Markus Lipowicz

Pedagogical Anthropology as Existential Risk Prevention: A Critical Take on the Techno-Progressive Discourses of Artificial General Intelligence and Moral Enhancement

Studia Paedagogica Ignatiana. Rocznik Wydziału Pedagogicznego Akademii "Ignatianum" w Krakowie 20/1, 73-101

2017

Markus Lipowicz
Jesuit University Ignatianum in Krakow, Poland

Pedagogical Anthropology as Existential Risk Prevention: A Critical Take on the Techno-Progressive Discourses of Artificial General Intelligence and Moral Enhancement

Pedagogiczna antropologia jako prewencja ryzyka egzystencjalnego – krytyczne ujęcie techno-progresywnych dyskursów wokół sztucznej inteligencji i moralnego ulepszenia

KEY WORDS
existential risk, Artificial General Intelligence, moral enhancement, technology, pedagogical anthropology, techno-progressive discourse

SŁOWA KLUCZOWE
ryzyko egzystencjalne, Ogólna Sztuczna Inteligencja, ulepszenie moralne, technologia, antropologia pedagogiczna, dyskurs techno-progresywny

SPI Vol. 20, 2017/1
ISSN 2450-5358
e-ISSN 2450-5366
DOI: 10.12775/SPI.2017.1.004

ABSTRACT

The article undertakes the problem of AGI (Artificial General Intelligence) research with reference to Nick Bostrom's concept of existential risk and Ingmar Persson's/Julian Savulescu's proposal of biomedical moral enhancement from a pedagogical-anthropological perspective. A major focus will be put on the absence of pedagogical paradigms within the techno-progressive discourse, which results in a very reduced idea of education and human development. In order to prevent future existential risks, the techno-progressive discourse should at least to some extent refer to the qualitative approaches of humanities. Especially pedagogical anthropology reflects the presupposed and therefore frequently unarticulated images of man within the various scientific disciplines and should hence be recognized as a challenge to the solely quantitative perspective of AGI researches and transhumanism. I will argue that instead of forcing man to adapt physically to artificial devices, as the techno-progressive discourses suggest, the most efficient way of avoiding future existential risks concerning the relationship between mankind and highly advanced technology would be—as John Gray Cox proposes—making AGIs adopt crucial human values, which would integrate their activity into the social interactions of the lifeworld (*Lebenswelt*).

ABSTRAKT

Artykuł podejmuje problematykę badań nad ogólną sztuczną inteligencją (OSI) w odniesieniu do koncepcji ryzyka egzystencjalnego Nicka Bostroma oraz propozycji moralnego ulepszenia człowieka Ingmara Perssona i Juliana Savulescu z perspektywy pedagogiczno-antropologicznej. Główny nacisk zostanie położony na nieobecność paradygmatu pedagogicznego w dyskursie techno-progresywnym, co wiąże się z bardzo ograniczoną ideą edukacji i rozwoju ludzkiego. Aby móc zapobiec przyszłym ryzykom egzystencjalnym, dyskurs techno-progresywny powinien przynajmniej w jakimś stopniu odnieść się do jakościowego podejścia humanistyki. Antropologia pedagogiczna jest tą dziedziną wiedzy, która w szczególności podejmuje refleksję nad z góry przyjmowanymi i przez to często niewyartykułowanymi obrazami człowieka, funkcjonującymi w obrębie rozmaitych dyscyplin naukowych, i to właśnie ona powinna z tego powodu być postrzegana jako wyzwanie dla wyłącznie ilościowej perspektywy badań nad OSI i transhumanizmem. Będę argumentować, iż zamiast zmuszać człowieka do fizycznej adaptacji do sztucznych urządzeń, jak to sugerują dyskursy techno-progresywne, najbardziej skuteczna droga uniknięcia przyszłego ryzyka egzystencjalnego wynikającego z relacji między człowiekiem a wysoko rozwiniętą technologią mogłaby polegać – jak proponuje John Gray Cox – na sprawieniu, aby OSI przyjęła podstawowe wartości humanistyczne, które zintegrowałyby jej aktywność w społecznych interakcjach zachodzących w obrębie świata życia (*Lebenswelt*).

Introduction

Techno-progressive researchers and supporters constantly have to face the objection of referring rather to fantasy and fiction than to reality and science. Indeed, as Philipp von Becker argues, some of the speculations concerning transhumanism, robotics or Artificial Intelligence may—at least at first glance—seem to have lost touch with reality.[1] However, over the past two or three decades, scientific progress has made several things become "everyday reality" which not that long ago used to be legitimately called science fiction (with an emphasis on "fiction"). One should not overlook the fact that the economic and social infrastructures of modern societies are heavily shaped by the newest computer technologies, which not only enrich, but also invade the human body and mind.[2] In that sense the various displays of transhumanism could be seen as very radical extrapolations of the contemporary technological progress, delivering thereby an important image of the commonly shared *Zeitgeist*, which tends to define human condition nearly entirely in technological terms.[3] One way or another, we are heading towards technologically modified posthuman future. The question is not if mankind will change its basic features of existence, but how and to what extent these changes will overcome or even overthrow crucial humanistic values and axiological systems, which have been developed throughout many centuries and have become the legitimate structure of our modern culture.

In this paper I will discuss some aspects concerning the problem of Artificial General Intelligence with reference to the concept of Moral Enhancement from a pedagogical-anthropological perspective. My main hypothesis is the following: the techno-progressive discourse does not incorporate any coherent pedagogical perspective and follows therefore a very reduced and one-dimensional understanding of education, human development and also—quite paradoxically—technological improvement, as well. My pedagogical-anthropological suggestion will therefore be the following: instead of trying

[1] P. von Becker, *Der neue Glaube an die Unsterblichkeit. Transhumanismus, Bioethik und digitaler Kapitalismus*, Wien 2015, p. 18.

[2] Ibidem.

[3] Ibidem.

to keep up the pace with technological progress through biomedical enhancement, i.e. the merging of human beings and technological devices, the opposite direction of making technological devices become more human, might eventually turn out to be more fruitful in the nearest future. However, in order to revert this contemporary development the techno-progressive discourse would have to integrate pedagogical thinking on a fundamental level and hence abandon the solely quantitative perspective and naturalistic ground, on which it stands.

Artificial General Intelligence—the upcoming ultimate Existential Risk?

The term Artificial General Intelligence was introduced for the first time by Ben Goertzel and Cassio Pennachin in the Preface of their 2007 published book *Artificial General Intelligence*.[4] In order to clarify the terms: AI refers to the idea of simulating human intelligence through technological devices. In that sense, as Goertzel underlines, AI research was originally aiming to develop AGI: "AI began in the mid-twentieth century with dreams of artificial general intelligence—of creating programs with the ability to generalize their knowledge across different domains, to reflect on themselves and others, to create fundamental innovations and insights."[5] The only reason for distinguishing between AI ("narrow AI") and AGI ("strong AI") is rooted in the history of the 20th century: in the 70's researchers finally had to realize their incapability to simulate humanlike or over-humanlike general intelligence through computer technology.[6] Hence the idea of artificial intelligence had to be reduced to "the ability to carry out *any particular task* that is typically considered to require significant intelligence in humans."[7] Nevertheless, the general progress of computer science and—in particular—nanotechnology has led to a strong revival of the new/old the dream

[4] B. Goertzel, C. Pennachin, "Preface", in: *Artificial General Intelligence*, eds. B. Goertzel, C. Pennachin, Heidelberg 2006.

[5] B. Goertzel, C. Pennachin, "The Novamate Artificial Intelligence Engine", in: *Artificial General Intelligence*, op. cit., p. 72.

[6] Ibidem.

[7] Ibidem.

of "building AGIs with general capability at the human level and beyond."[8] AGI expresses the concept of developing

> AI systems that possess a reasonable degree of self-understanding and autonomous self-control and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation.[9]

In contrast to the "narrow" or "weak" AI, the AGI researchers aim at creating cybernetic systems which would be capable of learning and—what comes along with that feature—teaching.

In view of the above I would like to highlight the following sentence of Goertzel and Pennachin: "The audience we intend to reach includes the AI community, and also the broader community of scientists and students in related fields such as philosophy, neuroscience, linguistics, psychology, biology, sociology, anthropology and engineering."[10] It is worth noting that although the AGI researchers seek to generate essentially pedagogical machines, they do not intend to reach the pedagogical community in the first place. Should we recognize this—probably unintentional—omission to be a mere coincidence?

We might identify this neglect as pure coincidence if it weren't for the fact that techno-progressive discourse seems to be in general far detached from the mainstream pedagogical discourse, which particularly stresses the non-quantitative and irreducible to any measurable factors characteristics of the human being, i.e. personhood. In her precursory dissertation *Menschenbilder in Erziehungswissenschaft, Neurowissenschaft und Genetik* Katharina Schumann analyzes and interprets the various concepts of man in the framework of various scientific discourses.[11] As Schumann argues, in a vast opposition to the personal image of man drawn by mainstream pedagogy, neuroscience and genetics rather promote a concept of man being a bio-chemical

[8] B. Goertzel, "Superintelligence: Fears, Promises and Potentials", *Journal of Evolution and Technology* 2015, vol. 24(2), p. 56.

[9] B. Goertzel, C. Pennachin, "Preface", op. cit., p. VI.

[10] Ibidem.

[11] K. Schumann, *Menschenbilder in Erziehungswissenschaft, Neurowissenschaft und Genetik. Eine vergleichende Analyse*, Weinheim – Basel 2015.

and measurable being.[12] Moreover, while educational science usually stresses the sociability of the human being, neuroscience and genetics rather concentrate on the individual as a single organism and its body functions.[13] In short: pedagogy—although interdisciplinary by its very nature—generally promotes a qualitative and ontologically exclusive concept of humanity, while the techno-progressive paradigms and movements rather stick to the picture given by robotics, computer science and natural sciences that promote a quantitative and naturalized image of man. From an interdisciplinary perspective, Schumann emphasizes that educational science does refer to the results of neuroscience, while the latter only invokes genetics, whereas genetics in turn do not make references to neither of the previous two.[14] This, as Schumann argues, seems to indicate that educational science has reached a significant level of self-reflection and self-criticism, while neuroscience and genetics have not reached a saturation state of knowledge, which would legitimately dispose them to relate their results to qualitatively different scientific disciplines as—for instance—pedagogy.[15]

Bearing in mind the statements made above, one may embrace the fact that—in contrast to the vast majority of previous AI discourses—the contemporary researches on AGI take into account the social nature of the human mind and intelligence in general.[16] "In reality," as Pennachin and Goertzel admit, "the mind is social—it exists, not in isolated individuals, but in individuals embedded in social and cultural systems."[17] As a matter of fact, AGI researchers are actually forced to incorporate this social approach into their own investigations since their aim is not only to develop a fixed program, which would be able to fulfill specific functions, but the creation of a self-reflective sovereign being that would simulate human self-development. From this perspective, the most salient approach is to enable the machines to interact, compete and cooperate with each other

[12] Ibidem, p. 254.

[13] Ibidem.

[14] Ibidem, p. 82.

[15] Ibidem, p. 251.

[16] C. Pennachin, B. Goertzel, "Contemporary Approaches to Artificial Intelligence", in: *Artificial General Intelligence*, op. cit., p. 24.

[17] Ibidem.

in games[18]. Finally the objective would be to provide the machine's "ability to relate to humans on a mind-to-mind rather than a software-program-to-mind level."[19]

However, this scenario—regardless of how fascinating it might seem to many of us who grew up reading fantasy novels or watching science fiction blockbusters—bears certain risks that have been one of the major subjects within the techno-progressive discourse for the last decade. Probably the most salient and popularized concept is the one by Nick Bostrom considering the development of technologically enhanced super-intelligence as an "existential risk": "An existential risk", as Bostrom argues, "is one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development."[20] Current technological progress, especially future scientific breakthroughs—as the AGI, for instance—unfolds new levels of global threats, "to which we cannot assign precise probabilities through any rigorous statistical or scientific method."[21]

In short: there are many indications that within a few decades we might find ourselves living in a world being void of any form of ontological security—not in the sense of postmodern existential anxieties or identity crisis, but in the very literal meaning of being uncertain whether we will not cease to exist any moment. Needless to say, humanity has endured many forms of natural risk since its very beginnings—"asteroid impacts, supervolcanic eruptions, earthquakes, gamma-ray bursts, and so forth."[22] However, as Ingmar Persson and Julian Savulescu state: insofar as humanity always had a naturally greater ease of destroying and harming than building or doing good, one cannot ignore the fact that mankind's destructive powers will only increase through developing new technologies.[23] Only that through those new artifacts the consequences of human

[18] Ibidem.

[19] Ibidem, p. 27.

[20] N. Bostrom, "Existential Risk Preventing as Global Priority", *Global Policy* 2013, vol. 4(1), p. 15.

[21] Ibidem, p. 16.

[22] Ibidem, p. 15.

[23] I. Persson, J. Savulescu, "Moral Transhumanism", *Journal of Medicine and Philosophy* 2010, vol. 35(6), p. 663.

destructiveness might be irreversible. Eventually, the misuse of new technologies might end up in an irreparable breakdown and hence downfall of mankind.[24] A good example of the irreversibility of existential risks might be the development of autonomous weapons systems. Michał Klincewicz argues that all strategies for minimizing the insecurities deriving from such military computer infrastructures would be "overly optimistic."[25] The only "safe" way would be to remove the human factor completely from the field of warfare—however, that in turn would lead to an ethical controversy concerning moral responsibility and freedom. As a result, Klincewicz argues that the only "safe" option would be not to develop this kind of technology at all, since all our hopes and strategies to reduce the associated risks would remain merely "wishful thinking."[26] Therefore, the key feature of existential risk is the following: contemporary or future manmade disasters might be simply too big to be handled afterwards, *post factum*—they rather have to be prevented.

The problem of emerging technologies is two-fold. First of all, questions can be raised whether mankind has reached the necessary moral state that would allow them to make responsible use of those artificial organs. Secondly one might stress, that through the development of AGIs, i.e. super-intelligent, self-controllable and self-evolving mechanical entities, we might ultimately find ourselves in a situation of being physically and mentally enslaved or erased by our own products. As Ben Goertzel states: many prominent thinkers[27] "publicly raise an alarm regarding the potential that, one day not necessarily that far off, superhuman AIs might emanate from some research lab and literally annihilate the human race."[28] Even if we were to reject the rather discouraging any further debate vision of human extinction, we should notice that this *Matrix*-scenario is not that detached from reality as it might seem in the first place. The affiliation between humans and their

[24] I. Persson, J. Savulescu, *Unfit for the Future: The Need for Moral Enhancement*, Oxford 2012, p. 9.

[25] M. Klincewicz, "Autonomous Weapons Systems, the Frame Problem and Computer Security", *Journal of Military Ethics* 2015, vol. 14(2), pp. 171–173.

[26] Ibidem, p. 173.

[27] Amongst those personages as Elon Musk, Stephen Hawking and Bill Gates.

[28] B. Goertzel, "Superintelligence: Fears, Promises and Potentials", op. cit., p. 56.

technological equipment has ceased to be a simple subject-object relationship. In fact, the statement that the "intricate relationship between technology and pedagogy has not been adequately explored" made by Mabel CPO Okojie, Anthony A. Olinzock, and Tinukwa C. Okojie-Boulder a decade ago, hasn't lost any of its topicality.[29] Given the recent progress of computer technologies and robotics it seems to be quite reasonable to assume that along with the expanding of a "strong AI" the mutual conditioning between man and machine will even become considerably more complex and ambiguous.

The question then arises: how can we reduce the odds of further irresponsible usage of high technology or the forthcoming rise of hostile artificial super-intelligence?

Moral Enhancement instead of Moral Education?

At the very beginning of their widely recognized book *Unfit for the Future*, Ingmar Persson and Julian Savulescu point out that our ethical standards came into existence thousands of years ago, as humanity was living in relatively small societies with a low level of technological knowledge.[30] Throughout the centuries we have therefore learned to be responsible only for the nearest environment and future.[31] In other words: our benevolence and destructiveness both had a local character. The progress of civilization resulted in an overcoming of these limitations by forcing the human being to live in globalized high tech societies. However, our moral standards, as Persson and Savulescu argue, didn't go through a comparable progress. On the contrary, "since the time of Confucius, Buddha or Socrates" human's ethical codes haven't significantly improved—"despite moral education."[32] Consequently, Persson and Savulescu suggest that we "need to speed up the pace of

[29] M. CPO Okojie, A.A. Olinzock, T.C. Okojie-Boulder, "The Pedagogy of Technology Integration", *The Journal of Technology Studies* 2006, vol. 32(2). Available at: <http://scholar.lib.vt.edu/ejournals/JOTS/v32/v32n2/okojie.html> (access: 26.08.2016).

[30] I. Persson, J. Savulescu, *Unfit for the Future: The Need for Moral Enhancement*, op. cit., p. 1.

[31] I. Persson, J. Savulescu, "Moral Transhumanism", op. cit., p. 661.

[32] I. Persson, J. Savulescu, "Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement", *Bioethics* 2013, vol. 27(3), p. 130.

moral improvement urgently to prevent the powerful output of technological progress being misused with catastrophic results."[33]

The statement concerning the necessity of overcoming the fundamental chasm between technological progress on the one hand and moral-ethical stagnation on the other, would hardly cause any controversy if it wasn't for the promoted technique of achieving this goal itself. Instead of moral education, which arguably has rather been quite ineffective in preventing human disasters throughout the past centuries, the two philosophers suggest an alternative form of moral improvement: biomedical enhancement. Since the traditional forms of moral education have proved to be not sufficient, Persson and Savulescu proclaim the necessity to inculcate social norms through biomedical techniques, which should eventually change our human nature.[34] In other words: since humanism was not able to prevent people from falling into inhuman barbarism, the next step could be to overthrow human limitations and promote "moral transhumanism."[35]

Consequently, this approach is posed a considerable problem. As John Harris stresses: the very proposal to "sacrifice freedom for survival" cannot be perceived as an option worth considering, since "that sufficiency to stand is worthless, literally morally bankrupt, without freedom to fall."[36] Moreover, apart from the controversial nature of using biotechnology in order to stimulate a particular direction of human development, the idea of moral enhancement itself seems to be a very inchoate one. Except for the statement on the unquestionable "mismatch" between technological progress and moral development, Persson and Savulescu offer no ethical paradigm or system. Therefore

[33] Ibidem.

[34] I. Persson, J. Savulescu, "Moral Transhumanism", op. cit., p. 667.

[35] These techniques would incorporate, as—from a rather critical point of view—Masahiro Morioka notes, pharmacological and non-pharmacological methods as well, combining altruism-enhancing drugs on the one hand and deep-brain stimulation or genetic manipulation, on the other. In short: according to the supporters of biomedical moral enhancement the only way how humanity could deal with technological progress on a moral dimension would be to overcome the limitations of humanity itself and force the human to become posthuman. M. Morioka, *Why is It Hard for Us to Accept Moral Bioenhancement: Comment on Savulescu's Argument*, Ethics for the Future of Life: Proceedings of the 2012 Uehiro-Carnegie-Oxford Ethics Conference, 2013, p. 97–98; I. Persson, J. Savulescu, "Moral Transhumanism", op. cit., pp. 667–668.

[36] J. Harris, "Moral Enhancement and Freedom", *Bioethics* 2011, vol. 25(2), p. 110.

it is not surprising that Persson's and Savulescu's arguments for moral enhancement—although widely discussed—fall short of convincing even many techno-progressive thinkers, including those working on developing AGI and transhumanists. I would therefore like to refer to the transhumanist philosopher Stefan Lorenz Sorgner, who names two reasons for rejecting the proposal for "moral transhumanism."

The first difficulty concerns the lack of scientific knowledge, which would be indispensable for the realization of biomedical moral enhancement within a short time.[37] If scientists could find a specific "Gene X" responsible for making people embody virtues like justice, freedom, equality and solidarity as universal norms, one could—from a transhumanists' perspective—embrace genetic enhancement as an appropriate tool for moral improvement.[38] But scientist are still far away from detecting such a specific gene.[39] Therefore, as Sorgner indicates, moral bio-enhancement is not, at least for now and the near future, very likeable to be realized. The second difficulty of moral enhancement is the absence of unambiguous, metaphysically grounded ethical criteria for morality, which in turn legitimizes a moral pluralism that couldn't be granted through biomedical means.[40] Given the rather

[37] S.L. Sorgner, "The Stoic Sage 3.0 – A Realistic Goal of Moral (Bio)Enhancement Supporters?", *Journal of Evolution and Technology* 2016, vol. 26(1), p. 88.

[38] Ibidem.

[39] Ibidem, p. 89.

[40] Through referring to Molly Crockett's research on citalopram, Sorgner emphasizes that reduction of aggressive dispositions does not have to be necessarily regarded as a moral progress *per se*. In fact one could argue that certain situations demand from us to be able to inflict harm on other individuals. Sorgner points out the events at the Northwest Airline Flight 253 on Christmas Day, 2009. If Jasper Schuringa—a passenger on the plane—would had been morally "enhanced" through citalopram (or another medication rising the serotonin discharge and reducing, in consequence, the predisposition for violent behavior) he maybe wouldn't have been that eager to stop violently "Underwear Bomber". Yet, Schuringa's violent act can be recognized not only as morally legitimized but morally right. Also the biochemical promotion of pro-social behavior, for example through oxytocin, cannot be seen as a even partial solution to the problem of human destructiveness. Just as not all forms of violence should be preventively suppressed not all kinds of pro-social behavior should be promoted. It might be worth noting that also criminals and terrorist groups promote particular forms of pro-social behavior within their own membership, which—at least from our point of view—cannot be legitimized and should therefore not ne promoted by any means. Ibidem, pp. 84–87.

pessimistic view on the actual feasibility of moral enhancement, which would equalize or at least partially counterbalance the existential threat deriving from the contemporary rise of technologies, one may seriously wonder, whether "we are doomed"?[41]

Sorgner rejects this discouraging view by claiming that morality "is related to the recognition of norms […] that developed during the Enlightenment."[42] It was indeed the modern era of Enlightenment which gave birth to a socially widely accepted moral system promoting the values of freedom and equality as fundamental human rights.[43] A closer historical view seems to indicate that the proceeding social recognition of modern values took place exactly at the same time as natural science and new technologies became relevant features of the Western liberal and democratic civilization. Hence Sorgner argues that there has to be a significant reliance between moral improvement and scientific development.[44] In other

[41] Ibidem, p. 88.

[42] Ibidem, p. 89.

[43] Ibidem.

[44] Ibidem. In order to confirm this standpoint Sorgner relates to the relevant, however very problematic historical analysis of Steven Pinker *The Better Angels of Our Nature*, in which the latter tries to prove the decline of violence throughout the recent centuries. However, from a historical and philosophical point of view the results of Pinker's analysis are highly questionable—to say the least. Even the acknowledgement of the significant progression of modern values like social justice, freedom and equality, cannot dismiss the atrocities and genocides of the 20th century, which in turn make it both cognitively and morally inappropriate to speak about a factual moral improvement of the human species. I believe this is the core of Persson's and Savulescu's argument that for 2,500 years our morals have not fundamentally improved—moreover, I would like to suggest that it is exactly this historical consciousness of inertia and of the cruel heritage of the 20th century which makes it contemporary so hard to believe that the rise of new technologies in the near future will not cause even bigger disasters. In short: even if we reject the idea of moral bio-enhancement for a multitude of reasons—the fear of the forthcoming outburst of human and inhuman violence in the near future can and actually must be recognized as plausible. Pinker's book has been widely commented and criticized as a rather ideological than scientific accurate account on the history of human destructiveness. John Gray stresses that for "liberal humanists"—as Steven Pinker, for instance—"the role of science […] to explain away" the evidence "that humans are violent animals." S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined*, New York 2011; J. Gray, "Delusions of Peace", *Prospect. The Leading Magazine of Ideas*, October Issue 2011. Available at: <http://www.prospectmagazine.co.uk/features/john-gray-steven-pinker-violence-review> (access: 26.08.2016).

words: the advancement of cognitive capacities through scientific progress might itself contribute to a positive development of moral attitudes in the future.[45] Instead of looking out for new, abstract models for moral improvement, technological advancement might itself become the crucial ethical and pedagogical force we are searching for.

Sorgner does not generally exclude any form of biochemical moral enhancement—he solely diagnoses its contemporary impracticability and incompatibility with the modern value of individual freedom. Unfortunately—besides his very dubious faith in the correlation between technological, cognitive and moral improvement[46]—the German philosopher falls short of delivering any alternative to Persson's and Savulescu's proposal of moral enhancement, which in turn leaves the most essential question concerning the appropriate usage of emerging technologies still unanswered. This rejection of the idea of an obligatory and global moral bio-enhancement and renouncement of Bostrom's suggestion for a governmental regulation or restriction of emerging technologies to a small elite of politically approved scientific group[47], makes the promotion of values reducing existential risks in the near future even more indispensible. Otherwise the bitter scenario of scientific progress exacerbating the human condition might eventually become true.[48] So what are the options? Obligatory moral enhancement? Governmental restrictions of emerging technologies? Insurmountable existential insecurity?

In the next section I will try to indicate that if the techno-progressive discourse would open up for pedagogical anthropology, we might all benefit significantly on both a cognitive and moral level.

[45] Ibidem.

[46] Not only were the concepts of justice, freedom and solidarity heavily misused in the past—one shall not forget that the worst disasters of the 20th century had a lot to do with Enlightenment itself. See M. Horkheimer, T.W. Adorno, *Dialectic of Enlightenment: Philosophical Fragments*, transl. E. Jephcott, Stanford (CA) 2002.

[47] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford – New York 2016.

[48] I. Persson, J. Savulescu, "Moral Transhumanism", op. cit., pp. 666–667.

Friendly AGI? Education and dialogue might be the answers

In line with Bostrom's concept regarding the rise of existential risk factors through emerging technologies, John Gray Cox points out that one of the biggest threats in the nearest future will be the emergence of AGI systems capable of performing all human cognitive activities, or even beyond. The assumption that these new artifacts could not only be more intelligent but might be also hostile towards their 'creator', i.e. mankind, forces us to reflect on preventive mechanisms: "We must find," says Cox, "ways to balance the odds in favor of future Artificial Super Intelligences becoming ethical and we must also bias the odds in favor of becoming, ourselves, ethical enough to be viewed favorably and be treated well by such ethical ASI."[49] Instead of trying to keep up the pace of technological progress through becoming posthuman cyborgs or morally enhanced slaves, we should rather focus on forming humanlike machines—not only in a cognitive, but first and foremost in an ethical sense. The ultimate objective should therefore not only be development of solely "super-intelligent," but also "friendly AI." For this purpose Cox proposes a fundamental reorientation in ethical theory, pedagogy and legislation that might first apply to us and then likewise to our self-reflecting products. I will focus on the first two aspects—ethics and pedagogy.

Through relying closely on Cox's critique of the "current mainstream ethics"[50], I would like to point out three problematic aspects of contemporary ethical discourses: (1) The exclusion of dialogue in favor of mono-logical homiletic style of enforcing principles instead of elaborating them; (2) The designation of a "Golden Rule," which is recognized as the only, central and universally obligatory moral code; (3) The legitimization of moral principles on the basis of one, absolute, metaphysical truth. I will not discuss each point separately but rather elaborate the whole context as a pedagogical issue.

Leaving aside the philosophical-historical problem of whether humanity has ever been capable of formulating universal moral

[49] J.G. Cox, "Reframing Ethical Theory, Pedagogy, and Legislation to Bias Open Source AGI Towards Friendliness and Wisdom", *Journal of Evolution and Technology* 2015, vol. 25(2), p. 49.

[50] Ibidem, pp. 49–50.

principles, Cox points out that each pursuit of such an objective ends up finally in being "mono-logical," which leads to the problematic assumption "that given the principles and specific conditions, one person can determine what is the ethical thing to do. No dialogue is necessary."[51] One may argue whether the axiological plurality of social standards is rather functional or dysfunctional for a collective.[52] Machines, however, which would be programmed for obedience to one moral code, would be completely void of human features like uncertainty or remorse—even if the realization of installed principles would cause harm and bane. The rise of super-intelligent, yet morally mono-logical artifacts would undoubtedly turn out to be a disaster for human beings. Also the installment of various moral codes wouldn't be very promising, since the moral dilemmas could still only be solved by choosing finally one rule over all the others. The very humanlike situation, in which free individuals discuss and mutually criticize their standpoints, would simply not be an option. Hence Cox argues that forming new moral strategies should not focus on already existing ethical systems (the Kantian Imperative or Bentham's Greatest Happiness Principle for instance). Instead of developing mono-logical systems, which would merely be derivates of human authoritarianism, AGI researchers should rather focus on fabricating a social artificial being, i.e. entities willing to learn, negotiate and seek a consensus.[53]

It is very important to stress that every dialogical process is based on the continual creativity of its participants—not only in the sense of solving a particular ethical problem, but primarily as an activity which notoriously redefines the very root and essence of the problem. It is exactly this process of continuous reflection which enables the individual to recognize one's own interests, motivations and respectively modify and correct them in regard to the various interests and

[51] Ibidem, p. 41.

[52] On the one hand it seems to involve the problem of moral relativism—yet, on the other hand, plurality always rises the odds for innovation and therefore improvement.

[53] In this context one may certainly acknowledge the possibilities of using the achievements of game theory (as practiced in the field of cybernetics)—nonetheless this adoption of game theory should not be understood as a process of decoding an algorithm by one individual "gamer", as Cox underlines. Ibidem, p. 41.

motivations of others.[54] If we now once again take into consideration that AGI researchers are aiming to develop machines that would be capable of simulating all human cognitive activities, it seems to be essential to emphasize that these technological entities must also be suited for communicative action, which would serve as a forum for disputing various moral propositions. In other words: the only strong and yet friendly artificial intelligence would be the one capable of understanding the ambivalent and ambiguous character of life in general. Only then would the intelligent machines tend to reflect their ways of functioning and have an inclination to interact with other machines and humans.

To sum up: the necessary condition for developing strong, yet friendly AI is the creation of artificial dialogical entities which would not only perceive communication as an instrument to achieve practical goals and solve particular tasks, but which would also recognize interaction and mutual understanding as a fundamental value—not solely a means, but also an end. Even if AGIs would finally become super-intelligent machines one could then legitimately count on their willingness to adapt or even improve the communicative, personal and subjective principles of the human "lifeworld" (*Lebenswelt*). Friendly AI can only be understood as being integrated within the everyday social framework, in which individuals address each other personally through "I" and "You"—a dimension, which most definitely cannot be expressed through quantitative data, yet, as pedagogy proves, can be learned and practiced.

To enable machines to participate in social life seems to be the major objective yet it is not the only one. Since existential risks have to be prevented, the other necessary condition for friendly AI would be for mankind evolving—prior to the machines—towards a dialogical way of understanding morality and its ideational foundations. This might be the profound moral challenge for our contemporary pre-AGI times, certainly difficult to implement—yet, not that improbable as Persson's and Savulescu's proposal of an obligatory, global and functional biomedical moral enhancement, whose realization would *inter alia* violate the value of freedom, which in turn would have to be acknowledged as morally

[54] Ibidem.

illegitimate. How could we educate ourselves towards a dialogical conception of morality?

First of all, evolving towards a dialogical form of ethics would include the abandonment of a certain understanding of epistemological objectivity. Obviously, as Cox argues, we are entitled to hold certain truths that cannot be dismissed as mere subjective beliefs, e.g. the spherical shape of our planet. However these truths, although objective in the sense of being empirically proven features of our reality, cannot be seen as "necessary, absolute, universal truths."[55] In other words: even though our knowledge might be—at least to some extent—objectively true, we cannot claim to hold incontestable metaphysical truths with a capital "T". Nonetheless, moral principles have quite often been portrayed as being founded on absolute and universal truths, which in turn guaranteed them social indisputability. On the other hand: the departure from any metaphysical foundation of morality—the (in)famous "Death of God"—has induced many philosophers to reject morality *per se* as a misconception of values.[56] If we would assume morality to be solely an enforcement of certain principles unto the human through the process of education, then, indeed, morality would have to be rejected as mono-logical system. However Cox argues that we might be able to educate ourselves to repudiate any "Golden Rule" in favor of approving of "Rainbow Rules."[57] What would that mean for the very process of education and for the human self-understanding as moral beings?

Cox argues that most obligating moral principles in the past and in the present as well could be summarized by the phrase: "Do unto others as you would have them do unto you."[58] Cox argues that this

[55] Ibidem, p. 46.

[56] Still the probably most prominent rejection of morality on behalf of rejecting metaphysics and approving life-enhancing values, is Nietzsche's first essay of his *Genealogy of Morality*—"'Good and Evil', 'Good and Bad'". F. Nietzsche, *On the Genealogy of Morality*, transl. C. Diethe, Cambridge – New York 2006, pp. 10–34.

[57] J.G. Cox, "Reframing Ethical Theory, Pedagogy, and Legislation to Bias Open Source AGI Towards Friendliness and Wisdom", op. cit., pp. 43–45.

[58] Ibidem, p. 43.

very biblical[59] principle may apply quite well to homogenous societies and cultures, where one's own needs and desires frequently coincide with those of the other. However, heterogeneous and culturally diverse societies do not provide this sort of axiological stability—in fact it is only here where the "other" really becomes the *other* in an emphatic sense. There is no other way to find out what the other really needs or desires than by engaging oneself in a dialogue with the other individual. The "truth" about the other can never be found out through any kind of mono-logical reasoning—hence no "Golden Rule" could be successfully applied here. The foregoing phrase has to be reformulated into: "Do unto others as they would have me do unto them."[60] "This second rule," as Cox indicates,

> is one that recognizes and embraces the diversity in the world. It might be called the "Rainbow Rule", in that sense. It is a "rule" that is widely applied in successful ways by people in settings where there is considerable diversity in the interests and outlooks of people involved.[61]

Cox adds that this "Rainbow Rule" should be rather seen as a piece of advice or a guideline than a principle in a strong sense.[62]

Interestingly enough Cox finds traces of his rainbow-approach in ancient philosophical and religious traditions, such as Confucianism or Christianity with its crucial message: "love your enemies": "Enemies are people who are different and do not belong to a homogeneous population in our own community."[63] I find Cox's interpretation highly interesting in the context of the heated discussions about AGIs, moral enhancement and transhumanism, since there are many indicators that we already perceive the forthcoming super-intelligent machines, robots and cyborgs to become our enemies in the nearest future. But does this *Matrix*-scenario necessarily have to be true?

One may say that the odds are not very good if we notice the contemporary development of the techno-progressive discourse.

[59] Mt 7:12.

[60] J.G. Cox, "Reframing Ethical Theory, Pedagogy, and Legislation to Bias Open Source AGI Towards Friendliness and Wisdom", op. cit., p. 44.

[61] Ibidem.

[62] Ibidem, p. 45.

[63] Ibidem, p. 44.

If, and only if, we would succeed in developing a strong, yet dialogical AI, we might also assume the opposite scenario to come true: at the beginning the evolving machines might recognize us as their parents, later as their colleagues or friends, and finally as childlike, naïve animals.[64] Nevertheless even the last stage would not necessarily imply that they couldn't still preserve a friendly relationship, maybe even a caring and protective attitude towards mankind.[65] Since the basic features of forthcoming technological entities depend on our own moral condition, the best way to prevent hostile AGIs would be a reorientation of education and pedagogy. Instead of aiming to keep up the pace with technological improvement through an immorally moral biomedical enhancement of the human body, we should rather strive towards an opposite direction: rather than creating the technological man we should aim at developing humanlike technology. The reason why we are heading towards the technological man is probably rooted in the fact that we do possess a quite precise image of technology, which cannot be stated about our image of man. That in turn makes it easier for technology to be a reference point for further (in)human development and progress. The opposite way could only be adopted through a conceptualization of the "human."

Although transhumanism—even in its most Gnostic trends and versions—seems to be rooted outside metaphysical categories, it forces us to recap the fundamental question concerning essential attributes of "being human." This might be a very risky tendency, since—as I will argue from a pedagogical-anthropological perspective in the following chapter—the intellectual debate about the meaning of "being human" should never be closed. Paradoxically, despite the openly proclaimed openness for the various possibilities of (post)human development, transhumanism seems to follow a very fixed, reduced and naturalized image of man—devoid of any pedagogical dimension. In order to become a more promising conception of further human development than the existential, transhumanism should incorporate pedagogical notions to a larger extent than it currently does.

[64] Ibidem, p. 45.

[65] Ibidem.

In search for the (post)human? Maybe some questions should remain unanswered

Unlike in the case of technology, of which—as stated above—we have a rather definite notion, there are nearly endless ways and possibilities to approach the problem of humanity. It is not my objective to add one more idea, which would only intensify the rather obscure condition of being certain of the uncertain and still not being able to express the obvious obviousness—that I am "human." Instead I would like to refer one more time to an argument made by Cox. In order to develop in a human-friendly direction, AGI should eventually not only be intelligent in the sense of being effective, but also intelligent in the meaning of being wise.[66] Unfortunately the idea of "wisdom" plays a rather insignificant role in techno-progressive societies, which doubtlessly poses a direct consequence of the marginal role of philosophy—the "love of wisdom," not efficiency—in our contemporary culture.[67] It shouldn't be ergo a huge surprise that wisdom doesn't also play a key-role in the mainstream AI and AGI discourses. The proposal of Goertzel, according to which "General Intelligence" should be seen as the ability "to achieve complex goals in complex environments with insufficient knowledge and resources,"[68] only seems to confirm this view.

According to Cox wisdom could be seen as a systematic form of intelligence "that responds appropriately to the full range of values we should hold in the context in which we live."[69] The deciding difference between intelligence and wisdom should be therefore made through distinguishing the capability of effective realization of a specific value form the ability to built up a balanced axiological order.[70] However, this axiological order refers to a more general, philosophical

[66] Ibidem, p. 47.

[67] M. Nowak, "Między wiedzą naukową a mądrością w pedagogice – w obszarze filozofii wychowania", *Studia Paedagogica Ignatiana* 2016, vol. 19(1), p. 33.

[68] B. Goetzel, C. Pennachin, "The Novamate Artificial Intelligence Engine", op. cit., p. 74.

[69] J.G. Cox, "Reframing Ethical Theory, Pedagogy, and Legislation to Bias Open Source AGI Towards Friendliness and Wisdom", op. cit., p. 47.

[70] Ibidem, p. 48.

problem, i.e. the permanent need to confront and consider the problems concerning the accurate understanding of the empirical reality and the meaning of life in a dialogue, as Marian Nowak notices.[71] Unfortunately, our Western modern civilization has significantly narrowed the activities of the human mind to an "instrumentalist understanding of social life that came with industrialization, capitalism and modern bureaucracy."[72] The contemporary research on AGI only reaffirms this reduced concept of intelligence that seems to be devoid of any wisdom. As it seems to be already hard to bear humanity deprived of wisdom (human-made ecological or economic crisis come immediately into mind), the rise of super-intelligent artifacts being solely effective and functional in "achieving goals in a complex environment" would probably constitute an ultimate catastrophe. The rise of an inhuman, merely efficient power— that, as I believe, seems to be the ultimate existential risk deriving from contemporary techno-progressive researches. The question then arises: How could we address the danger of AGIs becoming super-intelligent, yet narrow minded specialists? Colloquially speaking: how could we prevent forthcoming AGIs from, despite being geniuses, antisocial and unwise "nerds"?

I believe this is the moment where the lack of pedagogical reflection, together with the marginalization of philosophy within the techno-progressive discourse, should be noticed, criticized and—finally—overcome. Through adopting a very reduced concept of intelligence the techno-progressive also wields a very limited idea of education. According to Sorgner, education can be described as a multiplicity of "processes that can be described as the general transmission of culture by parents, whereby culture is closely connected to an ideal of the good."[73] Although, as Sorgner stresses, that this approach stays "open to various ideals of the good, so it can be valid for various historical and contemporary settings," all educational procedures aim "at an improvement of the life of the child."[74] Consequently Sorgner

[71] M. Nowak, "Między wiedzą naukową a mądrością w pedagogice – w obszarze filozofii wychowania", op. cit., p. 33.

[72] Ibidem.

[73] S.L. Sorgner, "The Future of Education: Genetic Enhancement and Metahumanities", *Journal of Evolution and Technology* 2015, vol. 25(1), p. 33.

[74] Ibidem.

argues that the results of education could be achieved through applying liberal forms of genetic enhancement. Sorgner's line of reasoning is very coherent: if we decide to conceptualize education solely as a transmission of data (values, ideas, information, etc.), then it shouldn't sound awkward to anyone that various kinds of these transmissions are parallel, compatible or even interchangeable. The problem, however, lies not in the line of argumentation itself, but in the presumptions of the techno-progressive point of view concerning education. Sorgner points out the parallels between education and genetic enhancement (both bring about epigenetic alterations)—but he does not recognize the fundamental difference: the implied image of man.

A very characteristic and—as I believe—very representative remark expressing the philosophical foundations of the techno-progressive movement is the following statement made by Sorgner:

> Given recent biological research, given that human beings and great apes have common ancestors, and given a basically naturalist understanding of the world, it is more plausible to hold that there is merely a gradual difference between human beings, great apes, plants, and maybe even stones.[75]

Obviously, one may argue that the assumption concerning the qualitative difference between man and animal is not grounded in the biological, natural sphere but rather in the symbolic and cultural. However, this argument could only be seen as valid if our cultural meanings would be shared by all contemporary, globalized societies and therefore legitimized on a universal level. But it is a constitutive feature of our late-/postmodern era that we do not share a common sense on the most fundamental symbols. All cultural meanings can

[75] Sorgner heavily criticizes Jürgen Habermas, for example, for clinging "to an anthropology within which human beings have a special status, since they and only they are supposed to be rational and independent subjects. Although Habermas claims explicitly that he is in favor of a 'soft naturalism', he uses the concept of a special subject that is beyond any empirical analysis." In a different fragment Sorgner respectively rejects Immanuel Kant's "moral prohibition of treating a person solely as a means", since this moral imperative is also grounded in a (post)metaphysical worldview, which in turn rests on a distinction between persons and things." S.L. Sorgner, "The Future of Education: Genetic Enhancement and Metahumanities", op. cit., p. 39.

therefore be taken into question—among those the concept of the human being along with its alleged superiority and transcendent dignity.

The pedagogical implication of this anti-anthropological standpoint couldn't be more fundamental: If we consequently reject the idea of a qualitative difference between human beings and animals, any standpoint which does not empirically prove the factual existence of humanity as a distinctive, irreducible and physical quality, cannot be approved as a legitimate limitation for technological progress and human enhancement. The ethical question of whether we should work on genetic enhancement, AGIs, etc. has to be reversed: why shouldn't we?[76]

In order to refer to this problem I would like to hark back to the proclaimed openness of the techno-progressive discourse for other scientific disciplines, as stated after Goertzel before. Interestingly enough, Sorgner also proclaims the necessity of having "informed intellectual reflections, insights concerning our place in cultural history, and an awareness of the great plurality of philosophical, ethical, and religious positions that have been dominant in human history."[77] Unfortunately, Sorgner doubts whether

> specialists, experts, and scholars from the humanities in its traditional form possess all the necessary skills. I doubt that they do, and I also doubt that they start from appropriate premises, because they assume that solely human beings are categorically ontologically separate from all other natural beings.[78]

In other words: in order to be taken seriously by techno-progressive researchers, the humanities should give up their own premises and adopt the solely naturalistic standpoint which perfectly legitimizes human enhancement. This demonstrates quite clearly that the techno-progressive discourse considers itself to hold a privileged position which does not show any interest in a dialogue with disciplines that work with qualitative, non-measurable characteristics and aspects of the human life. However, if the stake were not as crucial as it is, i.e. the future of humanity and mankind, one could perhaps somehow tolerate

[76] Ibidem, p. 37, 39.

[77] Ibidem, p. 44.

[78] Ibidem.

this stalemate—but, as a matter of fact, humanities shouldn't accept a situation in which human enhancing technologies remain paradoxically blind to any humanistic point of view. This disheartening condition, in which technological improvement has lost nearly all of its philosophical ties to the fundamental premises of the humanities, I believe constitutes the crux of contemporary existential risks.

So are there any ways for humanities to refer to the problems of AGI and transhumanism without losing its own premises? As a matter of fact I believe there are several—however, I would like to sketch one that I find the most promising, i.e. pedagogical anthropology. As Christoph Wulf and Jörg Zirfas point out: since the last decades of the 20th century, educational science has renewed its interest in anthropological topics and from this moment pedagogy has ceased to doubt that explicitly expressed or silently implied images of man constitute an essential aspect of both pedagogical theory and praxis.[79] The central function of pedagogical-anthropological research, as Dieter Lenzen suggests, should therefore be the elicitation of the various mythologies concerning mankind, that have aroused within the different historical discourses.[80] Rudolf Lassahn then expects those mythologized images of man to be criticized from a pedagogical perspective.[81] In that sense Wulf and Zirfas argue that pedagogical anthropology has to be understood as critical anthropology, because many of the culturally implied images of man have turned out to be featured by ideological distortion, indifference to historical changes, strict norms, homogenization, ethnocentrism and violence.[82] These results of pedagogical-anthropological researches finally led to a very important conclusion, which contains a very important moral implication: in order to prevent further human disasters it would be very helpful if we could reject our tendency to search for complete and closed definitions of the human being. The reason is simple: every

[79] C. Wulf, J. Zirfas, "'Homo educandus'. Eine Einleitung in die Pädagogische Anthropologie", in: *Handbuch Pädagogische Anthropologie*, eds. C. Wulf, J. Zirfas, Wiesbaden 2014, p. 10.

[80] K. Schumann, *Menschenbilder in Erziehungswissenschaft, Neurowissenschaft und Genetik*, op. cit., p. 35.

[81] Ibidem.

[82] C. Wulf, J. Zirfas, "'Homo educandus'. Eine Einleitung in die Pädagogische Anthropologie", op. cit., p. 11.

time we presume to have found the definitive concept of the human, we also consequently detect the "inhuman," which can be then evaluated as less valuable or straight away as worth- and useless. In order to avoid these scenarios, which have led to a multitude of disasters in the past, it seems to be highly important to understand the human being as *homo absconditus*—an insolvable question, an inexpressible mystery, that cannot be subsumed to any essential, metaphysical category.[83] At first glance it might seem as if the techno-progressive discourse would embrace such a non-metaphysical, unessential approach to the human being concerning through promoting various ways of human enhancement. Unfortunately, the opposite seems to be the case and I would like to elaborate upon this critical assessment shortly.

It is undoubtedly true that technology has always been a significant part of human life—in fact one may even argue that technology is an anthropological quality, because only through creating artificial tools was mankind able to built its own world of existence, i.e. civilization and culture.[84] The inevitability of technology in human life indicates a very important feature of humanity itself: man is a non-specialized being. Paradoxically speaking: the nature of man is characterized by the lack of a fixed nature. Mankind builds its identity through a multitude of various technologies. At the same time, the human condition is also marked by a dangerous mechanism which Georg Simmel described as one of the features of the "tragedy of culture": means tend to become ends.[85] The extensive growth of the amount of technological devices generates a tangled mass of instruments which pushes humanity away from its original intentions. Finally the human being also becomes part of this network and becomes another thing among others.

Over the last century the clean dichotomy between man as subject and the artificial as object fell into obscurity, since bio-, neuro- and nanotechnology made it possible to see the human body as an object

[83] Ibidem, p. 13.

[84] J. Ahrens, "Technik", in: *Handbuch Pädagogische Anthropologie*, op. cit., p. 633.

[85] G. Simmel, "Der Begriff und die Tragödie der Kultur", in: G. Simmel, *Gesamtausgabe*, vol. 14, Frankfurt am Main 1996, pp. 385–416.

for technological manipulation.[86] Ahrens emphasizes that through the application of biotechnologies, mankind finally transgressed its own identity: obviously the development of media, transportation technologies or the internet essentially altered of living conditions—however, all them solely widened and transformed our naturally given senses and sensory perceptions and did not *per se* question the subjective role of the producing and operating individuals.[87] Biotechnologies, on the contrary, aim directly at merging the human subject with an artificial object, which in turn implies a fundamental change within the species-dispositive of mankind.[88] Eventually the difference between person and thing would not only vanish in the techno-progressive theory—it would also perish on the level of social life. This, however, would necessarily lead to the eradication of the most fundamental feature of human life: the existential openness, irreducibility and inexpressibility of the individual. That is exactly—at least from a pedagogical perspective—why education alone, and not genetic enhancement, can be seen as the only appropriate way of "enhancement." Not because technology poses an inhuman or evil force, but because only through education can mankind learn to use artificial organs consciously in a way that will not devour the specifically human existential openness. This is perhaps not the metaphysical foundation of humanity which Sorgner and other transhumanists are correct in rejecting—but it nevertheless constitutes the social reality and necessary moral condition for any human way of life which might be worth appreciating.

Conclusions

In this article I tried to reflect on the complex relations between Artificial General Intelligence, biochemical enhancement, transhumanism and pedagogical anthropology. Topical as they are, the various propositions of the techno-progressive discourses seem to consciously ignore certain implications of humanities social sciences—in particular those dealing with the fundamental concepts of education and upbringing.

[86] J. Ahrens, "Technik", op. cit., pp. 636–637.

[87] Ibidem, p. 638.

[88] Ibidem.

Humanities should play a key role within the techno-progressive discourses—but not for the price of foregoing its own fundamental premises concerning the concept of humanity. Indeed, the image of man has changed throughout the centuries notoriously. Yet, critical pedagogical anthropology indicates that the probably most important qualitative feature of humanity is its non-specialized, existential openness of being. Therefore the intrinsic existential risk deriving from emerging technologies is not characterized by the danger of reducing the human being to a measurable (and hence fully disposable) thing among other devices and organs. Unfortunately, unlike other horror scenarios resulting from technophobia, this existential risk poses a very realistic threat for mankind in the near future. It seems as if the techno-progressive discourse is intellectually grounded in a philosophy that rejects all forms of knowledge which would go beyond measurable data and hence identifies humanity on a qualitative level. This situation can only be countered through an integration of pedagogical anthropology as one of the crucial perspectives and correctional instances for techno-progressive monologues.

In reference to John Gray Cox I also tried to indicate that the existential risks of further AGI development can only be prevented through pedagogical techniques, which would not aim at making man more technical, but technology more human. As a consequence, it will probably be important to integrate the upcoming super-intelligent artificial organs into the social structure of human interaction and communication. Through this we could not only prevent human subjects from being physically reified, but develop artificial objects to become self-reflective subjects. Arguably this direction might be considered to be very risky—however, I believe that it should be recognized as the least risky of all of the existential risks that are yet to come.

Bibliography

Ahrens, J., "Technik", in: *Handbuch Pädagogische Anthropologie*, eds. C. Wulf, J. Zirfas, Springer, Wiesbaden 2014, pp. 633–641.

Becker P. von, *Der neue Glaube an die Unsterblichkeit. Transhumanismus, Bioethik und digitaler Kapitalismus*, Passagen Verlag, Wien 2015.

Bostrom N., "Existential Risk Preventing as Global Priority", *Global Policy* 2013, vol. 4(1), pp. 15–31.

Bostrom N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford – New York 2016.

Cox J.G., "Reframing Ethical Theory, Pedagogy, and Legislation to Bias Open Source AGI Towards Friendliness and Wisdom", *Journal of Evolution and Technology* 2015, vol. 25(2), pp. 39–54.

Goertzel B., Pennachin C., "The Novamate Artificial Intelligence Engine", in: *Artificial General Intelligence*, eds. B. Goertzel, C. Pennachin, Springer, Heidelberg 2006, pp. 63–129.

Goertzel B., "Superintelligence: Fears, Promises and Potentials", *Journal of Evolution and Technology* 2015, vol. 24(2), pp. 55–87.

Gray J., "Delusions of Peace", *Prospect. The Leading Magazine of Ideas*, October Issue 2011.

Harris J., "Moral Enhancement and Freedom", *Bioethics* 2011, vol. 25(2), pp. 102–111.

Horkheimer M., Adorno T.W., *Dialectic of Enlightenment. Philosophical Fragments*, transl. E. Jephcott, Stanford University Press, Stanford (CA) 2002.

Klincewicz M., "Autonomous Weapons Systems, the Frame Problem and Computer Security", *Journal of Military Ethics* 2015, vol. 14(2), pp. 162–176.

Morioka M., *Why is It Hard for Us to Accept Moral Bioenhancement: Comment on Savulescu's Argument*, Ethics for the Future of Life: Proceedings of the 2012 Uehiro-Carnegie-Oxford Ethics Conference, 2013.

Nietzsche F., *On the Genealogy of Morality*, transl. C. Diethe, Cambridge University Press, Cambridge – New York 2006.

Nowak M., "Między wiedzą naukową a mądrością w pedagogice – w obszarze filozofii wychowania", *Studia Paedagogica Ignatiana* 2016, vol. 19(1), pp. 19–38.

Okojie M. CPO, Olinzock A.A., Okojie-Boulder T.C., "The Pedagogy of Technology Integration", *The Journal of Technology Studies* 2006, vol. 32(2). Available at: <http://scholar.lib.vt.edu/ejournals/JOTS/v32/v32n2/okojie.html>.

Pennachin C., Goertzel B., "Contemporary Approaches to Artificial Intelligence", in: *Artificial General Intelligence*, eds. B. Goertzel, C. Pennachin, Springer, Heidelberg 2006, pp. 1–30.

Persson I., Savulescu J., "Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement", *Bioethics* 2013, vol. 27(3), pp. 124–131.

Persson I., Savulescu J., "Moral Transhumanism", *Journal of Medicine and Philosophy* 2010, vol. 35(6), pp. 1–14.

Persson I., Savulescu J., *Unfit for the Future: The Need for Moral Enhancement*, Oxford University Press, Oxford 2012.

Pinker S., *The Better Angels of Our Nature: Why Violence Has Declined*, Penguin Books, New York 2011.

Schumann K., *Menschenbilder in Erziehungswissenschaft, Neurowissenschaft und Genetik. Eine vergleichende Analyse*, Beltz Juventa, Weinheim – Basel 2015.

Simmel G., *Der Begriff und die Tragödie der Kultur*, in: G. Simmel, *Gesamtausgabe*, vol. 14, Suhrkamp, Frankfurt am Main 1996, pp. 385–416.

Sorgner S.L., "The Future of Education: Genetic Enhancement and Metahumanities", *Journal of Evolution and Technology* 2015, vol. 25(1), pp. 31–48.

Sorgner S.L., "The Stoic Sage 3.0 – A Realistic Goal of Moral (Bio)Enhancement Supporters?", *Journal of Evolution and Technology* 2016, vol. 26(1), pp. 83–93.

Wulf C., Zirfas J., "'Homo educandus'. Eine Einleitung in die Pädagogische Anthropologie", in: *Handbuch Pädagogische Anthropologie*, eds. C. Wulf, J. Zirfas, Springer, Wiesbaden 2014, pp. 9–26.

ADDRESS FOR CORRESPONDENCE:

Markus Lipowicz, PhD
Jesuit University Ignatianum in Krakow, Poland
markus.lipowicz@ignatianum.edu.pl