

## РАЗВОЈ ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ И БУДУЋНОСТ ДРУШТВА<sup>2</sup>

- 1. Роботи не сме да њовреди људско биће, нити да уздржавањем од делања дојуси да људско биће буде њовређено;*
- 2. Роботи мора да извршава наређења која му дају људска бића, осим ако се њо не коси са Првим законом;*
- 3. Роботи мора да води рачуна о свом ојсијанку, осим ако се њо не коси са Првим и Другим законом.*

Исак Асимов (Asimov, 2012, str. 8)

Пре равно осам деценија, познати писац научне фантастике Исак Асимов (Isaak Asimov) представио је свету своја три закона роботике – нормативни кодекс који би се морао уградити у сваку машину, да би се спречила могућа, али не и неизбежна доминација робота над људима.<sup>3</sup> Уколико нема унапред програмираних ограничења за машине, роботи и/или вештачка интелигенција постаће супериорнији од људи. У том случају, независно од свог претпостављеног еволутивног и/или можда божанског порекла, човечанство ће изгубити стечено преимућство над другим облицима

<sup>1</sup> vcvetkovic@fb.bg.ac.rs

<sup>2</sup> Уређивање темата и објављени радови део су Плана рада НИО Факултета безбедности за 2023. годину.

<sup>3</sup> Асимовљеви аргументи су једноставни, односно здраворазумски: наспрот природним потребама људи, машинама (роботима) нису потребни храна, одмор, сан; машине нису лимитиране ограничењима органске физиономије, оне су јаче, брже и способније за учење, а делови који се сломају или униште, лако се надокнађују. Најзад, способност учења машина далеко превазилази људске способности. О идеји да вештачка интелигенција може постићи тзв. суперинтелигенцију, дефинисану као „било коју интелигенцију која у значајној мери превазилази когнитивне перформансе човечанства у свим доменима“ (Bostrom, 2014, p. 39; в. Muller & Bostrom, 2016).

живота на Земљи, више неће бити највиши облик интелигенције, нити главни преда-тор на Земљи. Слично пређашњим *hoto* облицима живота, људи ће нестати у борби, такмичењу или напросто суживоту са савршенијим обликом живота.

У времену када је Асимов осмишљавао и писао своје кратке приче о роботима (Asimov, 2012) које су садржале спекулације о (не)очекиваним исходима модерног доба, већ је увелико била развијена филозофска критика модерне која је поставила дијагнозу о нихилистичким последицама техницизоване савремености, која подједнако успешно поништава „нагонске“ (Ниче), „аутентичне“ (Хајдегер), „духовне“ (Берђајев) и све друге потребе/могућности човека. Вероватно да нико од великих мислилаца нихилизма није могао да претпостави границе модерног сцијентизма и његове главне полуте моћи оличене у савременој технологији. Постојала је развијена филозофска, политичка и друга дебата у којој се критиковала метафизичка позиција модерног човека, који наивно верује да може преузети место и улогу свог Творца (Достојевски и др.), али готово да нико, осим романописаца научне фантастике, није очекивао да ће машине дословно заменити људе у креирању (не само њихове) стварности. Могу ли машине (људски производи) доспети у ситуацију да дословно контролишу свог творца – то је ново, не толико метафизичко, колико емпиријско питање, које јесте и које ће бити предмет нових филозофских, религијских, политичких и других дискусија, тим пре што је употреба модерне технологије већ довела до тога да човек постане „сувишан“ у многим облицима рада, као и у самом одлучивању о томе „шта да се ради“.

Ипак, савремене машине, односно работи као њихов најсложенији облик, још увек нису отелотвориле свет Асимовљевих визија. То не значи да нису на добром путу. У последњих двадесетак година, од времена тзв. четврте индустријске револуције, интеграција све комплекснијих машина у нашу свакодневицу поприма контуре света које превазилазе филозофске и ине страхове, па чак и машту аутора научне фантастике. Данашња сложена хардверска и софтверска решења више сличе Асимовљевим роботима него машинама у уобичајеном смислу те речи. Довољно је само присетити се наше зависности од мобилних телефона па схватити у којој мери је техника наш „продужетак“, а колико израз зависности која надилази „основне“, али и „највише“ људске потребе. Међутим, истински искорак у свету технике, односно науке,<sup>4</sup> не огледа се толико у нашој (стеченој) зависности од машина, колико у могућности да их учинимо заиста интелигентним и самосталним ентитетима мимо наше воље. Алудирајући на Асимовљеве законе роботике, није сувишно упитати се када ће интелигентна, не само „рационална“ већ можда чак и духовита или емотивна машина (!), „одлучити“ да не ради по наређењима творца, већ да дела по своме, онако како налаже „разум“, тј. логика проблема и постојећи избор делотворних поступака који воде ка „најбољем решењу“. Када и уколико тако нешто постане могуће, тј. буде реалност, машине ће постати посебан, нови облик „живота“. Он би онда могао да створи властити, самостални свет нељудске, штавише – вештачке (не-природне)

---

<sup>4</sup> Историјски је тешко, а вероватно и непотребно одређивати првенство у односу наука–техника с обзиром на то да се одговори увек формирају у складу с произвољним почетним премисама о правилима (законима или тенденцијама) друштвеног развоја и претпоставкама о темељима људске природе.

егзистенције. Како би онда изгледао „суживот“ људи (природе) и машина (не-природе), не можемо ни да претпоставимо.

Савремени ниво развоја вештачке интелигенције, машина које су способне да саме закључују и изналазе „најбоља“, тј. најрационалнија (и најделотворнија) решења по свим могућим питањима, очигледно представља корак ближе „Асимовљевом добу“ – времену у коме човек и природни феномени (земљотреси, суше, поплаве и сл.) престају да буду једини покретачи *грушїивених* збивања. Модерни филозофи су ту могућност одавно препознали и дијагностиковали, с тим да вероватно нико од њих није могао ни да наслути куда то (заиста) води. Данашњи програми „за разговор“ (ChatGPT и др.), који су засновани на милијардама података који су увезани у јединствени виртуелни космос на интернету, још увек обликују одговоре и текстове само на основу унесених података. И даље не знамо да ли ће они колико сутра бити у стању да произведу нешто уистину „ново“ – реалност која би излазила из домена постојећег и (тренутно) стварног. Ако би дошло до промене у том смислу, машине би закорачиле у властити свет емоција и/или идентитета, што би значило да оне нису људи, али и јесу засебан облик живота, а он, ако је веровати науци, подразумева борбу за опстанак која би у овом случају била крајње неравноправна.

Вештачка интелигенција има потенцијал да једног дана уништи *praxis* и већину послова (активности, делатности итд.) којима се људи баве. То може обесмислити све што знамо о друштвеној историји и политичким заједницама, сукобима и солидарности, ратовима и сарадњи итд. Свагда несавршен људски свет коначно може постати „застарео и превазиђен“, а самим тим и *сувишан* ако га посматрамо из аспекта одржања природе или онога што под природом или стварношћу (још увек) подразумевамо. Тако бар виде будућност многи савремени аутори, који размишљају о стварности коју развој вештачке интелигенције собом носи (Yudkowsky, 2011, str. 316–320; Bostrom, 2014, pp. 39–47; Muller, 2016; Muller & Bostrom, 2016). Међутим, као и увек у људским стварима, све то „може да буде, али не мора да значи“. Паралелно са песимистичким визијама будућности у којима машине господаре људима, опстојавају, не без аргумената, и пројекције у којима је вештачка интелигенција само нови облик превазилажења многобројних проблема с којима се човечанство одувек суочавало – почев од сфере медицине (здравља) и екологије, преко привреде (производња, саобраћај и транспорт, урбано планирање, пружања јавних и комуналних услуга итд.), до политике (Savina, 2021; Sokolov *et al.*, 2017; Hassankhani *et al.*, 2021; Kim, Jung & Choi, 2016; Kissinger, Schmidt & Huttenlocher, 2021). Употреба вештачке интелигенције, све док се њоме може управљати, препуна је уобичајених друштвених „плусева“ и „минуса“, и у том погледу, можда, нема ничег новог.

У сваком случају, вештачка интелигенција већ увелико обликује наш свет и то ће наставити да чини и убудуће. Другим речима, неупитно је да смо већ добрано загазили у „Асимовљево доба“ у којем симбиотички однос између машине и човека одређује мноштво друштвених интеракција и процеса: од начина комуникације, пројектовања привредног развоја и политичких визија, до вођења ратова и рефлексивања истине и вредности. Тематски блок *Социолошкої ѝреїледа* посвећен је тек неколиким, специфичним друштвеним употребама вештачке интелигенције које

обухватају стварност тзв. паметних градова), војно и/или политичко коришћење, као и употребу у контексту актуелних нанотехнологија.

Упркос ламентирању различитих врста скоројевића о „урбаном“ или „цивилизованом“ и оном другом „не-урбаном“ (дакле „не-цивилизованом“, тј. примитивном) становништву, чињеница је да већину човечанства данас чине „урбани грађани“. Према подацима УН, више од половине данашњих осам милијарди људи живи у градовима, а очекује се да ће до средине овог века најмање 75% укупног светског становништва бити насељено у милионским градовима. То је податак који потенцијално може више утицати на промену живота и тзв. људске природе него сви досадашњи цивилизацијски процеси заједно. Они су управо и започели са урбанизацијом (пре око десетак хиљада година), која је нераскидиво везана са развојем технологије. Данас је живот у граду апсолутно немогућ без технологије и науке, укључујући ту и њихов највиши израз – вештачку интелигенцију. О различитим аспектима употребе и друштвеним последицама коришћења вештачке интелигенције у концепту „паметних градова“<sup>5</sup> више у текстовима Д. Кесића, М. Томића и Л. Димитријевића, док радови М. Копане и Д. Поповић Јовановић, третирају, између осталог, два екстремна пола употребе вештачке интелигенције – ратовање<sup>6</sup> и лечење (у контексту нанотехнологија).<sup>7</sup> Речени избор тема има скромну амбицију да буде својеврсни домаћи „увод у проблем“ коришћења или употребе вештачке интелигенције, теме која ће вероватно одлучујуће утицати на будући смер људске историје, каква год она била.

---

<sup>5</sup> Емпиријско праћење функционисања „паметних градова“ на основу употребе софтверских решења у пружању локалних услуга (Smart City Index) спроводе организације из Швајцарске и Сингапура – Међународни институт за развој менаџмента (International Institute for Management Development) и Универзитет за технологију и дизајн (University of Technology and Design). Видети <https://www.smartcitiesindex.org>

<sup>6</sup> Видети извештај НАТО заједничког сајбер одбрамбеног центра за изврсност (The NATO Cooperative Cyber Defence Centre of Excellence) где се утврђује степен имплементације вештачке интелигенције у оружаним снагама НАТО држава (Gray and Ertan, 2021).

<sup>7</sup> Кључни нанотехнолошки индикатори (Key nanotechnological indicators), које развија Организација за економску сарадњу и развој (Organization for Economic Co-operation and Development), могу се наћи на <https://www.oecd.org/sti/nanotechnology-indicators.htm>

## ARTIFICIAL INTELLIGENCE DEVELOPMENT AND THE FUTURE OF SOCIETY<sup>2</sup>

(Translation *In Extenso*)

1. A robot may not injure a human being or, through inaction allow a human to come to harm;
2. A robot must obey the orders given it by human beings, unless such orders would conflict with the first law;
3. A robot must protect its own existence, as long as such protection does not conflict with the first or second law.

(Isaak Asimov, 2012, p. 8)

As many as eight decades ago, well-known science fiction writer Isaak Asimov presented his Three Laws of Robotics to the world – a normative code that should be built into every machine in order to avoid the potential, yet not inevitable, domination of robots over people.<sup>3</sup> If there are no previously programmed limitations for machines, robots and/or artificial intelligence will become superior to humans. In that case, regardless of its assumed evolutionary and/or perhaps divine origin, humanity will lose the precedence acquired over other forms of life on Earth; it will no longer be the highest form of intelligence or

---

<sup>1</sup> vcvetkovic@fb.bg.ac.rs

<sup>2</sup> The edited thematic units and published papers are part of the 2023 Work Plan of the Scientific and Research Board of the Faculty of Security Studies.

<sup>3</sup> Asimov's arguments are simple, i.e., sensible: in contrast to people's natural needs, machines (robots) do not need food, rest or sleep, machines are not limited by the restrictions of organic physiognomy; they are stronger, faster and more capable of learning, while parts that get broken or destroyed are easily replaced. Finally, the ability of machines to learn exceeds human abilities. For the idea that artificial intelligence may achieve so-called superintelligence, defined as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Bostrom, 2014, p. 39, see Muller & Bostrom, 2016.)

the main predator on Earth. In a similar way to the previous *homo* forms of life, people will disappear in struggle, competition or simply in cohabitation with a more perfect form of life.

At the time when Asimov devised and wrote his short stories about robots (Asimov, 2012) that contained speculations about (un)expected origins of the modern era, there had already been a developed philosophical critique of the modern diagnosing the nihilistic consequences of technicized contemporary time, which equally successfully annuls the “instinctive” (Nietzsche), “authentic” (Heidegger), “spiritual” (Berdyayev) and all other human needs/capabilities. None of the great thinkers of nihilism could not have assumed the boundaries of modern scientism and its main lever of power embodied in modern technology. There were extensive philosophical, political and other debates criticizing the metaphysical position of modern man, who naively believes that he can assume the place and role of his Creator (Dostoyevsky and others), but almost no one, except for science fiction writers, did not expect machines to virtually replace humans in creating (not only their) reality. Can machines (human products) get into a position to virtually control their creator – that is a new question, not so metaphysical as empirical, which is and will be the subject of new philosophical, religious, political and other discussions, even more so because the use of modern technology has already led to man becoming “superfluous” in many forms of work, as well as in the very decision-making about “what should be done”.

However, modern machines, i.e., robots as their most complex form, have not yet embodied the world of Asimov’s visions. It does not mean that they are not on a good road towards it. In the past twenty years or so, since the time of the so-called Fourth Industrial Revolution, integration of ever more complex machines into our everyday life has assumed the contours of the world that overcome philosophical and other fears, and even the imagination of science fiction writers. Today’s complex hardware and software solutions resemble Asimov’s robots more than machines in the common sense of the word. It is enough to remember our dependence on mobile phones to realize to what extent technology is our “extension” and to what extent it is the expression of addiction that overcomes the “basic”, but also the “greatest” human needs. Nevertheless, the true step forward in the world of technology, i.e., science,<sup>4</sup> is not reflected so much in our (acquired) dependence on machines as in our ability to make them really intelligent and independent entities, regardless of our will. Alluding to Asimov’s Laws of Robotics, it is not superfluous to ask when an intelligent, not only “rational”, but perhaps even spiritual or emotional machine (!) “will decide” not to work by the orders of its creator, but to function as required by the “sense”, or the logic of the problem and the existing choice of effective procedures leading to the “best solution”. When and if something like that becomes possible, i.e., real, machines will become a separate, new form of “life”. Then it could create its own, independent world of non-human, or even – artificial (non-natural) existence. We cannot even assume how “cohabitation” of humans (nature) and machines (non-nature) would look.

The modern level of the development of artificial intelligence, or machines capable of concluding on their own and finding the “best”, or the most rational (and the most

---

<sup>4</sup> Historically, it is difficult, and probably useless to determine the precedence in the relationship science-technology, having in mind that the answers are still being formed in line with the arbitrary initial premises about the rules (laws or tendencies) of social development and the assumptions about the foundations of human nature.

effective solutions to all possible questions, is evidently a step closer to “Asimov’s era” – the time when man and natural phenomena (earthquakes, droughts, floods etc.) cease to be the only drivers of *social* happenings. Modern philosophers recognized and diagnosed that possibility a long time ago, but probably none of them could even anticipate where it (actually) led. Today’s “chatting” programs (ChatGPT etc.), which are founded on billions of data connected into a unique virtual universe on the internet, are still shaping the answers and texts on the basis of entered data. We still do not know whether as soon as tomorrow they will be able to produce something truly “new – a reality deriving from the domain of the existent and (currently) real. In the event of any change in that sense, machines would step into their own world of emotions and/or identities, which would mean that they are not humans, but also that they are a separate form of life which, if we are to trust science, involves fight for survival that would be extremely unequal in this case.

Artificial intelligence has the future potential to destroy *praxis* and the majority of jobs (activities, actions etc.) people deal with. It may make senseless everything we know about social history and political communities, conflicts and solidarity, wars and cooperation etc. The ever-imperfect human world can finally become “old-fashioned and obsolete”, and thus *superfluous*, if observed from the aspect of the preservation of the nature or what we (still) refer to as the nature or reality. At least this is how the future is seen by many modern authors thinking about the reality brought along by the development of artificial intelligence (Yudkowsky, 2011, pp. 316-320; Bostrom, 2014, pp. 39-47; Muller, 2016; Muller & Bostrom, 2016). However, just as it always occurs with human matters, all that “is possible, but not necessarily”. Parallel to the pessimistic visions of the future in which machines are masters to humans, there are also projections, not without arguments, in which artificial intelligence is just a new form of overcoming numerous problems humanity has always encountered – starting from the sphere of medicine (health) and ecology, via economy (production, traffic and transport, urban planning, provision of public and utility services etc.), to politics (Savina, 2021; Sokolov *et al.*, 2017; Hassankhani *et al.*, 2021; Kim, Jung & Choi, 2016; Kissinger, Schmidt & Huttenlocher, 2021). The use of artificial intelligence, as long as it can be managed, is full of usual social “pluses” and “minuses” and, in that respect, there might be nothing new.

In any case, artificial intelligence is already shaping our world and it will continue to do so in the future. In other words, it is indisputable that we have already stepped far into “Asimov’s era”, in which the symbiotic relationship between the machine and man determines a number of social interactions and processes: from the manner of communication, projecting economic growth and political visions, to warfare and reflections of truth and values. The thematic block of the Sociological Review is dedicated to several specific social uses of artificial intelligence that include the reality of the so-called smart cities, military and/or political use, as well as the use in the context of current nanotechnologies.

Despite the lamentations of all sorts of upstarts about the “urban” or “civilized”, and the other, “non-urban” (i.e., “uncivilized”, or primitive) population, the fact is that today’s humanity is mainly composed of “urban citizens”. According to the UN data, more than a half of eight billion people live in cities nowadays, and it is expected that minimum 75% of the world’s total population will be living in the cities with millions of inhabitants by mid-21<sup>st</sup> century. This information may potentially affect life and the so-called human

nature more than all civilization processes so far collectively. These actually began with urbanization (about ten thousand years ago), which is inextricably related to the development of technology. Today, life in the city is absolutely impossible without technology and science, including their highest manifestation – artificial intelligence. Further details about different aspects of the use and the consequences of the use of artificial intelligence in the concept of “smart cities”<sup>5</sup> can be found in the texts by D. Kesić, M. Tomić and L. Dimitrijević, while the papers by M. Kopanja and D. Popović Jovanović treat, among other things, two extreme poles of the use of artificial intelligence – warfare<sup>6</sup> and treatment (in the context of nanotechnologies).<sup>7</sup> The said selection of topics has a modest ambition of being a specific domestic “introduction to the problem” of using or utilizing artificial intelligence, the topic that will probably have a decisive influence on the future direction of human history, whatever it is like.

#### REFERENCES / ЛИТЕРАТУРА

- Asimov, I. (2012). *The Complete Robot*. Beograd: Čarobna knjiga. [In Serbian]
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Gray, M. & Ertan, A. (2021). *Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment*. Tallinn: The NATO Cooperative Cyber Defence Centre of Excellence. [https://ccdcoe.org/uploads/2021/12/Strategies\\_and\\_Deployment\\_A4.pdf](https://ccdcoe.org/uploads/2021/12/Strategies_and_Deployment_A4.pdf)
- Hassankhani, M., Mehdi A., Sharifi, A. & Azhdari, A. (2021). Smart City and Crisis Management: Lessons for the COVID-19 Pandemic. *Environmental Research and Public Health*, 18, 1–18, [doi.org/10.3390/ijerph18157736](https://doi.org/10.3390/ijerph18157736).
- Kim, K., Jung, J-K. & Choi, J. Y. (2016). Impact of the Smart City Industry on the Korean National Economy: Input-Output Analysis. *Sustainability* 8, 1–19. doi: [10.3390/su8070649](https://doi.org/10.3390/su8070649)
- Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2021). *The Age of AI: and our human future*. London: John Murray.
- Muller, V. & Bostrom, N. (2016). Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In: V. Muller (ed.) *Fundamental Issues of Artificial Intelligence* (555–572). Cham: Springer.

---

<sup>5</sup> The empirical monitoring of the functioning of “smart cities” based on the use of software solutions in the provision of local services (Smart City Index) is conducted by the organizations from Switzerland and Singapore – the International Institute for Management Development and the University of Technology and Design). See; <https://www.smartcitiesindex.org>

<sup>6</sup> See the report of the NATO Cooperative Cyber Defence Centre of Excellence), which determines the degree of artificial intelligence implementation in the NATO member-states’ armed forces (Gray and Ertan, 2021).

<sup>7</sup> The key nanotechnological indicators developed by the Organization for Economic Co-operation and Development may be found at: <https://www.oecd.org/sti/nanotechnology-indicators.htm>



- Muller, V. (2016). Editorial: Risks of Artificial Intelligence. In: V. Muller (ed.) *Risks of Artificial Intelligence* (1–8). London and New York: CRC Press.
- Savina, S. V. (2021). The Smart City Concept. *Ekonomika: vchera, segodnya, zavtra*, 11 (1A), 403–409. DOI: [10.34670/AR.2021.17.93.006](https://doi.org/10.34670/AR.2021.17.93.006). [In Russian]
- Sokolov, I, Drozhzhinov, V., Raikov, A., Kupriyanovsky, V., Namiot, D. & Sukhomlin, V. (2017). On artificial intelligence as a strategic tool for the economic development of the country and the improvement of its public administration. *International Journal of Open Information Technologies*, 5(9), 76–101 [In Russian]
- Yudkowsky, E. (2011). Artificial Intelligence as a positive and negative factor in global risk. In: N. Bostrom, M. Ćirković (eds.). *Global Catastrophic Risks* (308–345). Beograd: Helix [In Serbian]