TO DEVELOP OR NOT TO DEVELOP, THAT IS THE QUESTION

Few topics stimulate and enliven the collective imagination today to the extent issues related to the development of artificial intelligence (AI) do, simultaneously adding a tinge of sensation to our lives. People tend to follow media reports on the new technologies and on the new areas of their application, as well as those related to the controversies over the—not always predictable or easily controlled—implementation of AI technologies. There is a widespread belief that, in the near future, radical and fundamental changes will be taking place, both in the immediate human environment and in the human way of being and acting. Some scenarios, sketched by thinkers such as Nick Bostrom,[1] Max Tegmark,[2] Ray Kurzweil,[3] and Kevin Warwick[4], imply that very soon we shall witness an intelligence explosion on a scale making the superintelligence capable of controlling the world. On March 22, 2023, an open letter was published on the website of the Future of Life Institute, calling on all AI labs to immediately pause, for at least the period of six months, the developing and training of AI systems more powerful than GPT-4 (Generative Pre-trained Transformer 4) created by OpenAI[5]. The letter was signed by, among others, Elon Musk, Yuval Noah Harari, Steve Wozniak, Jaan Tallinn, and by many other AI researchers. It names the fears and concerns ignited by AI development which have been made public by mass media, books, and movies: "We must ask ourselves: *Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should* we risk loss of control of our civilization?"[6]

Just as often as we hear about the risks, we also hear about the benefits of AI, ranging from rapid access to information, increased productivity, and greater security (as well as other social values, such as education or democracy), to care for the elderly or the sick. Reports on accomplishments made possible by AI are as numerous as they are impressive. Following news from the AI world, we oscillate between fascination, accompanied by hopeful anticipation, and anxiety which occasionally turns into fear for the future. This is no coincidence for, in the face of the continuing AI development, both hopes and fears are entirely legitimate. The fundamental factor triggering such conflicting attitudes is that already now one can experience the impact, in practically every aspect of human life, of the ever increasing saturation of the environment with AI-equipped objects. Moreover, the prospect of a further development of AI only contributes to the polarization of views regarding its ubiquity.

Various prognoses concerning the place and role of the human being in the world controlled by a self-perfecting superintelligence are considered: from optimistic visions that AI's growing intellectual and causative potential will be used for the benefit of humankind up to entirely catastrophic visions of the annihilation of humanity and the colonization of the entire universe by a

---

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

2 See Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (New York: Knopf Publishing Group, 2017).

3 See Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York and London: Penguin Books, 2000).

4 See Kevin Warwick, *March of the Machines: The Breakthrough in Artificial Intelligence* (Champaign: University of Illinois Press, 2004).

5 See "Pause Giant AI Experiments: An Open Letter," The Future of Life Institute, https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

6 Ibidem.

new form of intelligent life liberated from the bonds of dependence on the human being.[7] It is difficult to determine which prognosis is closer to fulfillment. Yet it seems very probable that the techno-scientific civilization has paved the way to autonomous techno-evolution (predicted already in the 1960s by Stanisław Lem in his *Summa technologiae*[8]) which may get out of human control. In order to carry on systematic reflection on AI, various global organizations were created, such as Consortium for the Benevolent Consciousness of Artificial Intelligence or the Future of Life Institute. The United Nations in turn considers creating an agency—inspired by the International Atomic Energy Agency—for monitoring AI development.[9]

The question posed in the title of this essay is in fact rhetorical. Artificial intelligence will definitely become a permanent factor in the human world, influencing the shape of our lives and values. The current epoch is called that of the fourth industrial revolution. As much as it affects our emotions and stimulates the collective imagination, the awareness that the entire humanity and the present world order are now facing inevitable substantial transformations calls for a responsible, rational focus on long-term planning. The sum of all these elements, or more precisely, their interplay and emergent interference, produce a precious cultural capital of which we should make the best possible use with a view to a better and safer future for the entire planetary population and for its natural and civilizational environment.

One may think and write about AI-related problems in many ways. They can be approached in a strictly technical way, by formulating specific construction tasks and then by looking for means of their implementation; economic, ecological, legal, ethical, pedagogical and other aspects of those problems might be considered; one can ask about the possibilities of AI applications in many important areas of life and work, such as industry, science, art, health care, education, and the military. All these issues are important and burning. To consider and solve them, intensive conceptual work and the best possible organization and synchronization of activities are indispensable.

Two other areas in need of reflection should also be indicated. They are equally important but more difficult to grasp, for they do not fit into the framework of a particular discipline or a specific set of competencies. The first one is broadly understood cybersecurity. It includes, among other things, considering ways to effectively protect humanity against the use of AI resources in bad faith or for wicked purposes, for instance by criminal or terrorist groups, or by individuals or communities seeking to gain advantage over others through unethical manipulation of AI technology, by using it against their competitors in the fight for scarce resources. Another aspect of cybersecurity which must not be neglected is the development, beforehand, of the most effective countermeasures to protect us against launching (consciously or accidentally) into undesirable and dangerous AI developmental paths that would end in the autonomous and uncontrolled creation of systems, programs and technologies directly or indirectly threatening people. All these risks are real and it would be tantamount to unforgivable recklessness to overlook them in the public debate or in specialized scientific discourse. The second issue that requires deep consideration is the relationship between humans and artificial intelligence systems. First of all, the issue in question concerns developing functional and culturally well-embedded models of thinking about, behaving in, and referring to the newly emerging and sometimes surprising (positively or negatively) civilizational space of interactions between human and non-human intelligence; models which should enhance positive interactions within that space and, as far as possible, allow us to avoid disturbing and dangerous ones.

---

7 See Eliezer Y u d k o w s k y, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, in: *Global Catastrophic Risk*, eds. Nick Bostrom and Milan M. Ćirković, Oxford University Press, Oxford 2008, 308-362.

8 See Stanisław L e m, *Summa technologiae*, trans. Joanna Żylińska (Minneapolis: University of Minnesota Press, 2013).

9 See Jason N e l s o n, "Take AI Warnings Seriously, Says UN Secretary-General," Decrypt U: News, https://decrypt.co/144692/take-ai-warnings-seriously-un-secretary-general.

The world around is to ever greater extent being filled with complex, unintelligible, and unpredictable devices which may serve the humans and broaden their horizons but which may also formulate and realize their own tasks, even those contrary to the best interest of humanity. The more the world is changing, the more we should care to create a safe zone of psychological comfort, based on reliable knowledge, as well as on wisely constructed cultural texts,[10] helping ordinary people to overcome the feeling of alienation, or perhaps even the weirdness of AI, and to feel comfortable in the environment of entities so similar to us and at the same time so different, equipped with "almost human" intelligence and simultaneously outstripping us in ever new areas of competence. This is not an easy task. Yet we must not fail in realizing it for such a failure would be tantamount to an alienation of human beings from the world in which setting rules and regulations of conduct will gradually cease to be their exclusive competence.

One must also come to terms with the unavoidable process of reshaping human identity caused by the implantation of AI advanced products into the human body or by changes in functioning of the cerebral cortex resulting from the brain's continuous contact with digital information-communication technologies. Moreover, what is at stake here it is not just the mental and behavioral adaptation of individuals to new aspects of the external and internal reality of artificial intelligence for we need to create new cultural frameworks, codes, and idioms in which artificial intelligence could be "naturalized." The term "artificial" as such embraces a disturbing ambiguity. One of its meanings refers to an artifact, an object which is not part of the natural environment but is created by means of tools in accordance with a prior project. The opposition in question is that between the artificial and the natural, i.e., between created by human beings and created by nature. This contrast remains valid, even if a growing number of elements in the environment of our life become artifacts. For what is, for instance, a garden in which the place for each plant has been carefully planned, and any naturally growing one is ruthlessly removed? Is it not a natural artifact?... Yet the term "artificial" conveys another potential opposition: what is artificial is non-natural, i.e., it is directed against nature. This emotional-evaluative component of the meaning of the term "artificial" is often negative, as it may be noticed in various contexts, for instance, when we criticize somebody's behavior as artificial or complain of getting artificial flowers instead of real ones. In a figurative sense, such mental associations, even if made unconsciously, mortgage artificial intelligence for they instantaneously evoke distrust, distance, and reserve (if not outright aversion) towards it. It seems the time has come to overcome such biases. This does not mean that any AI development should be welcome. On the contrary, we should not repeat negative stereotypes but take consciously critical approach and carry on reliable—as far as it is possible—analysis of dangers and risks connected to AI. Perhaps the fear of AI ruling the world and eliminating humans is unfounded. As Jobst Landgrebe and Barry Smith argue, the creation of the so-called strong artificial intelligence is mathematically impossible, and only such an intelligence could surpass human intelligence in all aspects.[11] This does not, however, mean that existential risks created by the development and use of "ordinary" AI in many areas of life do not deserve considerations, also ethical ones, or legal regulations[12].

The authors of the papers included in this volume of *Ethos* attempt to responsibly reflect on many of the issues raised above. They focus, among others, on the transformations of language,

---

10 See Anna M a j, *Przemiany wiedzy w cyberkulturze: Badania nad kulturą, komunikacją, wiedzą i mediami* (A Transformation of Knowledge in Cyberculture: Research on Culture, Communication, Knowledge, and Media) (Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2021).

11 See Jobst L a n d g r e b e and Barry S m i t h, *Why Machines will Never Rule the World: Artificial Intelligence without Fear* (New York and London: Routledge, 2023).

12 By the end of 2023 The European Union should regulate the use of AI with the AI Act, the world's first comprehensive AI law. See "EU AI Act: first regulation on artificial intelligence," European Parliament: News, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

which has become a tool for the cultural "domestication" of artificial intelligence. They also address the problem of the cultural functions and meaning of literary texts devoted to the relationship between humans and AI. The question of "machine ethics" is a gripping theme in this context, and it frequently recurs throughout the volume as the authors address the question of whether robots and other AI objects will assimilate the ethical values and norms inherent in human culture or rather create their own morality, perhaps devoid of human sensitivity.[13] In the case of medical robots, a problem one of the articles specifically explores, this question becomes crucial. Equally gripping is the issue of the connections between AI development and the imperative to protect our natural environment. Will AI save the world thanks to implementing ecologically optimal solutions on a planetary scale or will it, contrary to such expectation, accelerate the ecological disaster?

Any initiative that engages intellect, emotions, and imagination in working on a generally recognized project of optimal co-existence (co-habitation?) of the human being with any AI—whether already present or developed in the future—should be welcome with goodwill and satisfaction. The editors of this volume hope that papers authored by thinkers representing diverse academic circles will contribute to the deepening and dissemination of such a holistic, i.e., integrating various points of view and involving all dimensions of human perception of reality, approach to artificial intelligence. What will come next cannot be accurately predicted. But it is quite obvious that the worst possible solution is remaining indifferent to the coming future and passively waiting for future developments. As long as we have any influence on the directions of AI development, we must do everything possible to maximize the chance that a benevolent and friendly artificial intelligence will emerge from intensive scientific research.

Agnieszka Lekka-Kowalik and *Krzysztof Wieczorek*

13 See Kevin W a r w i c k, "Cyborg Moral, Cyborg Values, Cyborg Ethics," *Ethics and Information Technology* 5, no. 3 (2003): 131–7.