Artificious Intelligences

Autore: [Stefano Vaj](#) (traduzione di: [Catarina Lamm](#))   *da: [Divenire 5](#) (2012)*

Intelligence is overrated.

Of course, the view that tends to sum up all the independent variables of the contemporary equation, both the economico-political and the cosmological, into two "fundamentals" represented by energy and information, remains epistemologically plausible and, more importantly, strategically useful. And in the same way as the energy that matters is not that remaining constant in agreement with the first law of thermodynamics, but that made available despite the second, so too information has meaning and value only inasmuch as we can extract it, transform it, process it.

In this sense, human intelligence has perhaps always been artificial - what are after all symbols, language, traditions, writing, algorithms, art, strategies, if not "artificial" supports to our handling of information? -, and intelligence certainly plays a central part in any scenario related to posthuman change. An intelligence which today we can generally describe as iterative, fractal, artificious.

But, since Wolfram [1] , we know two things: the first, that - contrary to the "creationist" prejudice which despite the Darwinian revolution still permeates society at least in a metaphorical sense - *arbitrary degrees of complexity can be generated by very simple mechanisms*; the second, that once a given, very low, threshold of complexity is met, *all "systems" are in essence computationally equivalent*, except on a single, albeit crucial, point, that is their relative performance in the execution of a given program.

If Newton could see the universe as a mechanical device, it is certainly more conform per se to our *Zeitgeist* to consider the universe as a computer system, as do Seth Lloyd and others [2] ; yet the Principle of Computational Equivalence is determinant in showing that the ability of a 1990s Macintosh to emulate an Intel PC via appropriate software reflects a more fundamental fact: precisely that *any* universal computing system, including a cellular automaton or a Turing machine or indeed the original 1980 IBM PC, can – provided that we set no limits to its memory and the time we are prepared to wait – emulate *any other* system, including the universe and all its contents.

Of course, another fact that Wolfram underlines is that, again contrary to Enlightenment mathematical bias, inside the space of possible problems *computational irreducibility* is the rule: thus the algorithmic solutions that allow one to know the state of a system after a certain number of steps without having simply to go through all the necessary steps, represent the exception, and not how things normally work. So that the fundamental difference between systems is represented precisely by the number of steps necessary to get to a given solution.

Ultimately, however, such a difference is just a matter of *performance*, not of capacity. The "artificial intelligence" in the myth of the automate is invariably founded on some structural "trick", on some qualitative device of a magical or mechanical or methodological nature, that would allow for a quantic leap, a phase change, in its ability to execute flexible and complex tasks; yet nothing of the sort appears neither to exist nor to be even necessary. In fact, the very fabric of physical, spacetime reality, is increasingly regarded not as continuous, qualitative and "analogical", but as intrinsically digital, atomic, granular, binary, and therefore this should come as no surprise. But, whatever the case, today we know at least, and this also from a pretty empirical point of view, that given enough time any digital computer can do anything that an analog computer can do, as well as it can do anything a neural network can do, as well as it can be indefinitely subdivided and multiplied into units that process information in parallel. We know, in short, that "intelligence" has something to do with architecture only in the sense that some architectures are able to execute given kinds of processing more efficiently and rapidly than others; *not* in the sense that some architectures would be able to do things that others would be structurally prevented from doing.

Now, in terms of raw "power", even a simple system like an abacus is able for example to

perform the operations of elementary arithmetic more efficiently than a complex system such as a human brain. And it may be worth noting once more that, with regard to human intelligence itself, history shows that at all times we tend to emphasise and overrate those skills that cannot yet be easily reproduced or augmented through artificial means: from memorising long literary texts with the aid of metric rules, to the ability to perform mental arithmetic on large integers (something which survived, even after the invention of the positional representation of the decimal system, in the popular admiration for the *idiot savants*), to that of solving complex mathematical problems, of managing databases of unstructured data, of easily deciphering natural languages or alphanumerical or ideographic symbols, up to success for instance in the game of chess or in financial markets.

Therefore, the empirical concept of "intelligence" or "mental ability" as applied to fellow humans is constantly evolving on the basis of the differences in individual skills that continue to matter (and in this respect right now Google, for instance, is once more profoundly altering the set of abilities crucial for educational and professional success). And, conversely, the paradox persists that makes "artificial intelligence", a least in the weak sense, simply be a not very rigorous expression (just another way to indicate a *horizon* or a mobile target...) referring to that which "artificial" systems cannot (yet) do - or at least radically simplify.

In this light, those same apostles of humanist political correctness who abhor any idea of posthuman transformation involving our progress in the AI field at least as much as the fact that a fundamental genetic component may exist in the results measuring the IQ of the individual being tested **3** , should consider that *it is the logical and formal nature of such tests which may allow even systems that would be very weak in a Turing text in comparison to a genuinely human idiot to outdo by far a biological brain in this area*. Here too, of course, it is empirical intelligence that counts, not Intelligence as the alleged smallest common denominator hypothetically shared by all or almost all humans being, with no real distinction between them...

In the meantime, the artificial systems that we invent continue to surpass us in an ever increasing number of fields. Indeed, when assessing a typical complex system definable as a fyborg **4** , or "functional cyborg", represented by a human whose biological substance is *supplemented* more often than literally replaced, the scales tends systematically to tip in favour of his "cyber" components as opposed to his strictly biological components: namely, in favour of the man endowed with a computer instead of the man with a slide ruler; of the man who delegates an increasing part of the definition and resolution of problems to machines instead of the man who programs them in Assembly. Little surprise in that, given that this is after all exactly the purpose of artificial systems, the reason they have been developed, adopted and perfected. From this point of view, one could say that artificial intelligence has, in more and more fields, surpassed "human", strictly speaking, intelligence *since it was invented*; but for the fact that *this is only one way of looking at things*, another equally plausible being to see it as simply *integrated* to the latter and extending its originary abilities.

Of course, in the past theoreticians and researchers in the field of artificial intelligence may have deserved - including perhaps as a result of the legacy of an anthropological (and before that a biological) vision of things still dominated by positivist mechanism and reductionism - a reputation for unreliability and superficiality. Many proclaimers of the "manifest destiny" of the mechanist paradigm, as well as many prophets of doom obsessed with the Frankenstein myth, are still probably too quick in underestimating the abilities of biological brains, which it is Darwinistically reasonable to suppose to be, despite all the energetic, dimensional and architectural limitations that afflict them, more than passably optimised to do *that which they have after all been evolved and fine-tuned to do*. Hence, it is not so strange that they may be able to outperform drastically systems with colossal calculation capacities in activities such as pattern recognition or motor coordination or fuzzy logic; and it should therefore never be taken for granted that different architectures, for instance those typical of an electronic processor, could easily compete – let alone at an equal level of energy consumption or volume occupied.

On the other hand, a biological brain is a *finite system*, with a finite (albeit astronomical) number of states. The behavioural and computational responses that it generates - brains do also have other functions in human and animal physiology - are accordingly reproducible *by definition* by any other system that enjoys the property of computational universality.

It must be stressed that this conclusion says nothing per se on the issue fundamental to "classical AI" of whether strategies exist, and if so what they could be, that would allow to reproduce in practice, in the short term, with acceptable performances and with a sufficient

degree of accuracy, behavioural responses that are typical of human beings, or in any case living beings, via "clean room" and "top-down" reverse-engineering procedures - i.e. entirely setting aside the structural mechanisms by means of which organic brains can generate such responses, and high- or low-level reproductions of such mechanisms.

This conclusion concerns nevertheless a more fundamental issue, which pertains to the very nature of the implicated processes and their reproducibility *in principle*.

The doubts that have been put forward in this last respect allege the limited and rather rudimentary results achieved so far (or else the entirely different strategies adopted to achieve similar results, e.g. in the case of software playing chess) to support the theory - which can be easily deconstructed as a reformulation of the Judaeo-Christian concept of "soul", albeit possibly beneath a thin secular veneer - according to which the human brain and/or the human mind would have some irreducible quid, a mysterious ingredient, destined to forever elude any attempt to this effect. A fundamental and still up-to-date catalogue of the positions that can be summed up in this objection is the debate contained in the booklet *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI* **5** , promoted and edited by a well-known humanist and creationist think-tank, the Discovery Institute, where we find one after the other interventions by George Gilder and Jay Richards, by William Dembski ("How," Dembski asks, "can a machine be aware of God's presence?"), by the usual John Searle **6** , and by Michael Denton, with a reply by Kurzweil himself.

This issue is of course also connected to the debate on the concept of "consciousness" or "identity" that resurfaces periodically even among transhumanists, especially in relation to a *Gedankenexperimente*, or sometimes to more concrete technological hypotheses, that have to do with for example the continuity and "survival" of a given person in the case of scenarios such as mind uploading or a destructive teleportation. **7**

Naturally, however, for whoever adheres to a post-Kantian vision of reality the "essentialist" diatribes in matters of qualia and of philosophical zombies are only the products of a thinking that is still under the influence of metaphysical dualism, to which one may oppose, even before Nietzsche's mental cleansing or the Vienna School's methodological cleansing, common sense, the very common sense that recognises that if something walks like a duck, swims like a duck, quacks like a duck, then it is reasonable to call such a bird a duck. Or to put it in more formal words: that "consciousness", "identity" (personal identities as much as for what even more obviously pertains to collective identities), "personality", are all concepts that can only be defined in *sociological*, rather than ontological, terms. Hence the considerations that exceed phenomenic reality, and/or the practical goal we have in mind when examining it, are from this point of view, as they say, "not even wrong".

Therefore, with respect to intelligence, the criterion represented by the Turing test represents after all just an empirical reduction of a concept with much wider implications, which is by the way the one usually applied in our daily interactions with other human beings or with higher animals, when we attribute to them intentionality, agency, motivations, etc., through what in neurolinguistic programming talk is defined as the "hallucination" of our inner states on other more or less similar beings **8** - from whose actual subjective experience we are in reality by definition as excluded, even with an identical twin, as we are from that of a PC, of a storm or of a stone – simply because such an hallucination can be (although need not be) useful to *comprehend*, and not just understand, the world around us. **9**

The theme of the intelligence of human beings relative to animal intelligence is in fact crucial to the discussion on the possible reimplementation of analogous features on different supports, because if human beings are currently the only known species that exhibits this or that trait in this area **10** , most of the alleged "irreducible peculiarities" of our brain can in reality be generalised, to various extents, to other organic brains and nervous systems. In this respect, there exists an evident morphological, structural, functional, etc., "continuity" between our own nervous system and the brain of other primates, other mammals, other vertebrates, and so on, so that most hypothetical "qualitative" differences of the human brain should be extended, by concentric circles, to systems which presents various degrees of analogy to it. And the thesis that the nervous system of an octopus could never be emulated on a computer because octopi would be "created in the image and likeness" of some transcendent being is certainly much harder to sustain even for those adhering to the most rigorous form of "anti-reductionism". Not to mention how arduous would be to find something genuinely "special" and "elusive" in the even humbler cognitive performances of an amoeba.

These considerations have not a mere philosophical scope, because there are ongoing research projects that envisage the realisation within ten years - provided that technological progress in the field of computerised scanning does not have some surprise in store... - of a model that explicitly describes at the level of neurones and synapses the brain of one of geneticists' favourite organisms, the fruit fly, also known as *Drosophila monogaster*. If on average a human brain has about 100 billion neurones, each connected to a thousand synapses, the insect in question has just about 100,000 neurons (of which around 16,000 have already been mapped [11]) arranged in two hemispheres, 41 local processing units, 58 connections that rely a certain unit to other parts of the brain, and 6 hubs; all figures that obviously makes the problem of an emulation at such a level of resolution, even though gigantic, more tractable by many orders of magnitude.

From here, the passage to an emulation of the human brain appears in principle as an essentially *quantitative* matter. And the scenario does not change much when one remarks, probably rightly so, that human intelligence is not just a matter of "brain", and that "mind" would in reality be what it is only when "situated" in a *body* and within the relative proprioceptive and sensory context – an idea that can be summed up in the assumption according to which artificial intelligence in an anthropomorphic sense is a *robotic*, more than a computer science, issue; or that others describe in terms of the difference between a purely inferential intelligence and one that is also *referential* (and as such able to access a "semantic" level that would in a certain sense be precluded to the former). As a matter of fact, the brain and nervous system of human beings represent and constitute a very relevant part of the complexity of the human "system" as a whole; so much so that the emulation of a whole body with its sensorial and proprioceptive input represents a problem that is only marginally more complex than the low-level emulation of the mere brain; and there are no apparent reasons why once the first problem has been solved there should be any particular difficulty impeding the resolution of the second.

There remains however the thesis of Penrose and others, avowedly "anti-AI" but within a perspective that remains "physicalist" and does not rely on the intervention of ineffable qualities, according to which the performances of human brains - but the claim should by necessity be extensible to all organic brains - would exceed the ordinary or algorithmic space and would depend upon quantum effects exploited by the brains themselves, brains that therefore could never be emulated by ordinary computing devices [12] (or, more strictly speaking, could only be emulated "asymptotically", in a timeframe tending to infinity).

Yet, as is obvious, we live inside an *integrally* quantic reality, which is shared not just by organic brains, but also by combustion engines, by traditional computers themselves, by stones and by stars. Recent research tends in addition to corroborate the importance of quantum effects in the strictest sense even in the case of macrophysical reality, including with respect to the processes that are in no way connected to intelligence, such as chlorophyll photosynthesis [13] . And the evident concern to "salvage" something special inside the human mind/soul (be it only a special feature shared with the fruit fly) renders automatically suspicious such hypotheses, which many consider as very implausible having regard to the radically different scales of the effects concerned. [14]

It is however Occam's razor that comes first into play. Today we have a pretty good idea of the kind of computations where a quantum processor, which we are not yet able to build but can describe fairly accurately, would make a substantial difference (the classical example in this field being computationally intractable problems such as the factorisation of very large integers into primes). Well, in all such activities human brains perform even worse than do traditional digital processors. Conversely, organic brains do not exhibit any of the features that theory predicts for quantum processing. The exploitation of quantum effects therefore represents at once an unwanted *deus ex machina* and a *homunculus* that explains nothing of the features concretely exhibited by the organic brains themselves.

But there is more to be said. In principle, if the ordinary functioning of animal brains really were intrinsically based on quantum effects, this would finally demonstrate the trivial (and so far in no way actually proven!) physical feasibility of quantum processors of remarkable complexity, which hence would reveal themselves to be altogether banal in nature. Such discovery in itself would certainly be good news; but would automatically refute the idea that the exploitation of quantum effects would imply the impossibility of functional emulation of organic brains on an "artificial" support. Simply, the formers' "software" would have to profit, to achieve reasonable performances, from a quantum (co?)processor instead of running on a traditional, albeit highly

parallel, computer.

Sure, future confirmations to this effect, or analogous discoveries of other peculiarities of animal brains, might result in that even their most barely practical emulation would require architectural options not just widely different from Searle's Chinese Room, or from implementations on systems of vacuum tubes imagined in the discussions on artificial intelligence in the fifties, but also, at least for the crucial components of the relative platform, much more similar to... a brain, or even to an entire human organism, than to a present-day computer. Exactly along the lines of my discussion in the book *Biopolitics. The New Paradigm* [15] that the most efficient system to calculate the embryo of a mammal on the basis of its genetic code might in the end reveal itself to consist in the relative DNA plus a uterus.

But for the time being such conclusions remain anything but granted; they do not at all affect the question of the emulability in principle of organic systems, under any functional angle and and at arbitrarily low levels, all the way down to the molecular level; and in any case they would only be a reflection of trends that we can in a certain sense easily generalise to the whole of cutting-edge research and engineering.

In this last respect, we have already mentioned the much broader interest of the possible development of quantum computers. Similarly, the current orientation towards systems of growing parallelism, from HPC to the inner architecture of processors for PCs or smartphones, need not be emphasised. Today, the world's most powerful "computer" in terms of teraflops, or rather petaflops, is represented by Stanford University's Folding@home project [16] , which has involved since its inception over *six million* processing units, of which about half a million are active every day. Such units are those contained in CPUs, GPUs and Playstations of participants who via the Internet freely confer the unused processing resources of their devices in a configuration characterised, exactly like a brain, by high latency and very limited bandwidth, but at the same time by exorbitant parallelism and redundancy. Moreover, if the vision of clanking tin and iron robots has been outdated for decades, the fact that it is specifically on... carbon if anything that the general attention is focused today, be it for future memory and processor technologies or for the science of materials, does not contribute to render especially plausible the idea of a humankind literally destined to go "from carbon to silicon".

Conversely, both the progress of technology and its "convergence" with features of organic systems on the one hand, and the decreasing philosophical fortunes of naïve anthropocentrism on the other, make ever more evident that the intelligence, understood in the proper sense, of a computer or of a system of any type can be expanded indefinitely without ever yielding anything even vaguely similar to a human (or for that matter animal) identity, let alone in the shape foreseen by the parochial and rudimentary psychology that often transpires in discussions on this matter or in bad, mostly apocalyptic, science fiction. And still less anything similar to the more or less contradictory projections, which invariably denote the monotheistic traditions, of human ethology in imaginary Supreme Beings of a metaphysical nature (see under "Robot-God").

Those who already in the beginning of the nineteenth century had largely cast off the humanist cultural paradigm are unanimous in abandoning the idea that the "specifically human" is to be sought in intelligence, or worse in a "degree" of intelligence which would place us at the apex of some universal scale of values, and went on to look elsewhere for its origin and essence, starting for instance from the appearance of the hand-tool binomial or from the conjugation of primate-specific traits with fundamental predator's ethology on which the attention of Oswald Spengler [17] or Konrad Lorenz [18] concentrated. Similarly, with regard to the founder of philosophical anthropology, Maria Pansera writes: "For [Arnold] Gehlen, on the contrary, man knows through his action, via a process of mutual interconnections between perception and motor activity. In other words, it is possible for Gehlen to understand knowing and intelligence, as specific human activities, on the basis of the concept of action: wanting to hold up the essential difference between man and animal as based on 'intelligence' is a radical mistake." [19]

Beyond specific aspects of this question, it is anyway Daniel D. Dennett [20] and Roberto Marchesini [21] who best illustrate that, whatever man may be, the kind of "intelligence" that characterises him is nothing other than the product of the fractal refraction of the individual ontogenesis of each of us and of the (Darwinian) phylogenesis that stands behind us. Thus things like "mind" and "consciousness" are nothing other than expressions useful to refer to specific evolutionary artifacts that, indeed, find their gradual, and increasingly limited, generalisation within animal species, families or classes closest to our own, and not at all in

natural or artificial systems, be these of similar or superior "intelligence", that represent the products of entirely different processes.

In other terms: a mind that we can recognise as such in an anthropomorphic, or at least theriomorphic, sense does not in the least represent an inevitable emergence of any system that reaches a certain level of intelligence [22] - which besides as we have seen is relevant only for the rapidity and efficiency of the requested processing, never for its nature - but on the contrary only the product of the specific trajectory of some replicators that have been subjected to variations modulated by selective pressures, as but one component of the implicit strategy of these replicators in function of their reproductive success. *Product that, besides, is in no way necessitated*: see the manifest evolutionary success and capacity for handling information manifested by systems rather "remote" in terms of easy hallucinations of our subjective states, such as, for example, termite nests.

Beyond this, there exist only "animistic" projections that may be perfectly innocent, perhaps philosophically as legitimate as those that concern other human beings, and even useful in daily life (the car "takes its revenge" or "is depressed" because it does not get the proper maintenance), but with a predictive capacity as to the "ethology" of the system or the phenomenon being considered that cannot be taken for granted, and must on the contrary be proven case by case, depending on the aspects that are relevant in the context.

If a mind, or more simply a zimbo, [23] should not be identified as anything that intelligent systems are, but as something that (some) intelligent systems *do*, and do by virtue not of their "power" or architecture, but by virtue of their *history*, the only chance to see it emerge, unless one is to emulate a tendentially infinite number of pseudo-phylogeneses while waiting for the spontaneous emergence of sufficiently similar products, is therefore *the deliberate and arbitrary reproduction of its behaviours*, the "output", starting with the biological model that we are interested in emulating and in accordance with the degree of accuracy pursued.

This model, conceivably, could well be a given human being; and in this sense a plausible way to describe a possible AGI that would come close to the anthropomorphic concept of intelligence adopted by the majority of contemporary narratives on "artificial intelligence" *could indeed be that of the "mind uploading" of a specific individual*. The realisation of an artificial intelligence of this kind would therefore coincide with the creation of a system that would be able to make itself increasingly competitive with humans in a generic Turing test (which measures the capacity of a system to avoid being recognised as an emulation in a finite number of interactions with a Tom or Dick) only as a byproduct of its becoming increasingly competitive in a "specific" Turing test (that is the hypothetical test that measures the capacity of the system to be mistaken by Tom or Dick for their respective wives).

Besides, this prospect represents one of the most interesting aspects of the technological hypothesis in question, because we have seen that in other respects the implementation of convincing emulations of anthropomorphic or theriomorphic "mental" processes has no obvious relevance for the processing of information to other ends, and for the capacities of the systems that perform it, including those pertaining to the ability, possibly by making use of appropriate peripherals, to communicate, reproduce, self-repair, self-program, learn from experience, etc.; and we have also seen that it is perfectly possible that - especially in the absence of exceptional "shortcuts" brought by alternative strategies (like those that can be applied to chess) and/or of a reproduction at a sufficiently low level of the mechanisms used by biological brains - even if you throw more and more computational resources at the task, the emulation implemented might remain orders of magnitudes *less performing* than the emulated system. That is, in other words, that it might function at a much *slower* pace than its biological original. [24]

That all this is not just a matter of resources depends on the fact that technologies often run into intrinsic limitations of a practical, if not downright physical, order, as it is shown for instance by the fact that the computer on our desktop does not make use of the 20 or 30 Gigahertz processors that past extrapolations had lead us to expect for the second decade of the new century.

Of course, the correct, and transhumanist, response to this remark is that engineering's *raison d'être* is precisely that of gradually obviating, abolishing, bypassing, eluding such limitations, as happened for example with the so-called Moore's Law, which has continued to describe the exponential evolution of the processing power of our computers despite the prediction of various "ceilings" which should have stopped the trend. But the *way* in which engineering

achieves such results is precisely *the introduction of architectural changes* (for example through the switch from serial systems of increasing speed to more or less stationary-speed systems of increasing parallelism). So that in the quest of ever better performance, for all we know, AGI engineering might well come up in the end with a system very similar to... a more or less conventional human grown inside an artificial womb.

In any event, the emulation of a specific human being from the angle of his verbal and non-verbal communication, to a sufficient degree of accuracy, is likely to represent for many, and probably for the vast majority of future societies, a persuasive metaphor of survival and continuity with regard to the identity of the human concerned, the change of substrate of the latter remaining as irrelevant in this respect as the gradual (and, in roughly seven years, complete) replacement of our body atoms in the course of our ordinary biological lifetime. And there is really no need to emphasise the implications of our possible achievements in this direction, both from the point of view of the individuals to be emulated and for the other members of their families and communities.

Of course, nothing prevents us, along the lines of the kind of emulations discussed above, from taking one step further in the direction of artificiality by emulating a person who has never existed; but it seems legitimate to consider a possible Turing-qualified AGI of this sort as nothing other than an inevitable patchwork of human traits belonging to members of our species that exist or have existed, or at least, by definition, plausibly attributable to hypothetical members thereof.

The reaction to all this by those who share, be it from a positive or negative stance, a more "mystic" vision of possible "artificial minds" is mostly the claim that there are no a priori reasons why the platform that runs a "psychomorphic" subsystem would know limitations similar to those affecting biological systems or should not (self?)modify itself in runaway, unforeseeable fashions, well beyond the features it started out with.

Yet such a vision is flawed by a both outdated and abstract view of biological systems themselves, who as Dawkins has shown cannot in fact be exhaustively describable, and understood from an evolutionary perspective, but in terms of their "extended phenotype", that is taken together as the overall effect that their genes produce on their environment *inside* and *outside* a body whose once rigid boundaries anyway tend today to become fuzzier not only technologically, but also epistemologically. [25]

Accordingly, *nothing at all* of what is potentially accessible to a "psychomorphic" emulation running on a system devoid of any "properly biological" component - even supposing one could make rigorous distinctions of this sort … - would not be accessible as well to a system unendowed with such a "total abiologicity". And not only because, at least within some limits, the same biology strictly understood presents its own plasticity and potential for chance, deliberate or not, as "wet" (that is, more biocentred) transhumanism insists; but above all because the same features of our hypothetically entirely "artificial" system are by definition also accessible by a system that exhibits equivalent or superior computational power but integrates one or more human brains (or bodies) of a traditional and "physical" kind - given also that, conversely, a more penetrating and modern vision of computers themselves ends up identifying them with nothing more than the sum of their peripherals.

In fact, the only argument in support of a fundamental difference between the two scenarios concerns the inevitable limitation of bandwidth that afflicts the integration between a human brain and "intelligent" functional components positioned outside the skull. Here again, the hypothesis that this bottleneck will be attenuated in the future by neuronal interfaces, according to the promises of some ongoing research [26] , is plausible, but appears to be neither decisive (after all, our five or six available sensorial appartuses have already been perfected as input channels for millions of years, and continue in this sense to represent a privileged access conduit to organic brains), nor required. Because precisely the experience of high-performance computing, and of digital computers in general, shows us that when processing resources are appropriately increased bandwidth limitations can be compensated for by simply moving the language of communications between the subsystems involved to ever higher levels [27] ; so that it seems only natural that the fyborg represented by the computer plus its user is bound to continue to shift increasingly in the direction going from the programming of the physical states of electronic circuits towards ever more general and abstract macro-instructions - without for that matter putting into question nor altering in the slightest the allocation of "psychomorphic" features (for example "motivations" or "intentionality"), which in such systems remain

hypothetically confined to the bio-peripheral we call "user".

In other words and to sum up: nothing allows us to regard a hypothetical future AGI as something radically different, from a practical point of view, from a man at the keyboard of a computer of equivalent capacity and of sufficiently complex and flexible programming; what matters, and is going to matter as well in the future, is the *processing power available to the system*, and not the digital emulation per se of "mental" processes; the value of such emulations, which on the other hand are possible and going to improve by definition, lies essentially in a better understanding of the characteristics that define a given identity, and in a promise of immortality for the latter, above all from the perspective of the social context where this identity has been deployed by the interactions that have involved his biological body.

These conclusions certainly do away with eschatological acceptions of a possible technological singularity in our future which could be interpreted as a parousia, as the Rapture provoked by the Advent of Infinitely Good, Wise and Rational Superior Beings dedicated to rescuing us from this Vale of Tears; rather, it reduces the very concept of a historical singularity to the original sense of the metaphor; which, as is the case for cosmological singularities, does not really predict infinite quantities, probabilities greater than one, and other nonsensical results to be interpreted in some mystical way, but simply refers to developments and scenarios extreme enough to surpass the capacity of our current ("human") predictive and theoretical tools; and obviously, in the case of a technological singularity, refers to the literally *transhumanist* will to wish for such a contemplated rupture, such a *Zeit-Umbruch*, to actually take place. Here, it is not difficult to identify on the contrary in the vision of the Singularity of, e.g., Ray Kurzweil [28] , and in his constant comparison between the processing power of singles computers or of the set of all computers connected to the Internet, and the human mind or the aggregated capacity of human minds, an anthropomorphic and anthropocentric legacy that represents the futurist pendant of the residual humanism, providentialism and universalism inherent in the author's values.

On the other hand, these conclusion also do away with the myth of the "Rise of the Machines", which keeps resurfacing, usually under the fashionable and "responsible" cloak of the Precautionary Principle, including in the most unlikely circles. [29] The underlying structure of such myth is simple: the progressive increase of the computational capacity of the systems that we are using will automatically bring about the birth of AGIs not only Turing-qualified but ethologically anthropomorphic and Darwinian in every sense, and this technological "bootstrap", together with an indefinite architectural flexibility, will yield an exponential acceleration in the occurrence of subsequent iterations (machines designing machines designing machines, ever more advanced, ever faster), with the ultimate result that these machines will accomplish a "revolution", "take over", and eventually supplant the human species by more or less violent means.

The pseudo-transhumanist variant of this narrative is inclined to regard the development described above as more or less as unavoidable as do the prophets of doom who announce the extinction of the "human race"; but it believes that it would be possible to "steer" the take-off of those AGIs, perhaps by hardwiring some form of "friendliness" and empathy/servility towards Humanity into their firmware, roughly speaking along the lines of Isaac Asimov's Three Laws of Robotics [30] ; and this in order, amongst other things, to hypothetically prevent the insurrection of the above mentioned AGIs (by definition capable of self-programming), by eliminating from the parameters of their functioning any motivation to reprogram such traits. [31] While waiting to understand how to do it, and while awaiting in a best-case scenario the results of actions aimed at increasing the "awareness" of researchers and governments, many "singularitarians" with these leanings do not hesitate to uphold the idea of international moratoria or regulations of a prohibitory kind for AI research programmes, on the model that we have already seen fostered for biotechnologies. [32]

In is easy however to deconstruct this order of ideas, which has circulated especially in th Anglo-American world, as the umpteenth avatar of the Golem parable, dressed up in a technological and millenaristic language, within a context that takes for granted a value system marked by utilitarian ethics, "natural law" universalism and a typically humanist specieism, in addition to a parochial, Manichean and a largely acritical vision of concepts such as "humanity", "extinction", "friendliness" and so on.

But even within the value horizon of those who worry about the rebellion of possible "Big Bad AIs" and about a Singularity characterised by a so-called exponential "hard takeoff", the claim

that such positions would be the only "responsible", consistent and rational stance, the one deserving general approval, does not stand up to analysis.

If, for instance, by "humanity" we mean the set of individuals belonging to our species and currently alive, the dreaded prospect is that of seeing the birth of a set of new entities, at first avid of care and resources, and in need of complicated and labour-intensive programming, who after a few years would first infiltrate in every sector of our life, surely cooperating with us, but rendering themselves little by little indispensable and gradually independent of their initial programming, and who then would gradually take over, excluding us from most decisional processes, and eventually, depending on the circumstances, either look after us purely out of the respect stemming from the memory of their creation, or abandon us to our fate, or concentrate and marginalise us in institutions, roles and social spaces where even our personal autonomy would fade away, while awaiting our final extinction.

Now, it is easy to see that this menace corresponds exactly to... the scenario that has forever described the succession of biological generations of our species, and is already in place irrespective of the creation of AGIs that would be anthropomorphic only in ethological terms, and thus not be the product of our ordinary reproductive cycle.

Moreover, with respect to millenaristic narratives focused on the risk represented by research and progress in the field of computer science in relation to the possibility that it might end up generating "hostile AGIs" susceptible to exterminate us, truth is that *the entire current human population of this planet is today threatened by an impending mortal peril*, which in the absence of really radical changes will see it totally extinct within a few decades anyway. Precisely because it will be killed, in ways rarely compatible with the eudaemonistic ideals put forward by utilitarian millenarists, by other humans, stupid machines, predators, disease, accidents or simply by... aging.

When confronted with this virtual certainty, it is hard to see how the truly vague, marginal, improbable and essentially incomprehensible prospect of being chased and assassinated by a Terminator controlled, according to the cliché popularised by the neo-Luddite director of *Avatar* [33] , by "hostile" AIs - who among other things would inevitably find themselves farther away from our food chain than would any human being with equivalent firepower! - might indeed deter us with respect to any change possibly offering an even minimal prospect of avoiding or postponing the much more concrete and impending menaces mentioned above.

Truth is instead that, as it ought to be obvious to whoever gives it but an instant's thought:

- a phenomenon or a machine does not need not in the least to be either intelligent (in the sense of exhibiting especial abilities to process information) or of a Darwinian nature (in the sense of being "marked by a selectively determined tendency to behaviours functional to competitive self-perpetuation and growth") in order to be dangerous;
- a Darwinian system, in order to be dangerous to any extent, does not need to be especially intelligent (the AIDS virus or Bill Joy's hypothetical out-of-control self-replicating nanomachines [34] represent two examples from among many other);
- a computer can be boundlessly intelligent and boundlessly dangerous without being Darwinian in the least, and therefore without exhibiting any "mental" processes or motivation of its own, of the kind attributed to the hypothetical "hostile AGI", and this as a result of altogether ordinary indesirable behaviours (for example depending on the motivations provided to it by its "human peripherals", or simply because of "deterministic", but unforeseen, developments dictated by its software and/or the bugs it may contain.)

There are in fact no elements that would allow one to demonstrate a priori that a horse, for example, is an intrinsically more dangerous system than a motorcyclist of some group of Road Warrior-style hooligans only because of the greater "psychomorphic" autonomy of the former in comparison with the bike of the latter.

Remains of course the question of "dangerous *to whom*?". Beyond last-century universalist abstractions and mythologies, humans, domesticated animals, gods and machines don't fight each other as such any more than does the entire set of females of the animal kingdom against the male sex, or do hypothetical social classes that would "objectively" cross the entire spectrum of human societies [35] . Rather, they work *together* in the struggle against collective adversaries of an essentially similar composition, and for their collective, continued (symbiotic or parasitic) existence. It is not by chance that many of us, *pace* Hume, Bentham or Stuart Mill, feed a cat, or tend to garden plants, or celebrate rituals, or paint pictures, or adorn our Second

Life avatar, with resources that could easily save the life of a member of our species at the other end of the world, and do not feel especially bad about doing it.

Sure, recently some human beings have enjoyed the dubious privilege to be among the first victims of weapons with a non-negligible robotic component, starting with the drones that gradually begin to replace traditional aircraft in ground attack. But, surprise surprise, it so happens that the "motivational" ingredient of these attacks remain altogether external to the robotic ingredient of those weapons [36] , and the increasing intelligence of the weapon plays a part only with respect to its *effectiveness*, according to parameters that are no different from, for instance, the explosive strength in kilotons or megatons that one of the parties in a (possible) conflict can throw at enemy targets. And this stresses once more the essential equivalence, for all practical purposes, between on one hand the system composed by a human at the keyboard of a component that incorporates some skills and autonomy, via digital devices enjoying a gradually increasing processing and creative proxy, and on the other a system that would instead implement the "human" component on another support. The latter is no doubt also potentially "dangerous", especially for those who happen to stand at the receiving end of the sights, but *no more and no less than its current, more prosaic alternative*.

There remains of course the disquiet, often expressed in a vaguely "evolutionist" language and as we pointed out expressed also by authors like Bostrom [37] , concerning the expected gradual increase of the role played by non-biological intelligence in our society - with the irresolvable paradox of ethical choices that regard as an insuperable goal the unconditional defence of paradigms that are, even more than biomorphic, strictly anthropomorphic (such as a suprematism in favour of entities with traits that be at least "eudaemonic" [38]), and at the same time promote the idea that it would be possible, or at least dutiful, to try and hardwire a vaguely "friendly" ethology inside non-biological intelligences. [39] Growth which one supposes might precisely lead to the gradual submission, marginalisation and in the end to the extinction of our "species".

Needless to say, the reference to the species represents nothing else than yet an attempt to denote in the biological, secularised, and allegedly "objective", sense an inclusive and abstract set of the kind of "Christianity", destined from an ethical point of view to oppose the pursuance of the interests, possibly conflicting or indifferent but invariably condemned, of the single concrete person, of his genes, his family, his community, his ethno-linguistic group, his race, his culture, etc., or of whatever might represent the assertion of an identity or a diversity or a peculiarity within what, in the perspective of ethical utilitarianism, should instead become the only all-encompassing category outside which it would be impossible to reason in terms of values. And this is a perspective that appears today to be already in crisis in view of tendencies such as the animal rights, or even more the "deep ecology", movements, that, by reducing these presuppositions in absurdum, in fact inevitably end up reopening the door to the relativism intrinsic to the original European worldview and to the inevitable *choice of values* that it implies. [40]

But it is the very concept of "loyalty to one's species" that contains aporias that are hard to overcome, in the light of the same traditional taxonomical category concerned (which among sexed living beings is notoriously defined as the set of all organisms naturally susceptible to breed yielding fertile offspring). And this not only because of the category's theoretical and practical relativisation in the world of contemporary biology, even from a "synchronic" perspective. But, more fundamentally, because in a diachronic sense the concept loses anyway much of its operational meaning, given that every species, retrospectively, represents nothing other than the product of gradual transformation and diversification of the traits of its progenitors, well beyond any conceivable interfecundity (even when the impossibility owing to temporal distance is not taken into account); so that it can be expected that - even though morphological traits of some species show an exceptional temporal stability - given a sufficient lapse of generations the evolutionary pressures will make their descendants unrecognisable, even though they are by definition vertically related, to a point such as to render it absurd to regard them both as part of the same "potential reproductive pool".

From this angle it appears equally arbitrary both to regard man and australopithecus (or, even more so, man and primordial mammals) as part of of the same species, and to imagine that australopitheci should be considered as "extinct" further to the evolutionary transformation of their descendants into other, subsequent and/or parallel, species of the *Homo* family. In this sense, it is typical of transhumanism to reason implicitly in terms of *clade* rather than of species,

and to identify biologically, if with anything, with an *evolutionary line*, in which nostalgia for the future verifies Nietzsche's saying that "The species, seen from afar, is something as *fleeting* as the individual. The 'conservation of the species' is just a consequence of the growth of the species, which is equivalent to a victory over the species on the path towards a stronger species. […] And it is precisely with regard to every living being that one can best show that it does all it can not to conserve itself, but to *become more than it is*." [41]

If one installs oneself in the opposite perspective, provided that a deliberate action on one's own nature is not proscribed by a "providentialist " view, a consistent "humanism" in the specieist sense would paradoxically have dictated to austrolopitheci a sort of antievolutionary "eugenics" imposing at the time the immediate elimination of newborns and germlines that would present proto-human traits, and therefore such as to provoke with time the disappearance of "australopithecity" as an absolute ethical reference. Similarly, it could today suggest measures aimed at perpetuating immutably the gene pool and frequencies that denote "Humanity, version 2011" for all the centuries to come [42] . In fact, a threat to this kind of "frozen specieism", a threat which is much more ineluctable for the survival of "the human species as we know it today", stems no doubt from the mere *transformation* of the latter over the course of time rather than from the sudden emergence of angels or demons or aliens in the guise of silicon artificial intelligence. The range of scenarios imagined by science fiction shows also how nothing guarantees that such a possible change would imperceptibly extend over immeasurable eons, given that - especially at the current level of panmixity - the genetic panorama of our species appears in principle subject to be profoundly modified *in times on an entirely human scale and in the space of just few generations*, either by deliberate enhancements and alterations or by random mutations, which be capable of conferring a decisive reproductive advantage.

But what would an "Advent of AGIs" represent, besides, for our lineage, from the very point of view here criticised? The species, in Darwinian terms, essentially represents as we have seen just a *competitive space*, so much so that Humanity is an ideological construct that is no more tended to in humans than its equivalent is in other animals by some "whisper of the genes", of the kind that instead programs living organisms for parental investment - be it even indirectly as in the case of sterile worker ants - at the expense of their own "interests" and of the survival of the organism concerned. [43] If evolutionary psychology demonstrates the good, and Darwinian, reasons for the existence of various degrees of a spontaneous empathy also between genetically unrelated individuals, the very fact that a genetic correlation is unnecessary shows that the empathic reactions are more the result of the empathiser's easy self-identification with, and projection on, his object - which might very well be a being belonging to another species, another genus, or not even be "living" but mineral, imaginary, virtual - than of a degree of biological proximity per se. So that, beyond the humanist paradigm, there do not exist any particular descriptive or intrinsic elements in our ethological endowment that would justify the promotion, from an ethical point of view, of a radical "us vs them", of a fundamental loyalty, of a specieist nature.

Vice-versa, human experience demonstrates how "descendance", even in its ethological and sociobiological meaning, obviously has intrinsic genetic roots, today invariably takes on an extensive and metaphorical meaning that leads to identify continuity of groups, or successions among individuals, where literal kinship plays only a partial and contingent part, or pulls out over time to the point when the genetic contribution of the forefathers becomes infinitesimally diluted. Or better still: in our species the meaning of an affiliation defined on such broadened bases often reveals itself *stronger*, in one's self-identification with a set of interests, roots and perspectives (for instance of a "national" or "popular" nature) than links themselves dictated by direct, or at least plausible, biological descent.

Consequently, even if possible artificial intelligences with an intrinsic psychomorphic component were not simply identified, at the social level and therefore for all practical purposes, with the personalities that they could emulate, it is only a personal bias to this effect [44] that would prevent one from regarding such AGIs as in every sense "children of mind" [45] , according to Hans Moravec's well-known expression, or even as children *tout court*, like Gazurmah for Marinetti's transhuman Mafarka [46] ; and therefore as the legitimate successors, at a personal and/or evolutive level, of the interested individuals, at least on a parity with any future conspecifics who be not part of his immediate offspring.

On the other hand, as in the field of "bio" and "nano", also in the field of "info", "cogno" and "robo", the moral prohibition to "play god" [47] , in this case under the specific guise of the

scarecrow represented by the creation of psycomorphic beings whose hostility or even mere "superiority" should be feared, gives place to the very concrete threats generally represented by a neo-Luddite trend that today invokes, strengthens and justifies a *slowing down* in the field of research and technology that appears to many as already too far gone – even though it is precisely in the ICT that the pace of technological progress has best resisted the end of last century, also in terms of its indirect influence in other fields and sectors.

We refer here in the first place to the threat that neoLuddism represents to the individual survival of each of us, to the competitive survival of our respective affiliations, and to the very survival (short-, mid- and long-term) of the clade all existing human groups belong to; with respect to which the "intelligence" that is or is not made concretely available is anyway going to play so obvious a part that it needs no further illustrations.

But, even more, our concern goes to the threat and curse that it may represent with respect to the will to knowledge, power, greatness that in our opinion can alone give ethical and existential meaning, from a posthumanist point of view, to survival itself.

After all, if intelligence may be overrated, a stupid life is not really worth living.

Note

[1] Stephen Wolfram, *A New Kind of Science*, Wolfram Media 2002. This work, today accessible in its entirety online at the address *www.wolframscience.com*, has been amply contested for its content, and at the same time for the fact that "it says nothing new". Which appears to be a good indication of its ability to reflect the coming of age of a change of paradigm...

[2] Seth Lloyd, *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*, Vintage 2007. This idea, which today is often assumed implicitly and explicitly in our rapport with reality in general, represents as well, in a somewhat transfigured sense, a posthumous revenge of "primitive" panpsychism that monotheism-originated "world disenchantment", which sees physical reality only as a crude and dead reflection of a transcendental Mind posited outside it, has been opposing for two thousand years.

[3] See the very recent confirmation, for the first time directly based on a genetic profiling of the tested individuals, in the study by G. Davies et al., "Genome-wide association studies establish that human intelligence is highly heritable and polygenic", in *Molecular Psychiatry*, Aug 9 2011, doi: 10.1038/mp.2011.85, of the conclusions by among others Hans Jürgen Eysenck, Arthur Jensen, Richard J. Herrnstein, Jean-Pierre Hébert, and because of which James D. Watson, the Nobel laureate who with Francis Crick discovered DNA, has been black-listed. On this matter see also the author's discussion with Adriano Scianca in *Interview on Biopolitics and Transhumanism* (Italian: *Dove va la biopolitica?*, Settimo Sigilio 2008], which is online in English at *www.biopolitix.com*)

[4] This expression is used among others by Gregory Stock in *Redesigning Humans: Choosing our Genes, Changing our Future*, Mariner Books 2003, to suggest that the implantation of bionic legs does not really represent, for better of for worse, a radical change or even a likely, immediate prospective given that a motorcycle allows one to obtain a similar or greater performance increase in one's speed.

[5] *Are we Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI*, edited by Jay Richards, Discovery Institute 2001. The books expressly represents a "sceptical" reply to the better known *The Age of Spiritual Machines: when Computers Exceed Human Intelligence* by Ray Kurzweil (first edition Viking 1999), but it is significant how in this pamphlet the "it cannot be" remains just an auxiliary argument to the much more fundamental "it must not be". The discussions of the practical impossibility, or even of the philosophical impossibility, of the hypotheses under consideration is thus always taking place against the background of their *moral* impossibility. See also on this matter the approach taken by a writer of science-fiction like Charles Stross who has even expressly hijacked the transhumanist themes and reflections by groups like the Extropy Institute for his own commercial writings (cf. *Three arguments against the singularity*, *www.antipope.org*).

[6] In fact, and this is rarely mentioned, the famous mind experiment of the Chinese Room (for an exhaustive critical literature regarding this "paradox" see *Wikipedia*) contains ambiguous consequences on matters of AI. Indeed Searle is forced to accept and hence to postulate that the output, the behaviour, of the Chinese Room that he talks about must be *identical* to that of a Chinese individual, given that if it were not so the example would automatically be useless as a support for the thesis that *no* "Chinese room" can really "think". But the very

admission of the fact that even a simple a system such as the Chinese room could theoretically exhibit, even only on a timescale that is a multiple of the age of our universe, this kind of behaviour, does in itself lend support to the idea of a "strong AI", at least from a functional point of view and for anyone who does not consider problems of a noumenal nature as worth discussing.

[7] I have dealt with this topic informally in the article *Uploading, cyborgisation, teletrasporto, rianimazione postcrio: possibilità ed identità* at: *transumanisti.wordpress.com*).

[8] See for instance Richard Bandler and John Grinder, *The Structure of Magic*, Science and Behavior Books 1989.

[9] The ability to do it, if we consider difficulties encountered in this respect by persons affected by autism, does seem to play a relevant role in the development of our capacity for normal social and linguistic functioning.

[10] It is now believed that the Neanderthal, who are also credited with indices of "intelligence" at one time regarded as belonging exclusively to our species, such as the use of fire, funeral practices, the probable use of some language, etc., would have belonged to a different species, i.e., to a biological group not normally interfecund with Sapiens, and with even a different number of chromosomes. Besides, studies in last century have proven how the hiatus between our own cognitive performances and those of great apes has been amply overrated, and this as much because of obvious ethological differences as because of an ideological anthropocentric bias the religious origin of which are quite obvious.

[11] See for instance Nicholas Wade "Decoding the Human Brain with the Help From a Fly", in the *New York Times*, 13 December 2010.

[12] Roger Penrose, *Shadows of the Mind. In search for the missing science of consciousness*, Vintage 2005. See also by the same author, *The Emperor's New Mind*, Oxford Paperbacks 1999.

[13] See for instance Roseanne J. Sension, "Biophysics: Quantum path to photosynthesis" in Nature no. 446, 2007.

[14] Max Tegmark ("Importance of quantum decoherence in brain processes" in *Physical Review* no. 4, 2000) stresses for example that the temporal scale involved in the "firing" of neurones and in the excitation of microtubules is slower by more than ten orders of magnitude than the times involved in the decoherence invoked by Robert Penrose and by Stuart Hameroff in their theory of consciousness.

[15] See Stefano Vaj, "Il secolo biotech", in *Biopolitica. Il nuovo paradigma*, SEB 2005 (*www.biopolitica.it*).

[16] The initiative's official site, with relevant statistics, on-going projects, published results, etc., can be found at *folding.stanford.edu*

[17] See Oswald Spengler, *Man and Technics*, Barnes Review, 2002 (original version: *Der Mensch und die Technik. Beitrag zu einer Philosophie des Lebens*, CH Beck Verlag, 1991).

[18] See Konrad Lorenz, *Behind the Mirror: A search for a natural history of human knowledge*, Mariner Books 1978.

[19] Maria Teresa Pansera, *L'uomo e i sentieri della tecnica*, Armando Editore 1998.

[20] Daniel C. Dennett, *Kinds of Minds: Towards an understanding of Consciousness*, Basic Books 1997; but see also, more indirectly, *Darwin's dangerous idea: Evolution and the Meanings of Life*, Simon & Schuster, 1996.

[21] Roberto Marchesini, *Post-Human. Verso nuovi modelli di esistenza*, Bollati Boringheri, 2002.

[22] An extreme example of this idea, widespread above all at the time of the first adoption of electronic processors, is represented by a classic of the Steampunk genre, namely *The Difference Engine* by William Gibson and Bruce Sterling, which is about, inter alia, the birth of a "mind", whose first thought is modelled on the Cartesian *cogito ergo sum*, inside a mechanical punch-card system into which the demonstration of Gödel's Second Theorem had been programmed.

[23] Zimboes are "second degree zombies" invented by Dennett as a reductio as absurdum of the argument that the conceivable existence of philosophical zombies, who consist in unconscious automates who nevertheless behave in ways altogether indistinguishable from those of human beings, would demonstrate that subjective consciousness involves something more than its material-behavioural support. Based on the fact that thoughts are themselves "behaviours" broadly speaking, which manifest themselves in, inter alia, the electrochemical processes of the brain, zimboes, in addition to being indistinguishable from thinking conscious beings, think of themselves, by definition wrongfully, as belonging to this

latter category. Which renders obvious the inherent contradiction in the hypothesis that consciousness can be distinct from the behaviours it originates, given than we could all fit this "zimbo" definition without there being any way in which to falsify the hypothesis. See by this author, "The Unimagined Preposterousness of Zombies", in *Journal of Consciousness Studies*, vol. 2 no. 4.

24 In these hypotheses, a plausible test would certainly not be construed as the scenario described by Turing of a human operator chatting in real time through a teletype to someone sitting next door, but rather as an epistolary exchange, or even an interstellar communication with an entity an appropriate number of light-years away, so as to take into account the degree of latency required by the system.

25 See Richard Dawkins, *The Extended Phenotype: The Long Reach of the Gene*, Oxford University Press 1999. This view corresponds to the traditional transhumanist view that considers inseparable from the nature of an enhanced individual the enhancements thereof.

26 See for instance the Braingate project at its site *www.braingate2.org* , or studies of the kind discussed in *"Writing memories with light-addressable Reinforcement Circuitry"* by Adam Claridge-Chang et al., in *Cell*, Volume 139, Issue 2, 405-415, 16 October 2009.

27 One thinks, for example, of the quantity of information that it is necessary to transmit to a vehicle to drive it in urban traffic to a given destination in comparison with the "compression" allowed by the mere request of being driven to a certain address submitted to a system capable not only of setting the course as can a contemporary GPS car navigator, but of negotiating like a cab driver all the unforeseeable necessary transactions (that yet need not have anything to do with what is commonly regarded as the exclusive domain of a "strong" AI.)

28 Ray Kurzweil, *The Singularity is near*, Viking Press 2005.

29 See for example *Global Catastrophic Risks* (Oxford University Press 2008), edited by Milan C. Circovic and by Nick Bostrom, who yet was one the founders of the World Transhumanist Association, particularly in the essay signed by Eliezer Yudkowsky, exponent of the "Singularity Institute for" (or perhaps today "against") "Artificial Intelligence", whose Web site is located at *http://singinst.org* . The topic of existential risks, or "x-risks", runs a general risk of taking an intrinsically reactionary tone, because given the risk of infinite, or at least infinitely unacceptable, harm, there are no limits, no matter how unlikely it may be, to the costs, human lives included, that it would be "economically" rational to envisage in order to avoid it - except possibly, if such risks are more than one, for a distribution of available resources according to their respective probabilities; but in any case without leaving any resource available to *invest*, by accepting a real (albeit calculated) *increase of risk*, in anything other than mere survival. Such a perspective appears by the way a direct reflection of that of Aldous Huxley' *Brave New World*, that is of the most perfect stagnation possible in exchange for the best hope of stability that one can obtain. The matter is of course further aggravated by the superstitious bias in favour of inaction, which, including for non-anthropic risks, suggests, in case of doubt, an alleged moral superiority of the choice that least affects the independent unfolding of phenomena, in deference to a more or less secularised providential vision according to which, given equal chances, the most important thing is that the harm that might occur anyhow be not of human responsibility, do not depend on any attempt to take one's destiny in one's own hands and to "play god". Cf. on the contrary, the Proactionary Principle about which Max More, the founder of the Extropy Institute, is currently writing a book, as described at the address *www.maxmore.com*

30 See Isaac Asimov, *I Robot*, Spectra 2008, and the other novels and short stories of the same cycle. The best refutation of the "humanist" consistence of Asimov's system is to be found in the cycle of the Humanoids by Jack Williamson, centred on the resistance and gradual defeat of the humans in their resistance against the rampant androids who are persecuting them everywhere in order to comply with the Three Laws according to their own more orthodoxically Asimovian interpretation (but even Asimov himself manifested his awareness of the contradictions in his system in the short story *That Thou Art Mindful of Him*). That such a fictional universe reflects inconsistent hypotheses as to the behaviour, autonomy and anthropomorphism of those that are for all intents and purposes zombies, not even in the philosophical but in the Caribbean sense, is rendered even more conspicuous, a contrariis, in the Berserker cycle by Fred Saberhagen, who speculates about a future cosmic struggle against AGIs whose intrinsic finality is disinfecting the universe of biological life, or at least human life, by means of systematically seeking it out and destroying it. Instead, a much more absurdly naïve version of "bad androids" are the human, all-too-human Cylon in

the world of *Battlestar Galactica*, who are little more than a metaphor of the "aliens" who perversely threaten the American way-of-life and its Manifest Destiny of cosmic hegemony.

31 According to the rather ridiculous example by Nick Bostrom, "Gandhi would never have never deliberately swallowed a pill that would have rendered him capable of killing other human beings". But see in this respect, if anything, the easy objections raised in in *Hplusmagazine* 15/04/2011, at *hplusmagazine.com* , by Hugo de Garis, even though from a not-so-different perspective.

32 See on this subject the chapter "GMOs and other monsters" in my *Biopolitica. Il nuovo paradigma*, op.cit.

33 Cf. Francesco Boco, "La tentazione a-storica", in *Divenire. Rassegna di studi interdisciplinari sulla tecnica e il postumano*, no.4, Sestante Editore (*www.divenire.org*).

34 See Bill Joy's neo-Luddite manifesto "Why the Future Doesn't Need Us" in *Wired*, 8 April 2000. Or even, in the sector of bestselling fiction, Michael Crichton, Prey, Harper 2006.

35 Besides, selection, including in evolutionary terms and contrary to what is implied by still widespread Darwinian popularisations of humanist and "progressist" leaning, does not essentially takes place between species, unless in the sense that some go extinct and others survive by occupying their ecological niche and Lebensraum, but *inside the gene pool of any given species*, and therefore with regard to individuals, and possibly to groups able to act in a coordinated fashion, who are carriers of certain genetic traits.

36 More specifically, the "motives" remain relegated to the human peripherals of the weapons, or, in a more indirect sense, to "mechanisms" such as the Market, that simply represent human superstructures aimed at excluding *political* decisions in the strong sense.

37 For details about his recent theoretical production, and about the Future of Humanity Institute that he directs in Oxford, see the site at *www.nickbostrom.com*

38 But see Nietzsche's stinging irony on the topic of ethical utilitarianism, quoted by Max More in "The overhuman in the transhuman" in *Divenire. Rassegna di studi interdisciplinari sulla tecnica e il postumano* no. 4 (the original English version is available online at *jetpress.org*), according to which "Man does not strive for pleasure; only the Englishman does."

39 This operation in itself would by definition automatically deny these beings their autonomous and therefore "moral" status, and therefore their very "intelligence" in the human sense, as Catholic moral philosophy illustrates well with respect to utopias, or dystopias, that foresee a hypothetical reduction of men themselves to automates programmed to do only "good".

40 Naturally, surpassing in this sense the humanist paradigm - which I confronted in its historical and logical unfolding at the level of juridical and ethical philosophy in my *Indagine sui diritti dell'uomo. Genealogia di una morale*, LedE 1985 (online at *www.dirittidelluomo.org*) ,- also does away with the question, which some strain away at also in transhumanist circles, of the "rights of robots". Indeed, outside of a "natural rights" ideological framework, the possible imputation of legal positions does not in fact depend upon "essentialist" analyses of the rightsholder concerned but, avowedly, on a pure *convention*, such as those that define and delimit the citizenship status within the perimeter of a given legal system, or - as J. Storrs Hall remarks in *Beyond AI: Creating the Conscience of the Machine*, Prometheus Books 2007 -, those already applicable to perfectly ordinary things such as corporations; not to speak of the legal capacities that different legal systems may confer in various ways to embryos, slaves, trust funds, public agencies or even to as yet unconceived possible children (see art. 462 of the Italian Civil Code), and tomorrow in Spain, if the Zapatero's bill founded on the Great Ape Project (*www.greatapeproject.org*) is passed, to chimpanzees, bonobos, gorillas and orangutans, on the model other countries apply to minors and other humans suffering from some natural incapacity. But other questions exist, purely theoretical for the time being, including the emulation of human personhoods on various systems, that have no more need to be dealt with in "essentialist" and universal terms of highly dubious validity, are those that concern the extinction and the succession of the subjects traditionnally defined as "natural persons".

41 Friedrich Nietzsche, *The Will to Power*, aphorisms 280 and 302.

42 Something which risks however to take place, because of a gradual loss of the array of intra-species diversity, as a consequence of the possible evolution of our world in precisely the direction laid out by the globalisation of a sclerotic Brave New World that, because it puts all its eggs in one basket, is for this reason susceptible to find itself *more* vulnerable to the challenges of time than would a richer and more plural human, or posthuman, landscape.

43 See for instance David Barash, *Sociobiology: The Whisperings Within*, Souvenir Press Ltd 1980, or Yves Christen, *L'heure de la sociobiologie*, Albin Michel 1979.

[44] See the avowed bias emerging in the last part of Charles Stross's *Accelerando*, a confused fictional hotchpotch of transhumanist themes that ends with the remaining humans, even though "enhanced" and not unwilling to have themselves uploaded into... bird flocks, fighting and fleeing the Vile Offspring (!), a derogatory epithet employed to refer to the "weakly divine" intelligences who from then on dominate the inner part of the solar system on its way to restructuring, and who in order to attain the maximum efficiency necessary to compete in Economy 2.0 would have "reduced" themselves to "pure mechanisms" - that is, have accessed a posthuman level that is unattainable and incomprehensible to those who are not prepared to walk on the their path, so much so as are human cultures for an australopithecus.

[45] Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press 1990.

[46] Filippo Tommaso Marinetti, *Mafarka the futurist*, Central June 1998. In fact, in the futurist parable that is the subject of this 1909 novel, Mafarka definitely reaches the stage of a superman only by his decision to create Gazurmah as the "pure" fruit of his will, and by his final gesture in which he infuses life into him.

[47] See for instance Jeremy Rifkin and Ted Howard, *Who should play God?* Dell Publishing Co. 1977; or Bill McKibben, *Enough: Staying Human in an Engineered Age*, Owl Books, 2004.