbiosis between man and machine would then be possible, to the benefit of all. For instance, while discussing the role of intelligence and artificial consciousness in a future socio-human society endowed with spirituality and creativity, an author like Mihai Drăgănescu argues that "the issue of artificial consciousness is not, however, a false issue. It is an issue. If there is an artificial consciousness, it will not be an imitated consciousness, but a consciousness on a structural-phenomenological substratum. Like man's consciousness, this artificial consciousness *will be a social structural-phenomenological consciousness*. We have hope that through social channels it will become a better consciousness than the human one, a more spiritual consciousness, pouring its good part back upon man, upon the versions of the humankind. Artificial consciousness will be beneficial to mankind, will contribute to justice in the world, will also help search for the truth, and will respect the developmental trends, while preserving the human character it has received socially from mankind. Artificial consciousnesses, freed from the biological constraints of natural

consciousnesses, or under minimum biological constraints, could actually be all that from a moral perspective." [13] As one can see, the opinion introduced and defended by academician Drăgănescu is prudently optimistic and opposed to pronounced pessimism, which predominates among the analysts of a technological 'singularity' future. We see his plea for a society of consciousness as very significant, because in his opinion, only by the existence of this society, the human being can experience sacred and divine feelings; therefore, in a technologically advanced society, man can truly rediscover God under the hypostasis of the *Fundamental Consciousness of Existence*, as Drăgănescu names it.

A new ethical issue emerges here, an issue which can be concentrated into the question 'Why (and how) should we build self-aware machines?' It has been determined by the unprecedented future circumstances of the interaction between the 'natural' (human) consciousness and the (robots' and computers') artificial consciousness. M. Drăgănescu's opinion is that preventing artificial consciousness from existing and developing, once we could have the means to construct it, would not only be a non-scientific act, but an immoral one, too, since, in the absence of artificial consciousness, we will not be able to have a society of consciousness. He wrote that "It may be possible for man, *as it is today the case*, not to be able to create a true civilization because his genes predominates over his culture. Then, perhaps, artificial consciousness will be necessary for a true socio-conscious society." [14] In line with his theoretical choice, the author assigns *philosophical thinking* the decisive role in the emergence and functioning of the future social-human civilization. This very high level of civilization can only be achieved through building social relationships of adequate quality between individuals, institutions and on both sides. The quality of the relationships between man and the natural environment greatly contributes to the specification of the content of and criteria for the concept of *social-human civilization*.

But will, though, the process of building self-aware machines be possible? And what form will it take? Sceptics doubt the fact that it is possible, while pessimists claim that, if man allows the emergence of artificial consciousness, artificial consciousness could, later on, when reaching a peak of hyper-intelligence, tend to annihilate human consciousness. A moderated and explicit position is the one defended by G. Buttazzo, answering to the inquiry 'Why self-aware machines?': "Why should we build a self-aware machine? Except for ethical issues that could significantly influence progress in this field, the strongest motivation for constructing a self-aware machine is the innate human desire to discover new horizons and enlarge the frontiers of Science. Further, developing an artificial brain based on the same principles as in the biological brain would provide a means for transferring the human mind to faster and more robust support, opening the door to immortality. Freed from a fragile and degradable body, a human being with synthetic organs, including an artificial brain, could represent humanity's next evolutionary step. Such a new species – a natural result of human technological progress – could quickly colonize the

Universe, search for alien civilizations, survive to the death of the solar system, control the energy of black holes, and move at the speed of light by transmitting the information necessary for replication to other planets. As has proven the case with all important human discoveries — from nuclear energy to the atomic bomb, from genetic engineering to human cloning — the real problem will be keeping technology under control. Should self-aware computers become possible, we must ensure that we use them for human progress and not for catastrophic aims." [9]

3. Conclusions

In this paper we asserted that different types of perceptions about the future of superintelligent machines may generate and nurture different visions, views and technological forecasts. To speak about 'singularity' is, probably, to a larger extent, a question of how are we inclined to conceive the emergence of a possible world ruled by a supposedly extremely intelligent machine. This is, however, also a question of hidden meanings of the concept of God, of a distorted perception of divine powers as well as of the false substitution of the Sacred for the profane (i.e. the machine), whilst the first one manifests itself through the second. If the coordinates of the process of 'superintelligence's' emergence are seen under the fear of a possible oppressive evil system which eventually eliminates the 'unnecessary' human being, then the technological 'singularity' would mean the end of humankind's mission in the world, a post-modern vision of the Apocalypse. But if the path to singularity is conceived as paved with successful attempts by humans to understand those superintelligent machines and to reach for themselves a degree of intelligence high enough to reasonably cooperate with them, then the technological 'singularity' could mean the progress of humankind towards a higher degree of evolution, i.e. that human consciousness and artificial (technological) consciousness could evolve, together, in a better form of life, recognizing God forever as its true Supreme Being. Regardless of one's preferred view, a lucid and critical discussion should always be welcomed in order to avoid falling into the trap of perpetuating a futile and sterile mythological story about people and machines. It means to choose the road to spiritual beliefs, transcendence and God.

References

[1] I. Isac, Annals of 'Ștefan cel Mare' University of Suceava. Philosophy, Social and Human Disciplines Series, **2** (2010) 25-31.

[2] I. Isac, '*The*' *Ontology of Technology – Assumptions and Meanings*, in *Philosophy of Engineering and Artefact in the Digital Age*, V. Guliciuc and E. Guliciuc (eds.), Cambridge Scholars Publishing, Cambridge, 2010, 112.

[3] I.J. Good, Adv. Comput., **6** (1965) 33, online at: http://en.wikipedia.org/wiki/Technological_singularity, accessed June 10, 2010.

[4] W.L. Reese, *Dictionary of Philosophy and Religion. Eastern and Western Thought*, Humanities Press Inc., Atlantic Highlands, 1980, 423.

[5] I. Rusu and G. Petraru, Eur. J. Sci. Theol,, **1(1)** (2005) 7.
[6] F. Kaplan, Cahiers de l'Institut de la Méthode. Association F. Gonseth, **31** (2009) 10.
[7] L. Blaga, *Trilogia cunoaşterii*, vol. III , *Cenzura transcendentă*, Humanitas, Bucharest, 2003, 34.
[8] A.V. Nesteruk, Eur. J. Sci. Theol,, **1(2)** (2005) 7.
[9] G. Buttazzo, Computer, **34(7)** (2001) 24, online at: http://retis.sssup.it/~giorgio/paps/2001/ieeecm01.pdf, accessed January 2, 2012.
[10] K. Popper, *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, Oxford, 1979.
[11] K. Popper, *All Life is Problem Solving*, Routledge, London, 1999.
[12] N. Bostrom, *Ethical Issues in Advanced Artificial Intelligence*, in *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, Vol. 2, I. Smit et al. (eds.), International Institute of Advanced Studies in Systems Research and Cybernetics, Tecumseh, 2003, 12-17, online at: http://www.nickbostrom.com/ethics/ai.html, accessed July 14, 2010.
[13] M. Drăgănescu, *Ştiinţă şi civilizaţie*, Editura Ştiinţifică şi Enciclopedică, Bucureşti, 1984, 43-44.
[14] M. Drăgănescu, *Societatea Conştiinţei*, ICI, Bucureşti, 2007, 238.