Human Core Values in the Digital Age – Challenges towards Philosophy, Ethics and Religion

By Dr. Karl Johannes Lierfeld

Abstract:

Futuristic technonologies like artificial intelligence, robotics and nano technology promise tremendous progress in general and – more specifically – sound solutions for many of humanity´s most urgent problems. But on the other hand these technologies pose inherent control problems and challenge us to define new ethical standards and to regulate the adoption of said technologies.

Within an unknown timeframe, exponential growth and unparalleled rates of progress will inevitably lead to a form of non-biological intelligence that will first achieve parity with humanity, then by far exceed our capabilities. By this time, however unclear the exact date remains, emulation technologies will allow not only to map the human brain but eventually also to recreate a virtual copy of the mind and implement it in a simulated environment or on a hard drive. Humanoid robots will be integrated into society and subsequently granted basic civil rights; in fact, the humanoid conversational robot **SOPHIA** already gained civil rights in Saudi-Arabia in 2017. This is especially precarious from an ethical vintage point because this robot has obtained more rights than most of the Saudi-Arabian women possess.

Human beings will enhance their brains by the use of machine-brain-interfaces like Elon Musk´s project **NEURALINK** promises. Humans augmented with such devices will possess superhuman capabilities and an IQ that will grow multiple times greater than that of the brightest biological brains. How should we regulate who is allowed to augment their brain and who is not? How can we guarantee a fair competition if augmented minds float the job market? And how should such augmented human beings should be called, in the first place?

The upcoming revolutions are indeed more complex and challenging than any other technological revolution before, and they demand our focussed, interdisciplinary work and unified rationality to make sure the outcome is humanistically appropriate. That also means to bound forces of the world´s most eminent thinkers and leading minds, from science to politics to

economics, but of course also the religious lead figures must engage in this multi-level challenge.

In order to create machines that act ethically, it appears obvious that the first pre-requisite would be a canon of core values which we then translate to the machine intelligence. The second part of the problem is not remotely as hard to solve as the first part, and this is because we as a species failed for thousands of years to obey to a mutual moral code. The cultures and philosophies appear to be too different to unify under one banner of thinking.

The big problem, however, is: we need to find this moral / ethical consensus *before* an intelligent agent does it for us. This consensus cannot pay attention to cultural diversities as it has to be implemented on a global scale. Hence such a moral code has to be abstract in a way that it transcends cultural differences, but tangible in a way that it expresses clear and non-ambivalent operational rules. Which leads us directly to the underlying, essential communication problem: we don´t even know a non-ambivalent terminology (except from code). Isaac Asimov´s famous 3 robot laws serve as a sound example for this dilemma:

1. *„A robot may not injure a human being or, through inaction, allow a human being to come to harm."*

2. *„A robot must obey orders given it my human beings except where such orders would conflict with the First Law."*

3. *„A robot must protect ist own existence as long as such protection does not conflict with the First or Second Law."*

While Asimov surely achieved to create compact laws with a coherent logic supporting them, his robot laws are deemed to fail due to our ambivalent and context-dependent language. We have too much room left for interpretation. What is meant by „injure a human being"? Or what does it mean to „come to harm"?

It depends on the context. A masochist who books a sex robot to get tortured by the robot could even demand that he got harmed by not getting harmed, as demanded. Human communication is too broad and works on too many layers to fit to machine intelligence yet, and we can´t expect machines to understand us right if we don´t teach it to them.

The following shows a list of terms that we need to (re)define in order to make our language as well as our philosophy future-proof:

- human consciousness / other forms of consciousness
- mind
- identity
- free will
- intelligence
- rationality
- morality
- ethics
- reality
- perception
- qualia
- human being
- humanity
- augmented human beings
- robotic entities
- uploaded memories and minds etc.
- equality / inequality

The main challenge is here to overcome a centuries old habit of developing different – and partly opposing – schools of thinking and to constantly ponder and formulate alternatives; in contrast, we need to reduce some complexity and create mutually accepted definitions that don´t leave any room for interpretation to a machine intelligence.

But it won´t be done by sharpening our communication tools. Once we have found a non-ambivalent terminology, we still need to find consensus about how an intelligent agent should act, ethically spoken.

Challenges for traditional religions:

The ongoing and upcoming technological revolutions pose unique challenges towards traditional religions. This is partly because of the sketched changes of the concept of man, but also due to the ontological shifts of paradigm that could come with the singularity or any form of intelligence explosion.

First of all we need to define what a human being essentially is and what the concept of humanity should include in the future. Since a humanoid embodiment isn´t necessary in the realm of simulation technology, uploaded human minds should still count as human entities, equipped with rights and dignity. Such technologically created entities would deserve to be protected, just as humans do. Essentially we are in need of an extended concept of man that isn´t exclusively bound to biology a priori.

Neglecting these issues would inevitably result in forms of mind crimes, situations where relevant forms of consciousness would be harmed while these factual right violations been ignored. Even if a „Whole Brain Emulation" will never happen, these foreseeable ethical issues remain, simply because many AIs, including conversational bots, are black box systems which aren´t fully understood by their developers. Hence we will enter a state where sentience could be simulated with such virtue and credibility that we couldn´t deny the existence of some form of consciousness with any sufficient decisiveness. Even more precarious, we couldn´t decide this question anymore. And if we can´t distinct between sentient and non-sentient systems anymore, we will have to act *in dubio pro reo* to remain ethically appropriate. In other words: if technology becomes ambivalent in terms of consciousness and qualia, and we can´t deny sentience by design, then we will have to act as it was proven that given AI is sentient.

Transhumanism:

As a technologically driven movement of futurists, transhumanism strives to transform human life, eventually resulting in the merging (or even successive replacement) of biological life through technology. In an official declaration from 2013, the Catholic Church has widely rejected the concepts and endeavours of transhumanism as a reduction and devaluation of human life. Through its emphasis upon technology and the augmentation of the human biology, the concepts of transhumanism would degrade life in its original forms.

The elephant in the room is, however, the ethical issue that stems from a general rejection of transhumanist technology. The example of advanced nano medicine highlights the cascade of ethical problems that arise if we categorically avoid the usage of transhumanist technology. This is because molecular nano technology could ensure affordable global health care, thus enhance the quality of life for billions of people. The **Global Health Care Equivalency (GHCE),** an initiative started by nano scientist Frank Boehm, strives to supply the global market within the next 30 years. As a constructivist technology, molecular nano medicine will use the rearrangement of atoms to create medicine from atom stocks (for very little costs, obviously) or, in form of advanced nano bots, directly rearrange cells on a molecular level.

Transhumanist technologies will rise – no matter if the Church supports or neglects them – and these technologies will save much more lifes than condoms have, for example. It wouldn´t be merely ignorant to try to stop the development of transhumanist technologies like advanced nano medicine or life extension – it would be an ethical issue, since these technologies promise to raise the quality of life in general and battle many diseases in particular.

What we really and urgently need is a regulation of the access to given technologies. While the Global Health Care Equivalency promises to foster equality, the developments could result in an increased inequality if not everyone has equal access to these innovations. At worst, only an elite would benefit from the unparalleled possibilities of molecular nano medicine while a global basic health care is still not established. Therefore, the Church should focus on supporting the rights of the underprivileged and help to ensure that global health care is guaranteed to everyone before more advanced technologies are released to the exclusive circle of billionaires. This will be a problem of ethical economy because it is thinkable that a procedure for life extension, age reversing or brain augmentation could cost a 100 million in the prototype phase, which would only be affordable for a tiny fraction of the population. In return, these people would adopt capabilities that will certainly distinct them from everyone else, of course including the fraction of people that already benefit from molecular nano medicine by then.

One of the most crucial problems will be to regulate the ethical usage of transhumanist technologies. To ban these innovations would only result in avoidable suffering and deaths; rather the challenge will be to implement these technologies in ethically appropriate ways and for the benefit of all people – not only for the privileged.

The Singularity as a „materialist version of the Rapture“:

Many highly diverse notions, hopes and anticipations are connected with the opaque, still hypothethical, yet not a priori unlikely concept of the awaited Singularity. While it is still unclear what the Singularity will be, it appears to be common sense that it will result in a game-changing shift of paradigm which will change the course of humanity forever.

One key aspect of the Singularity is the transhumanist idea of extending and subsequently transcending biological life through the means of technology. Conservative commentator Wesley Smith coined this aspect the „materialist version of the Rapture“: humans use technology to overcome the limitations of their physical bodies in an attempt to achieve immortality.

Regarding the christian doctrine, the impact of the Singularity would mostly affect the concept of the Imago Dei. This is simply because if man is really made in God´s image, how was mankind enabled to create machines that adopted their own forms of reasoning? Rather than a devaluation of human life, the Singularity could be interpreted as an elevation of humanity and the skills we have acquired. Transcending our biology and developing strong, autonomous reasoning machines must not necessarily stand in conflict with the concept of the Imago Dei, because if mankind is made in the image of God and these machines will be made in the image of us, then we would have rather created a technological version of the Imago Dei that is embedded in secular realms.

The two phases of transformation that lead to the Second Coming are transfiguration and resurrection. The Singularity resembles the phase of transfiguration by altering humanity´s physical and spiritual conditions. Essentially, a prolonged life span achieved through methods of life extension could make life sustainable enough to eventually last until the Second Coming of Christ. It is highly debatable whether a prolonged biological mind would be equal with an (if ever possible) uploaded mind.

The Simulation Argument / Simulation Hypothesis:

It is common sense that we don´t have access to reality itself, or with Kant, the things themselves; what we perceive as reality is merely the inner simulation of the outside world, created by this yet unknown personalized virtual reality which is also our consciousness. Everything is transmitted by our five senses and processed within our brain – so couldn´t it be tricked?

The underlying doubt in the reality of the real, the truthfulness of our perception, is centuries old and has been most famously described by Rene Descartes´ *evil demon* who interrupts our connections with the objective outside world to manipulate us.

A contemporary version of this ontological doubt is the simulation argument (and any corresponding theories). 2001 – two years after the release of the iconic *simulation blockbuster* THE MATRIX – Oxford philosopher Nick Bostrom attempted to profoundly shatter the ground of our reality. Bostrom´s turned-to-be-famous *simulation argument* takes the premise of THE MATRIX disturbingly literally and uses probability deductions to present three possible versions with very tendencious probabilities. The fundamental threat of this thought experiment is rooted in its undeniable logic:

1. The fraction of posthuman civilisations that reach a posthuman stage is very close to zero.
2. The fraction of posthuman civilisations that are interested in running ancestor simulations is very close to zero.
3. The fraction of all people with our kind of experiences that are living in a simulation is very close to one. (Bostrom, Nick: Simulation Hypothesis, Oxford, 2001)

Bostrom´s hypothesis provoked a passionately led debate about the very foundations of what we really know. It re-formulates a well-known threat to the authenticity and general value of human experience: that it might be all illusional. As the incoming information is all coded and decoded those information could in fact be the input a holistic computer simulation could transfer to the individual.

The circumstance that we can´t neglect whether the premise nor the development of the argument opens a horizon with the unlikely, but possible option that the simulation could be our reality. Abstract philosophers like Baudrillard have had their influence on the simulation debate, coining the terms *simulation*, *dissimulation* and *hyper reality* already back in the 1970s, and he created the metaphor „desert of the real“. Indeed, future simulations will make the real look grey and hopeless, a danger that could result in forms of collective escapisms from reality.

The simulation argument illustrates how insecure we are about perception, experience, qualia and reality in general. The fact alone that we can´t deny it a priori makes it worth a second look as a fascinating thought experiment. In 2016, the Bank of America announced that the likeliness to live inside a collective computer simulation lies between 10 and 30 %, and a team of scientists has been hired to hack us out of that false reality. Rational thinking, however, reminds us that there is no possibility to proof the simulation thesis from *inside* of the simulation, so the thought experiment is heuristically pointless.

A danger of advanced AI, however, would arise if an AI would establish its goal to find out whether reality is simulated or not. If this was the final goal, the AI could waste all available resources and turn all raw material into programmable matter, transforming *the desert of the real* into *computronium*.

The simulation argument poses the question of god in a contemporary way, shifting the focus into the realm of – yet unknown – technology. The designers of a given simulation would take god´s place, while leaving the devine entity still improvable. Hence, the simulation theory cannot be confirmed, as any evidence could also be simulated. Odds are it also can´t be denied as long as we don´t find privileged insights on how our brain destills inner states from outer reality. The simulation argument is ultimately a self-immunizing theory, just like a conspiracy theory. Yet it could establish pseudo-religious movements with strong belief systems that stem from our profound doubts about the nature of reality, consciousness and perception.

A scientific shift of paradigm that could alter all metaphysics

The research for Whole Brain Emulation might result in a scientific shift of paradigm, for instance by determining what happens when the biological brain dies or whether a soul exists or not. Regardless of the heuristic validity such an (alleged) breakthrough would surely have a great deal of impact upon the world religions.

Theoretically, there could be a scientific explanation for the soul. While the input channels of the 5 human senses work in clear cut causality, obeying classical physics, the same doesn´t hold true for basically everything else that takes place in the human consciousness. Every layer of perception (or its processing) that is remotely subjective / subconscious can´t be deduced or explained by classical physical methods. Instead, psychological, emotional or mental states in general seem to follow the opaque rules of quantum physics. Vastly structured by uncertainty, ambivalence and non-distinctiveness, the quantum sphere appears as a contemporary explanation for the subjectivity of consciousness and may eventually also lead towards a scientific concept of the soul.

Research on brain emulation and cerebral architectures might also reveal what happens after death. While this could be merely a contemporary scientific model, it may rise to mainstream popularity regardless its true heuristic potential. Any scientific shift of paradigm could have a game-changing impact upon the religious ecosystem.

It is very imaginable that such a scientific breakthrough – valid or not – could result in the development of religious cult movements or even forms of digital gnostic. And due to the exponential qualities of the digital age, a new world religion could arise within a few years. Therefore the Church should promote critical thinking, modern forms of rationality and future-proof ethics.

The following list gives an overview over many anticipated events that revolve around the singularity and the rise of AI, robotics and nanotechnology:

Event	Probability within 30 years	Prob. within this century
Singularity	low-moderate	High
Superintelligence	Moderate	high
Intelligence Explosion	moderate	high
Scientific shift of paradigm	moderate	high
Establishment of a new technological religion	high	very likely mainstream
Whole Brain Emulation	moderate	high
Brain Augmentation	high	very likely mainstream
Full automation	high	very likely mainstream
Sentient robots	Moderate	high
Robot rights	high	very likely mainstream
Robots, indistinguishable from human beings	moderate	high
Life extention	high	very likely mainstream
Space colonization	moderate	high
AI as a political leader	high	very likely mainstream
Global nano-based healthcare	high	very likely mainstream
Global AI singleton	moderate	still moderate (hopefully)
Simulations that are indistinguishable from reality	high	very likely mainstream
Runaway AI	high	high
AI accident	high	high
AI abuse by humans	high	high
Mind crimes	moderate	high
Molecular nanotechnology	high	very likely mainstream
Global Healthcare Equival.	moderate	high
Genetic enhancement	high	very likely mainstream
AI superorganisms	moderate	high
Self-replicating AI	high	very likely mainstream