Marko Uršič
University of Ljubljana, Slovenia

Virtual reality and the human world

Paper presented at the symposium *Humanism and Humanities*
Slovenian Academy of Sciences and Arts, Ljubljana, November 2014

Abstract: Nowadays, the danger of modern science and technology – the attention to which was drawn already by Martin Heidegger – is not in any way smaller, in a sense it is even bigger than it was in the middle of the previous century, as the essence of contemporary technology is far more concealed to human thoughts and, on the other hand, because the majority of people accept and take for granted the development of technology, assuming it is human-friendly. Since the end of the Cold War between the two super powers, the danger of nuclear clash has, indeed, decreased, but new technologies are becoming latently more dangerous, particularly biotechnology and cyber-technology. Through cyber-technology, our "real world" becomes more and more penetrated by the virtual world, causing to disappear the differences between *sensory* and *virtual* reality (as if we were on the way back to Plato's cave of shadows), as well as the difference between live human *consciousness* and *machine*. The gradual, yet persistent process of "cyborgisation" of the human being, society and the world is taking place. However, that which might be the "saving power" (Hölderlin) does not imply the need to turn away from science, neither it is a way of the utopian return to some "genuine human being" and/or to "Nature" (Rousseau). Rather, the "saving power" lies only in human persistence in thinking, self-awareness, ethics and spirit.

Key words: virtual reality, cyber-technology, Martin Heidegger, Vilém Flusser, Jean Baudrillard, Vernor Vinge, Ray Kurzweil, Nick Bostrom

The main theme of this paper is to address the question of the role, influence and meaning of *virtual spaces* as assumed in our human, historical, "real spaces" (physical, living, social, etc.), and in our "real time," i.e., in the few decades following the first wave of cyber-technology, when we are overwhelmed by its second major wave – internet, the expansion of the "World Wide Web" in which virtual spaces/worlds have become *interactive* in relation to our own "real world" by penetrating into our human and historical present. We observe that *virtual reality*, originally confined to games, cyberarts, models and simulators (e.g. flight simulators), is gradually becoming a part of the global *reality* itself. However, the latter is not necessarily ontologically *real* in the sense that philosophy has been seeking since pre-Socratic times (and in this quest frequently despairs). Since I am a philosopher whose fundamental conviction and philosophical search is Platonic, I would like to emphasise that we have to distinguish between "real" in the sense of *actual*, and ontologically real in the sense of *true* being. With this distinction I endeavour to maintain "transcendental tension" in the world, i.e. the *possibility* of "transcendence-in-immanence".

A “danger” of modern technology, pointed out by Martin Heidegger, predominantly in his discussions in *The Question Concerning Technology* and *The Turning* (English translations quoted from: Heidegger 1977 and 1962) is now even greater and more concealed in its self-evident universality than it was in the previous century. Let us look at an extract taken from this discussion by Heidegger:

> “…poetically <*dichterisch*> dwells man upon this earth. The poetical <*das Dichterische*> brings the true into the splendour of what Plato in the *Phaedrus* calls *tò ekphanéstaton*, [that] which shines forth most purely… The poetical thoroughly pervades every art, every revealing of coming to presence into the beautiful.
> Could it be that the fine arts are called to poetic revealing? […]
> Whether art may be granted this highest possibility of its essence in the midst of the extreme danger, no one can tell. Yet we can be astounded. Before what? Before this other possibility: that the frenziedness of technology may entrench itself everywhere to such an extent that someday, throughout everything technological, the essence of technology may come to presence in the coming-to-pass <*Ereignis*> truth.”

Heidegger's words on the dangers of technology are – intentionally, of course – ambiguous. The essence of technology as well as the essence of art, and particularly poetry, is “the poetical,” the coming-to-pass of being. In the very same sense, we may also understand the renowned verses by Friedrich Hölderlin, emphatically cited by Heidegger: *But where danger is / grows the saving power also*. – The first wave of cyber-technology, reaching its peak in the 80's of the previous century brought about an interesting “turn” in the attitudes of many artistic creators and theorists towards art and technology (and science): for them, cyber-technology did not represent a danger for “poetical dwelling of man upon this earth,” but rather, the opposite – the most promising possibility of some new human creation and freedom. This period in the development of cyber-technology may be labelled the “time of cyber-optimism.” One of the main theoretical protagonists of this optimism was Vilém Flusser (1920–1991), philosopher and communicologist, expert on the theory of photography and, specifically, “techno-images”, born in the Jewish community in Prague, emigrated to Brazil in 1940, later returned to Europe, and after 1972 lectured predominantly in France and Germany. Flusser introduces his renowned essay *Digital Apparition* from the 1980's by stating:

> “Before our doubting eyes, alternative worlds begin to emerge from the computers: lines, surfaces, and soon also bodies and moving bodies, made up of point elements. These worlds are colourful and emit sounds, and in the near future they will probably also be touched, smelled and tasted. But that isn't all, because the moving bodies that will soon be realised through calculation and which are beginning to emerge from computation, will be equipped with the artificial intelligence of *Turing's man*, so that we will be able to enter into dialogical relationships with them …” (English translation quoted from: Flusser 2000).

These findings, written by Flusser a few decades ago, contain more than just a grain of truth and although contemporary computers still cannot perform the Turing test (i.e. their intelligence

is not yet equal to human intelligence), information specialists and futurologists predict that computers will surpass human intelligence sooner or later, perhaps as early as in the first half of this century (see below). Nevertheless, contemporary computer software in all its colourfulness, flexibility and above all *interaction* with us, members of the species *homo sapiens* is more and more "realistic," sometimes even three-dimensional, through holographs and similar applications. Flusser wonders "why do we disparage them as 'apparitions'? Why are they not real for us?" (*Ibid*.) For if "reality is measured by the density of distribution" of computed point elements, bites and pixels, then "the elements can be distributed as densely in the hologram of the table [as in a 'real' table] and our senses will no longer be able to distinguish between the two" (*ibid*.), of course, by assuming that they shall not only be visual, but also haptic (tactile) "pixels" of the virtual table as densely and sensually persuasive as visual stimuli are, received on the part of the real table by our eyes and then sent to the brains. But is this *really* going to happen? Although it is not very likely that such complete erasure of the difference between the virtual and the real world should happen in the near future, let us say that in this century, in principal, we cannot rule this out when we consider a more and more powerful technological network of cybernetics, biotechnology and nanotechnology. It may happen that sometime in the future the *consciousness* shall no longer be able to distinguish between the virtual and the real world, even if it maintained its own self-evidence in the Cartesian sense of *cogito ergo sum* (I think, therefore I am) or in the Husserl's phenomenological variant of *ego cogito cogitatum* (I think a thought). Flusser takes a step further by saying that from now on we shall have to live with –

> "… the digital world picture as it is being suggested to us by the sciences and presented to our eyes by computers […] whether we like it or not [… and] this imposes on us not only a new ontology, but also a new anthropology. We have to understand ourselves – our 'self' – as such a 'digital distribution', as a realisation of possibilities thanks to dense distribution […] This new anthropology, going back to Judeo-Christianity, where humans were conceived of as mere dust" (Flusser, *op. cit.*).

Here, it should be noted that the Christian tradition claims that a human *body* is dust and ashes, as in the context of decomposition, and not to do with the human soul. Yet, leaving the religious tradition aside, why on earth *should* our self be merely a "digital distribution" from the contemporary sciences' point of view? Isomorphism between computer and human consciousness is merely a certain unclear analogy that has not paid off well so far, except in completely determined cognitive activities: predominantly those more directly associated with rational, analytical thought. In short, the digital analogy is concerned with a certain *world picture* and not with the reality "itself". Nevertheless, there is a danger that we will only see this image once the living spirit has been almost completely concealed.

In his description and analysis of the digital image of the contemporary world, Vilém Flusser in many respects comes close to Heidegger's philosophical legacy, especially in his writings on technology in terms of "Enframing" <*Ge-stell*>, but Flusser's attitude towards this fatal "coming-to-pass of being" is – as stated above – different to Heidegger's: Flusser is an optimist regarding the development of technology and the digitalisation of the world. In his essay *Digital Apparition*, he says that "we are no longer the objects of a given objective world [i.e., no longer

subjected to it], but projects of alternative worlds. From the submissive position of subjection we have arisen into projection. We grow up. We know that we dream" (Flusser, *op. cit.*). With digitalisation and virtualisation of the world, the modern period of subjectivity is supposedly coming to an end with the transformation of "subject into project" (*ibid.*) and thus, also bringing an end to the period of the Galilean submission of a subject to an object. Flusser concludes his essay in a rather prophet-like conceitedness: "'Digital apparition' is the light that illuminates for us the night of the yawning emptiness. We ourselves, then, are the spotlights that project the alternative worlds against the nothingness and into the nothingness" (*ibid.*). Nevertheless, we cannot avoid the question of whether that very "projection into the nothingness" is not precisely the most accomplished form of nihilism of modern subjectivity, which Heidegger conceived in terms of "oblivion of being." He saw it as extreme "danger" brought about by modern technology: the "Enframing" of all being as merely available, left to the boundless autocracy of the human subject. Yet, Flusser is aware that the cybernetics increasingly covering science and art would eventually implement Nietzsche's maxim "Art is better than truth," thus establishing the primacy of "oughts" before facts. However, Flusser optimistically embraces this possibility also when he claims, for example in his essay *On Projection*, "that this opens up the field of projecting alternative worlds and people" (Flusser, *op. cit.*). Supposedly, this means that new energies would be released that would enable our civilisation to overcome decomposition and disintegration of traditional structures and values. But Flusser is rather gullible in believing – as stated in his essay *Conceiving of Technology* – that "future technology shall not change neither the world nor an individual, but shall make sense of life in front of the absurd and death" (*op. cit.*); and when he says, in the essay *On Projection*, that we are now "at the crucial point of gaining 'poetical' freedom."

Is it truly so as Flusser claims? After two decades of development of cyber-technology this "gaining freedom" increasingly reveals itself as a new form of slavery, or it is at least a serious *possibility* and danger of some new, very aggravating slavery. – Of course, not even Flusser was naive with his cyber-optimism. In his essays, he critically writes about, for example, contemporary "amphitheatres" in which we ourselves, without being forced from the outside, indulge in the capital or market domination of "techno-images" (Flusser, *op. cit.*). His essay *On Projection* has an ambiguous conclusion, claiming that "we begin – in fleeting moments of insight – to rise up from submission to conception fully aware of the fact of how uncomfortable, dangerous and unpromising this adventure might be. Is this optimism?" (*op. cit.*). Without a doubt, we may now add to those sceptical tones that our cyber "adventure" does show that technology is becoming more and more promising, but at the same time, it is also becoming more dangerous for our human existence. [1]

[1] In the Slovenian philosophical and broader cultural space, it was Janez Strehovec, philosopher, who wrote about Vilém Flusser and generally on cyber-technology and art in his books *Virtualni svetovi* [*Virtual Worlds*], 1994, and *Tehnokultura – kultura tehna* [*Techno-culture – the Culture of Techno*], 1998. Strehovec considerably followed Flusser's cyber optimism when he claimed: "An affirmative attitude towards technology also presupposes the overcoming of imperialistic domination over nature that was characteristic of traditional industrial societies, and classifies us into the paradigm of reconciliation between society, nature and technology; more precisely, it is the symbiosis of nature and society through technology, which is user-friendly and takes account of nature" (Strehovec 1998, p. 95, translated by M. P.).

During the time of the first great wave of progressive virtual reality, the renowned philosopher and sociologist Jean Baudrillard (1929-2007) was very critical towards it. In his now already classic book *Simulacra and Simulation* (1981), he quotes, as a motto, his own simulacrum of a sentence from the biblical *Ecclesiastes* (the Preacher): "The simulacrum is never what hides the truth – it is the truth that hides the fact that there is none. The simulacrum is true" (translation quoted from: Baudrillard 1999). Simulation constantly generates simulacra, but now "[s]imulation is no longer that of a territory, a referential being, or a substance. It is the generation by models of a real without origin or reality: a hyperreal" (*ibid.*) This means that –

> "[i]t is all of metaphysics that is lost. No more mirror of being and appearances, of the real and its concept. […] It is a hyperreal, produced from a radiating synthesis of combinatory models in a hyperspace without atmosphere […] It is a question of substituting the signs of the real for the real" (Baudrillard, *op. cit.*).

– And in this kingdom of signs of the real that is supposed to veil completely the real itself, lies the source of Baudrillard's fascination and at also of his "world-weariness," his postmodern *malaise général* that is present in his sharp irony, which refers to everything in the contemporary world, in a world where "for a long time the real has been absorbed in film (or television) hyperrealism" (*ibid.*). The hyperreality of simulacra is supposedly beyond the difference between true and false, between real and imaginary. He wonders: "Since the simulator produces 'true' symptoms, is he or she ill or not?" (*ibid.*). In any case, in this hyperreal "simulation" there is indeed something ill, a symptom of which is also Baudrillard's melancholy, which radiates from almost any of his pages. With the world he paints, there is something *generally* wrong, evil prevails in it, even if it is only virtually "simulated." (I guess that among "seven capital sins" of our time is also this all-embracing pessimism, a radical unhappiness with everything there is). – At the beginning of his later and best known book *The Perfect Crime* (1995), Baudrillard writes:

> "This is the story of a crime – of the murder of reality. And the extermination of an illusion – the vital illusion, the radical illusion of the world. The real does not disappear into illusion; it is illusion that disappears into integral reality. […] Alas, the crime is never perfect. Moreover, in this grim record of disappearance of the real, it has not been possible to pin down either the motives or the perpetrators, and the corpse of the real itself has never been found." (English translation quoted from: Baudrillard 2008, xi)

He then continues to kill the reality with the following words: "Reality is a bitch. And that is hardly surprising, since it is the product of stupidity's fornication with the spirit/mind of calculation – the dregs of the sacred illusion offered up to the jackals of science" (*ibid.*, 3); and further on: "On the horizon of simulation, not only has the world disappeared but the very question of its existence can no longer be posed" (*ibid.*, 5) (which is in fact *contradictio in adiecto* in relation to Baudrillard's own questioning in his book), for "we live in a world where the highest function of the sign is to make reality disappear and, at the same time, to mask that disappearance" (*ibid.*) while also "[a]rt today does the same" (*ibid.*) etc. – in the presence of all these catastrophic claims, we may shrug our shoulders and sigh in the face of such boundless

melancholy, but we may also protest: Baudrillard's "perfect crime" theory has a flaw in its own postulation (among all other things). Namely, in the very formulation of this imperfect perfection there is a simple, obvious flaw: in spite of all appearances, in spite of all simulacra, or "traces of nothingness," which surround me in the contemporary world, I am still here as *my-self*, i.e., my consciousness is still here, my "openness" to the world. I can still have a look at this tragicomic scene "from behind the scenes." And even when I am no longer here, there will be other people, my "other selves." We come into the world and we leave the world in order to make space for those who are coming behind us. Baudrillard, who called the erasure of the difference between virtuality and reality the "perfect crime" has, in my view, overlooked or not taken sufficiently into account that which is essential to reality: *the reality of consciousness, the soul, the mind*. It is precisely because of the mind (or, even better, of the spirit) that this crime over reality cannot possibly ever be perfect.

* * *

Now, in the "real time" of the first and the second decade of the new century, there passes through us (also "across" and "below" us) the *second* great wave of cyber virtualisation of reality: the sprawl and globalisation of the internet as the other and, at the same time, the same "parallel world" with countless virtual spaces. *Google* is considered to be the biggest machine made by the human race.[2] Since *Google*'s "search machine" was first released to the global public in 2004 with the purpose "to organize the world's information and make it universally accessible and useful" (see *Google Corporate Information*), we have already gotten so used to it that we "ask" the machine this and that, we read using the machine, correspond, virtually socialise, etc.; therefore, we can no longer imagine a time when "he" (or "she"?) was not here. And now there follows an astonishing question: what if one beautiful day Mr. (or Mrs.) *Google* "woke up", what if He or She became aware of Itself, what if It became a Person of some sort? This question, of course, not only sounds futuristic, but also science-fictional. Nevertheless – is this really so impossible? Is it definitely true that somebody like Mr. or Mrs. *Google* introducing himself or herself to me and talking to me like my "next of kin" does could never happen? A sceptic, of course, would think this is not possible, claiming that *even if Google* talked to me like my wife does, this would not mean that it is a conscious being, a person with self-awareness and its own will. However, this is no comfort, particularly if the awoken machine wanted something from me. Let us say that Mr. *Google* wanted something unexpected; for example, if He wanted me to leave my house immediately, because He knew there would be an earthquake in five minutes… Would I believe "Him"? Or worse still, would I believe that the almighty *Google* has no evil intentions? For although the basic maxim of *Google's Code of Conduct* is "Don't be evil!", who can guarantee me/us that it would follow this maxim even when awoken into a person with his or her own consciousness and will, perhaps even feelings (so much feelings already circulate in the electronic veins)? – If *Google* (or any other internet giant) "woke up," the futurists Vernor S. Vinge, Ray Kurzweil, James D. Miller and others would say that *singularity* has come (and if more of them woke up, we would probably, judging

[2] See for example (Bostrom 2014, 16).

by the human world, eventually see one of them prevail over the others).[3] Yet, the "singularity" would not necessarily be mean to the human race. As I prefer to be optimistic, I assume that there are even more chances for it to be good, let us say in the style of some cyber "good demon". Of course, all this is purely hypothetical, and I personally do not even believe that such a "singularity" will take place in this or the next few centuries, even less during my/our lives – but, for example, in Galileo's time, no one seriously thought (not even Galileo) that a human being, except for the legendary Menippus "Above the Clouds" in Lucian's satires, could *really* (i.e. in its earthly body) make a trip to the Moon. Anyway, *if* in the future, the computers' "awakening" takes place, that will mean the arrival of the *third* wave of virtual reality, much bigger than the first two, and it could wipe out the species *homo sapiens* from the face of the Earth. Or, in a slightly better scenario for us, our species would be "reset" or upgraded to "super-humans," or cyborgs who would be called, let's say *sapiens 2.0* (Nietzsche would turn over in his grave). Let us briefly look at how these scenarios are imagined by futurists…

Vernor Steffen Vinge (born in 1944), mathematician, computer scientist and writer of science-fiction books, is best known as a futurist with his article *The Coming Technological Singularity* (1993, a revised version is available on the internet, 2010). At the very beginning of his article he boldly states: "Within thirty years we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended." There follows a question: "Is such progress avoidable? If not to be avoided, can events be guided so that we may survive?" (Vinge 2010, 1). The starting point of this article is that we are faced with a change "comparable to the rise of human life on Earth" (*ibid.*), the cause of this change being "the imminent creation by technology of entities with greater than human intelligence" (*ibid.*). Vinge defines "singularity" as follows:

> *Singularity* is "a point where our old models must be discarded and a new reality rules. As we move closer to this point it will loom vaster and vaster over human affairs till the notion becomes a commonplace. Yet when it finally happens it may still be a great surprise and a greater unknown." (Vinge 2010, 2)

Futuristic technological and historical singularity is thought of in double analogy with physics: this Event is supposed to be *singular* in the sense of being torn out of our historical space and time; and secondly, it is *hidden* for us since it is now (still) beyond our experienced "event horizon". However, there seem to already be indicators according to Vinge that "give the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue" (*ibid.*): supposedly, a new era of "transhumanism" shall appear. Vinge has seen the signs of this singular change primarily in the ever accelerating development of computer science and nano- and biotechnology in the last few decades – if this trend is to continue with such a rapid pace in the future, the singularity shall

[3] The term 'singularity' comes from mathematics and physics. In mathematics, singularity is a point (or a surface) where a function is not defined (otherwise said, where it is not differentiable). Thus, for example, reciprocal function $y = 1/x$ yields a singularity for the value $x = 0$, since at this value $y = \pm\infty$. Singularity in physics is 'a black hole' in spacetime; however, considering black holes, we have to distinguish between a black hole as *an area of spacetime* beyond the 'event horizon,' which is not accessible from the outside and, singularity itself, in the middle of the black hole, which is merely a *hypothetical* point, a theoretical construct (i.e., the consequence of the General Theory of Relativity). – The futuristic use of the term 'singularity' is of course merely analogical.

be unavoidable as early as in this century. Vinge also draws the most plausible scenarios of the emergence of singularity, which may be summed up in three points:

1. computers are "awake" in terms of being conscious beings, and maybe also possess their own free will (i.e. the emergence of "strong artificial intelligence"); the systemic variant of this awakening has already been mentioned: (1a) large computer *networks*, along with their human "partners," wake up as superhumanly intelligent entities;
2. interfaces or connections or joints among computers and people (created by the connections of info+bio+nanotechnologies) become so intimate that people become *cyborgs* and acquire superhuman intelligence; one of the pioneers of this idea is the English professor of cybernetics, Kevin Warwick, author of "Project Cyborg" and the book *I, Cyborg* (2004);
3. biological science provides the means to enhance natural human intelligence, e.g. with genetic engineering and similar.

Vinge is convinced that the possibilities of (2) and/or (3), which he sums up with the term "intelligence amplification," are more likely than the possibilities of (1) and (1a), claiming that "it's very likely that intelligence amplification is a much easier road to the achievement of superhumanity than pure artificial intelligence" (Vinge 2010, 5), since something similar happened during natural evolution in the transition between non-human primates and humans. However, he assumes that the accelerated scenario of awoken "superintelligence" (perhaps awoken in an instant) is more likely than its gradual awakening. But the key question we may ask here is whether people will be able to have an essential influence on the transition into "technological singularity": are we able to prepare for it and direct its emergence according to our interests and wishes? Here, Vinge is a moderate optimist, since he hopes that we have at least "the freedom to establish initial conditions, make things happen in ways that are less inimical than others" (*ibid.*). Nevertheless, a key ethical and axiological question still remains unanswered: which objectives, goals or "maxims" should be involved in the "initial conditions" of superintelligence, if we could (co)create them before its awakening? The main ethical a priori principle would probably remain the same as the one inscribed in *Google's* code of conduct: "Don't be evil!" – but would the dividing line between good and evil drawn by superintelligence be still the same, or at least similar, as ours? Although Vinge puts immortality among his most highly appreciated motives of "strong superhumanity" (*ibid.*, 7), he realistically finds: "In fact, I think that the new era is simply too different to fit into the classical frame of good and evil" (*ibid.*). But in spite of all the unknowns, he concludes this prominent article in an optimistic way, almost theologically enthusiastic: "… and while mind and self will be vastly more labile than in the past, much of what we value (knowledge, memory, thought) need never be lost" (*ibid.*, 8).

Ray(mond) Kurzweil (born in 1948), a controversial theorist on cyber-technology and futurologist is the most prominent contemporary prophet on the near technological singularity and emergence of "transhumanism".[4] His most renowned book is *The Singularity Is Near: When Humans Transcend Biology* (2005). In it, he claims:

[4] Kurzweil's parents were Jewish immigrants from Austria. Ray passed as a *wunderkind*, studied at the prominent MIT and became famous as a computer scientist and inventor. To name only a few of his main innovations: optical character recognition, text-to-speech synthesizer, print-to-speech reading machine for the blind (his friend was

> "…we have already succeeded in modelling portions of our brain-neurons and substantial neural regions and the complexity of such models is growing rapidly. Our progress in reverse engineering the human brain, a key issue that I will describe in detail in this book, demonstrates that we do indeed have the ability to understand, to model, and to extend our own intelligence. This is one aspect of the uniqueness of our species: our intelligence is just sufficiently above the critical threshold necessary for us to scale our own ability to unrestricted heights of creative power and we have the opposable appendage (our thumbs) necessary to manipulate the universe to our will." (Kurzweil 2005, 21)

These words are highly optimistic and probably also presumptuous, but Kurzweil uses them just to reiterate, in his own radical way, the conviction that has been alive for decades among the followers of the "strong artificial intelligence" (i.e. computerised consciousness, self-consciousness, will, ... person): the conviction that it is possible artificially to create intelligence that is not only equal to the intelligence of humans, but surpasses it. If Kurzweil spoke only in favour of this conviction, he probably would not be so criticised by numerous eminent proponents of artificial intelligence. But: *if* strong artificial intelligence is possible, would that not be *eo ipso* "singularity" in Vinge's and/or Kurzweil's sense? Kurzweil, in his dramatic way only expresses the expectations of many computer scientists that it will be possible to create an "artificial mind". However, if we do not believe in this eventuality, we may find in Kurzweil's radicalisation of these expectations a kind of *reductio ad absurdum* of such projects. Anyway, his chronological announcement that singularity will take place in "real time" during the current century is surely questionable. Namely, in his book *Singularity Is Near* (2005), he predicted that:

- in 2029 a computer shall successfully pass the Turing test, which means it will become equal to human intelligence;
- in 2030's the boundary dividing humans and computers shall progressively become erased (numerous cyborgs, etc.);
- in 2045 Singularity will emerge, implying the end of the human era, the beginning of "transhumanism" and the emergence of a "new species": *sapiens 2.0*, and possibly later, *3.0*. To our last generation of humans, Kurzweil recommends the use of cryonics (low-temperature preservation of bodies) to overcome the time leading to "real immortality";
- around 2200, the universe will become a "gigantic supercomputer".

So what are the arguments for such a dramatic chronology of forthcoming events? In his books (and in his lectures and films), Kurzweil shows numerous statistical diagrams demonstrating the exponential growth of key technological factors in the last few decades, e.g. in the

---

blind singer Stevie Wonder), he invented a new type of music synthesizer, discovered new methods of machine pattern recognition, excelled in the field of computerised medical diagnostic systems, etc. – He wrote a series of notable, yet criticised books, the most prominent of them being: *The Age of Intelligent Machines* (1990), *The Age of Spiritual Machines* (1999), *Singularity is Near* (2005), *How to Create a Mind: the Secret of Human Thought Revealed* (2012). He cooperates with *Google* and with NASA, and he has established his own *Singularity Institute for Artificial Intelligence* (2004). In the previous century, he supposedly predicted the end of chess, the collapse of the Soviet Union and the expansion of the Internet; however, these 'prophesies' were questionable since they were quite rational predictions rather than any sort of clairvoyance.

development of computer hardware, knowledge in the area of biotechnology and similar. He expressly refers to "Moore's law" (Gordon Moore, 1965), which says that the ability of computer processors doubles approximately every two years. Yet, this has been true for a few decades, but the further validity of this "law" depends on the introduction of new hardware technologies (i.e., if cellular and quantum computers are realisable and similarly), since there are technological boundaries of classical hardware "architecture". Furthermore, statistical extrapolation to predict future events is often completely wrong, as it is hard to determine all the relevant indicators of future development. Also, more "intuitive" futuristic predictions often turn out to be false or incomplete because events frequently happen in a much different way than predicted.[5] Regardless of the validity of Moore's law in the future, it is not clear why and how the growth of the quantity of hardware (and software, in parallel) shall bring about new quality, i.e. strong artificial intelligence: computer consciousness, machine "soul" and/or "mind".[6]

* * *

The species *homo sapiens* is around 200 000 years old, but the historical humankind, only around 6000 years (let us say so, from the Sumerians onwards). Will human culture last for a further 6000 years and humankind another 200 000 years? Will we be able to last that long? Will "the Fates" allow it? Obviously, there were relevant "thresholds" in the development of the world, life and consciousness that needed to be overcome, or "filters" through which we needed to pass so that evolution – naturally driven or "designed" (in *some* way) – could make progress. Here, I do not attempt to tackle the question of whether overcoming such thresholds is the work of nature or of God's plans, nor do I provoke any metaphysical deliberation about whether these are merely turns in continuous development or some leaps "from quantity into (a new) quality". I simply note that the three following thresholds are still not adequately understood: the beginning of the world/universe (the "Big Bang"), the origin of life and the emergence of consciousness, i.e. thinking and self-awareness of *homo sapiens*. After those "great three" come several smaller, but for our development still very relevant thresholds: among them are the beginning of farming, the invention of the wheel and of writing, the Copernican Revolution and the first Moon landing and, last but not least, the invention of computers, of "thinking machines."

Nick Bostrom (born in 1973), philosopher at the University of Oxford and lucid analytical futurist, ponders in his latest book titled *Superintelligence: Paths, Dangers, Strategies* (2014) whether the "awakened" computers – of course, *if* they ever truly wake up (Bostrom feels this is quite likely) – might be quite alien to human mentality, maybe even hostile in the way that

[5] Many a futuristic expectations from decades ago, for example those from the famous science-fiction film *2001: A Space Odyssey* did not come true: in 2001 we were far from space flights with human crew to Jupiter's moons, yet, unfortunately, something completely different happened: the fall of the Twin-towers in New York. Orwell's negative utopia *1984* also did not come true as predicted by the writer in 1948, although we might say that now, unfortunately, it is coming true – in a *different* way. (If we take into account "the difference in sameness" of futurology, also Kurzweils' predictions are very interesting.)

[6] Eschatological expectations that may come true with the development of computers or "artificial intelligence" in view of 'singularity' are particularly bizarre. On contemporary cyber-eschatology, see social and anthropological study (Geraci 2010); I write more about that in (Uršič 2015), especially about the web portal SECOND LIFE.

their motives of operation would be quite different from our human motives, so that they might, sooner or later, attempt to "get rid of us" in the same way that computer HAL tried to murder the whole human crew of the spaceship Discovery in the film *2001: A Space Odyssey*. Perhaps people will not even know what the awakened machines are thinking before it is already too late? For this reason, Bostrom is convinced that we need to prepare ourselves for the possible arrival of "superintelligence" as much and as extensively as we possibly can.

In the middle section of his book, Bostrom analytically deals with the question of how quickly, i.e. in what length of time interval there will (when it will or *if* it will) the birth of superintelligence come: slowly, moderately or very quickly? This, of course, is also one of the elements on which the possibility of human control or at least influence on the "explosion of intelligence" depends. Bostrom assumes that a very speedy "take-off" – "over some short temporal interval, such as minutes, hours, or days" (*ibid*., 64) is more likely – and in the case of such a scenario, humanity's fate would depend, above all, on preliminary preparations for the event.

While pondering the real chance of superintelligence, the key question is what a "superintelligent factor" would be *capable* of doing and at the same time what it would *want* to do, what His (or Her) *motives* would be, what would its "final cause" (*télos)* be – and to what extent could people determine such motives in advance. Would it be possible to programme a value system into superintelligence, or a code of ethics? To this probably legitimate concern, and regardless of how computers shall develop in the future, I need to make a remark that it would be beneficial if we could first provide an answer to the more basic question: does any computer (or network and similar) can have *its own free will*? Can it execute an act *autonomously*, without some human "order"? However, it is even more difficult to answer this question than the question of whether a computer can have its own consciousness.

Bostrom's book *Superintelligence* is a smartly-conceived warning and guide about the need to at least start *thinking* about questions associated with the control of all contemporary high-technology, not only of nuclear technology and biotechnology, but also of computer and digital technology in the widest sense. The critical analysis of the dangers brought about by new cyber-technologies should also direct concrete measures in the planning and production of "thinking machines." In the chapter "Strategic image," Bostrom pleads for the *The principle of differential technological development*:

> "Retard the development of dangerous and harmful technologies, especially ones that raise the level of existential risk; and accelerate the development of beneficial technologies, especially those that reduce the existential risks posed by nature or by other technologies." (Bostrom 2014, 230)

The Oxford philosopher attempts to get through to us that the creation of cyber virtual reality is no longer a game, because now it involves the fate of humankind. "One important parameter is the degree to which the world will manage to coordinate and collaborate in the development of machine intelligence" (*ibid*., 246): we need to work together for the common good. This endeavour to analyse the "strategic" and ethical aspects of the development of artificial intelligence and the informational and other technological sciences need to associate with

humanistics and philosophy, because we need an integrated insight. Bostrom emphasises: it is urgent for the planning of computer science and technology also to include "social epistemology" (*ibid*., 258) – in order that "[b]efore the prospect of an intelligence explosion" we humans shall not be "like small children playing with a bomb" (*ibid*., 259).

Let me conclude this paper with the following question: why modern science seems not only to forget that *mind*, especially mind as (self)consciousness, is the *spiritual* essence of human beings, but often also fervently denies its ontological existence? (*cf*., for example, Dennett 1991). Why it is so hard "among the shadows of the world" to recognize the *reality* of mind, to acknowledge the *evidence* of spirit? To see that there are no "surfaces" without *their* own "depth"? And when we consider the *actual* reality of our world, whether its material or virtual reality, we have at least to admit that the world as we see it might be just the surface of *Reality*, of the ultimate *Mystery*.

REFERENCES

Baudrillard, Jean (1999). *Simulacra and Simulations*, http://www.bconradwilliams.com/files/7313/9690/1991/Baudrillard-Jean-Simulacra-And-Simulation2.pdf (last access: 2.2.2017).

Baudrillard, Jean (2008). *The Perfect Crime*, trans. Chris Turner. London: Verso.

Bostrom, Nick (2014). *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Dennett, Daniel C. (1991). *Consciousness Explained*. New York: Little, Brown and Co.

Flusser, Vilém (2000). *Digital Apparition.* http://web.mit.edu/uricchio/Public/Documents/media-in-transition/Flusser-Digital%20Apparition.pdf (last access: 2.2.2017).

Geraci, Robert M. (2010). *Apocalyptic AI*. Oxford: Oxford University Press.

Heidegger, Martin (1977). *The Question Concerning Technology*. http://simondon.ocular-witness.com/wp-content/uploads/2008/05/question_concerning_technology.pdf (last access: 2.2.2017).

Heidegger, Martin (1962). *The Turning.* https://monoskop.org/images/4/4d/Heidegger_Martin_1962_1977_The_Turning.pdf (last access: 2.2.2017).

Kurzweil, Ray (2005). *The Singularity Is Near. When Humans Transcend Biology*. New York: Viking, Penguin Group.

Miller, James D. (2012). *Singularity Rising*. Dallas: BenBella Books.

SECOND LIFE: http://secondlife.com/ (last access: 2.2.2017).

Strehovec, Janez (1998). *Tehnokultura – kultura tehna* [*Techno-culture – the Culture of Techno*]. Ljubljana: Študentska založba.

Uršič, Marko (2015). *O sencah* [*On Shadows*]. *Štirje časi – Zima*. Ljubljana: Cankarjeva založba.

Vinge, Vernor (2010). *The Coming Technological Singularity* (1st version 1993). http://mindstalk.net/vinge/vinge-sing.html (last access: 2.2.2017)

Translated by Mateja Petan