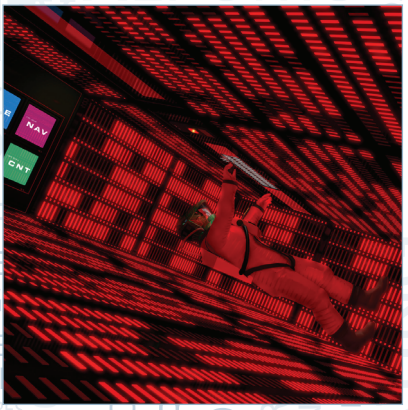
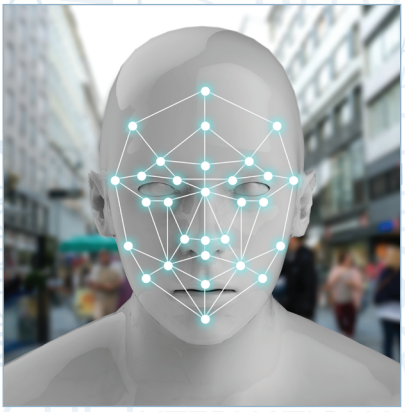
RIAC

RUSSIAN INTERNATIONAL AFFAIRS COUNCIL





WORKING PAPER

INTERNATIONAL AND SOCIAL IMPACTS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES

44 / 2018

RUSSIAN INTERNATIONAL AFFAIRS COUNCIL

MOSCOW 2018

International and Social Impacts of Artificial Intelligence Technologies. Working Paper No. 44 / 2018 [S.V. Karelov, M.V. Karliuk, A.G. Kolonin, N.M. Markotkin, D.R. Scheftelowitsch]; [I.S. Ivanov, Editor-in-Chief]; Russian International Affairs Council (RIAC). – Moscow: NPMP RIAC, 2018. – 52 pages. – Names of authors and editors are listed on reverse of title page.

ISBN 978-5-6041888-0-4

The Working Paper focuses on possible impacts of related technologies, such as machine learning and autonomous vehicles, on international relations and society. The authors also examine the ethical and legal aspects of the use of AI technologies. The present Working Paper of the Russian International Affairs Council (RIAC) includes analytical materials prepared by experts in the field of artificial intelligence, machine learning and autonomous system, as well as by lawyers and sociologists. The materials presented here are intended to contribute to the public dialogue on issues of artificial intelligence and the possible consequences of using this technology.

The full text of the working paper is available on RIAC's website. You can download it or leave a comment via this direct link – **russiancouncil.ru/en/paper44**

Contents

Introduction

Artificial Intelligence (AI), which until recently was the exclusive realm of science fiction, is now one of the most promising and rapidly developing technologies. Limited or "weak" AI technologies are regularly used in various fields, from mobile phones and consumer electronics to defense products.

At the same time, there are very few areas of scientific knowledge that are quite as mythologized as AI. Largely due to the images created by popular culture, artificial intelligence often appears in the media as a kind of Pandora's box which, when opened, will inevitably lead humanity to the brink of apocalypse. While such alarmism seems unfounded, the development of "strong" artificial intelligence that is capable of making informed management decisions is clearly on the agenda today.

The prospect of creating a technology of this kind largely calls into question not only the current system of the global division of labour, but also the existing world order and the international security architecture. Against the backdrop of the heightened tensions and the collapse of confidence between the great powers, there is a real danger that a new arms race in the field of artificial intelligence technologies could develop. Having said that, while control over the other types of weapons of mass destruction – nuclear, chemical and biological – is regulated by the relevant international agreements and conventions, the development of military artificial intelligence technologies remains in a "grey area" of international law.

At the same time, it is clear today that artificial intelligence will be increasingly used in the military and civilian spheres, including in terms of ensuring strategic stability. The temptation to obtain a new and more advanced weapon and pull ahead in the technological race is too great for countries to take humanitarian considerations into account.

One of the problems facing researchers is the massive conceptual and terminological confusion that exists in the field of AI. This working paper systematizes the current knowledge of AI technologies and evaluates the prospects for its development.

At the centre of our analysis is the possible impact of artificial intelligence and related technologies (machine learning, automated devices) on international relations and various spheres of public life. The articles contained herein also consider the ethical and legal aspects of implementing AI technologies.

The present Working Paper of the Russian International Affairs Council (RIAC) includes analytical materials prepared by experts in the field of artificial intelligence, machine learning and autonomous system, as well as by lawyers and sociologists. The materials presented here are intended to contribute to the public dialogue on issues of artificial intelligence and the possible consequences of using this technology.

Artificial Intelligence: Time of the Weak

D. Scheftelowitsch

The term "artificial intelligence (AI) normally denotes a computer program, i.e. an *algorithm*, that is capable of solving problems which an adult human brain can solve. The *International Dictionary of Artificial Intelligence*[1] denotes AI as the area of knowledge that deals with the development of technologies that enable computing systems to operate in a way that resembles intelligent behaviour, including the behaviour of humans. It should be noted that this definition is phenomenological; it leaves the notions of intellect and intelligent behaviour to be detailed by philosophy. Given the shortage of knowledge about the brain and the cognitive apparatus of biological systems, the notion of AI defies more precise mathematical formalization.

At present, AI research is normally understood as the use of algorithms to solve problems that require cognitive strain.[2] Such problems used to imply (and partially still imply) games such as chess and Go, handwriting recognition, machine translation and creative work. In the eyes of the general public, each of these problems was originally perceived as the last remaining obstacle to the creation of a "true" AI, one that would be able to replace humans in all fields of expertise. In reality, it turned out that teaching a computer to play chess is much easier than teaching it to play football, and that even if we know how to teach a computer to play football this knowledge is difficult to apply to machine translation. This is why, following an initial peak in enthusiasm, the scientific community subdivided the notion of AI into the *strong* and *weak* subcategories. *Weak AI* implies an algorithm capable of solving highly specialized problems (such as playing chess), whereas *a strong AI* can be used to solve a broader range of problems; ideally, it should be capable, at the very least, of everything an adult human is capable of (including arriving at logical conclusions and planning its future actions). Strong AI is used interchangeably in literature with *artificial general intelligence* (AGI).[3]

One thing worth noting is that the strong–weak classification of AI algorithms is not set in stone: as of early 2016, it was believed that playing Go required a strong AI. After the impressive win of the AlphaGo algorithm over the current human world champion in March 2016,[4] this problem was relegated to the domain of weak AI.

The following main trends in AI research can be identified as of the late 2010s.

[1] Raynor W. Jr International Dictionary of Artificial Intelligence. London, United Kingdom: Global Professional Publishing. 2008.

[2] Boulanin V., Verbruggen M., SIPRI Mapping the development of autonomy in weapon systems. SIPRI, Solna: SIPRI. 2017.
Floreano D., Mattiussi C. Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies. The MIT Press. 2008.

[3] Yudkowsky E. Levels of Organization in General Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg. 2007. pp. 389–501.

[4] Silver D. et al. Mastering the game of Go with deep neural networks and tree search // Nature. 2016. (529). pp. 484–503.

Machine Learning and Pattern Recognition

Searching data for consistent patterns, such as classifying objects in photographs as backgrounds, humans, vehicles, buildings or plants.

Planning and Inference

Proving logical statements and planning actions for the purpose of achieving a certain goal, based on the knowledge of the logical laws that allow this goal to be achieved. One example would be synthesizing sensor data to assess the traffic situation for the purposes of controlling a vehicle in an effective way.

Expert Systems

Systematizing data together with logical connections, knowledge mapping and answering semantic questions such as "What is the share of energy prices in the production costs for the Irkut MC-21 aircraft?"

Current Status and Prospects

The objective of creating a strong AI has yet to be achieved. In fact, the scientific community has gone perceptibly cold on AI following the so-called AI winter of the mid-1980s, which was brought about by the initially inflated expectations. This resulted both in the disillusionment among potential AI users and in the insufficient performance of software.[5] Considerable progress has been achieved in more specialized fields since the mid-2000s. This is primarily due to the continuous evolution of computing equipment: whereas in 2001 the industry's flagship Intel Pentium III processor demonstrated the highest level of performance across the board at 1.4 billion arithmetic operations per second, its descendant, the Intel Core i5 of 10 years later was capable of 120 billion operations per second, i.e. nearly 100 times more.[6] This growth was to a significant extent spurred by the video gaming and computer graphics industry, as the race for ever more realistic imagery turned GPUs from peripheral hardware into powerful computing systems capable of not just processing graphical data, but also of random parallel computing tasks (albeit of relatively low complexity). In 2011, the peak performance of the Radeon HD 6970 video card stood at 675 million operations per second.[7]

The other important driver of the industry was the capability to digitize and manually classify texts, images and audio files, and also to create comprehensive digital knowledge bases. The accessibility of massive volumes of properly classified data allowed for large samples to be used in the training of machine-leaning algorithms. The classifying precision increased,[8] turning machine transla-

[5] Russel S., Norvig, P. Artificial Intelligence: A Modern Approach. London: Pearson, 2009.
[6] Intel Corporation. Intel® Core i5-2500 Desktop Processor Series // Intel Corporation. 2011. URL: https://www.intel.com/content/dam/support/us/en/documents/processors/corei5/sb/core_i5-2500_d.pdf
Dongarra J.J. Performance of various computers using standard linear equations software // Netlib. 2014. URL: http://www.netlib.org/benchmark/performance.pdf
[7] Geeks3D. AMD Radeon and NVIDIA GeForce FP32/FP64 GFLOPS Table // Geeks3D. 2014. URL: http://www.geeks3d.com/20140305/amd-radeon-and-nvidia-geforce-fp32-fp64-gflops-table-computing/
[8] Saarikoski J. et al. On the Influence of Training Data Quality on Text Document Classification Using Machine-Learning Methods // Int. J. Knowl. Eng. Data Min. 2015. № 2 (3). pp. 143–169.

tion from a coarse tool into a broadly applicable solution. For example, Google Translate utilises a vast base of user-supplied *parallel texts*, i.e. identical texts in different languages, which makes it possible to teach the system and improve the quality of translation on the go.[9]

These two factors combined cut bulk data processing times substantially, thus facilitating computations of ever greater complexity within reasonable timeframes. By early 2018, AI had reached a number of widely publicized milestones and made its way into various sectors of the economy. It would, however, be a mistake to perceive AI as a silver bullet for all of humanity's problems.

AI Breakthroughs

Expert Systems and Text Recognition: IBM Watson and Jeopardy!

Perhaps the most widely known example of AI is the IBM Watson expert system, which combines a vast bulk of knowledge (i.e. data and the associated semantic relations between them) and the capacity for processing database requests in English. In 2011, IBM Watson won an impressive victory over the reigning human champions of the TV quiz *Jeopardy!*. This achievement, which demonstrated Watson's ability to process and structure information, helped IBM break into the expert system market.

Neural Networks: Google DeepMind's Win at Go

The growth in computational performance made a significant contribution to the development of artificial neural networks, which had been conceptualized back in the mid-1940s.[10] Contemporary technology makes it possible to teach large neural networks, and the complexity of problems that can be solved directly depend on the size of the network. These *deep learning* systems stand out for their structure, which makes it possible to first recognize the local details of input data (such as the difference in the colour of neighbouring pixels) and then, as the data is processed, their global properties (such as lines and shapes). Google engineers created and trained a neural network based on such an architecture that could play the game Go. The algorithm then unexpectedly beat the reigning human world champion.[11] Other research teams achieved similar results with video games like *StarCraft II*.[12]

Practical Applications of AI

As of early 2018, machine learning, pattern recognition and autonomous planning had spilled over from research labs into the commercial market. The first users of the new technology were the militaries on either side of the Iron Cur-

[9] DeNero J., Uszkoreit J. Inducing Sentence Structure from Parallel Corpora for Reordering // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2011.
[10] McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // The Bulletin of Mathematical Biophysics. 1943. № 4 (5). pp. 115–133.
[11] Silver D. et al. Mastering the game of Go with deep neural networks and tree search // Nature. 2016. (529). pp. 484–503.
[12] Vinyals O. et al. StarCraft II: A New Challenge for Reinforcement Learning // CoRR. 2017.

tain, which had been interested in automated planning solutions ever since the 1950s.[13] Economists embraced the planning technology at roughly the same time. Listed below are several examples of how AI algorithms are currently used.

- IBM offers its Watson solution to a variety of economic sectors: health professionals can use it to diagnose patients and recommend treatment; lawyers can use it to classify specific cases as per the legislation; and railway personnel can use it to detect fatigue in tracks and rolling stock.[14]
- In medicine, pattern recognition makes it possible to identify and classify internal organs when planning surgeries.[15]
- Online shops use machine-learning mechanisms to target products to regular clients.[16]
- Autonomous robotized museum guides[17] conduct guided tours and answer visitors' questions.
- The military is introducing the first elements of autonomous decision-making technology. Short-range air-defense and anti-missile systems leave humans out of the decision-making loop due to the long reaction times of human operators, and advanced anti-ship missile systems assign targets to missiles depending on their priority.

The Prospects of AI Application

It is expected that AI will be capable of solving even more problems in the 2020s. Listed below are some of these problems, complete with the projected R&D progress.

Driverless Vehicles

The greatest problem associated with the development of driverless vehicles is to make them navigate road traffic despite the manoeuvring constraints and the multitude of possible situations on the road. As of 2018, the most developed solution in the market is the Audi A8 autopilot, which can successfully navigate traffic jams on motorways.[18] In March 2017, BMW promised to roll out a vehicle by 2021 that would be able to deliver passengers to their destinations without human intervention.[19]

---

[13] Keneally S.K., Robbins M.J., Lunday B.J. A Markov decision process model for the optimal dispatch of military medical evacuation assets // Health Care Management Science. 2016. № 2 (19). pp. 111–129.
Tsyguichko V. Models in the System of Strategic Decisions in the USSR / V. Tsygichko, Moscow: Imperium Press, 2005.

[14] AI Stories // IBM. 2018. URL: https://www.ibm.com/watson/ai-stories/index.html

[15] Alpers J. et al. CT-Based Navigation Guidance for Liver Tumor Ablation // Eurographics Workshop on Visual Computing for Biology and Medicine. 2017.

[16] Diallo A.D., Gobee S., Durairajah V. Autonomous Tour Guide Robot using Embedded System Control // Procedia Computer Science. 2015. (76). pp. 126–133.

[17] Aggarwal C.C. Recommender Systems. Heidelberg: Springer International Publishing, 2016.

[18] Ross P.E. The Audi A8: the World's First Production Car to Achieve Level 3 Autonomy // IEEE Spectrum. 2017. URL: https://spectrum.ieee.org/cars-that-think/transportation/self-driving/the-audi-a8-the-worlds-firstproduction-car-to-achieve-level-3-autonomy

[19] Reuters. BMW says self-driving car to be level 5 capable by 2021 // Reuters. 2017. URL: https://www.reuters.com/article/us-bmw-autonomous-self-driving/bmw-says-self-driving-car-to-be-level-5-capable-by-2021-idUSKBN16N1Y2

Military Applications

The most promising military application for AI would be automated target recognition and tracking capability for robotic platforms and, consequently, the ability of such ability to engage targets autonomously.[20] As of late 2017, first-tier countries were conducting a number of research efforts in this field, primarily as applied to ground-based, sea-surface and submarine weapon systems.

Robotic Surgery

Surgery (including microsurgery, which requires precise manipulations) is expected to become automated in the near future. The robotized technology demonstrator STAR presented in 2016[21] is capable of performing soft-tissue surgeries. Further achievements in pattern recognition could result in all soft tissue surgeries becoming robotized. This would make surgical interventions more affordable while reducing the workload of personnel.

Problems Still Unsolved

Despite computer algorithms appearing to be omnipotent, there are still problems that they either cannot solve or are poor at solving. In particular, it is difficult, for obvious reasons, to make them solve problems that are hard to formalize, such as writing a novel or picking the most beautiful photo from a selection of pictures. In fact, even problems that can be formalized with mathematical precision do not necessarily yield themselves to AI solutions with guaranteed success. This may be due to the complexity of mathematical modelling as applied to lower-level tasks (say, modelling a robot's movements when teaching it to play football), the complexity of the problem itself (for example, there are no algorithms significantly better for making logical conclusions and proving mathematical statements than the complete exhaustion of all possible logical chains[22]), the vast number of parameters and the imprecision of the observable world (e.g. as applied to football), as well as the insufficient capacity of computational systems as compared to the human brain. Indeed, the interaction of the brain's 1×10^{11} neurons cannot be simulated by algorithms. As of early 2018, the best science had come up with was the simulation of $1.7\text{x}10^{9}$ neurons with a 2400-fold delay in 2013.[23] In fact, even simulating the required number of brain cells may not necessarily replicate actual brain activity in a computer model.

The drawbacks of machine learning merit special mention. As a rule, machine learning requires the existence of pre-classified training data that an algorithm would analyse for patterns. Data shortages may result in situations in which input data does not belong to any of the classes learnt by the algorithm. The recognition of a new phenomenon at input and the creation of a new class of ob-

[20] Boulanin V., Verbruggen M., RIKEN. Op. cit.
[21] Shademan A. et al. Supervised autonomous robotic soft tissue surgery // Science Translational Medicine. 2016. № 337 (8). p. 337.
[22] Hopcroft, John E.; Motwani, Rajeev; Ullman, Jeffrey D. Introduction to Automata Theory, Languages, and Computation. London: Pearson, 2013.
[23] RIKEN, Forschungszentrum Jülich. Largest neuronal network simulation achieved using K computer // RIKEN. 2013. URL: http://www.riken.jp/en/pr/press/2013/20130802_1/

jects is a difficult task to solve at best,[24] and will become even more complex if the algorithm is made to learn actively during the classifier's operation and the recognized classes are rendered temporarily changeable. One other significant shortcoming of machine learning is the extreme sensitivity of algorithms: for example, it is possible to cheat facial recognition software by putting on a pair of glasses.[25] In some instances, a photograph may be classified erroneously if it has been subjected to changes imperceptible to the human eye: for instance, a seemingly insignificant manipulation with a photo of a panda may lead an algorithm to mistake the animal for a monkey.[26]

Computers may be successful at solving such "complex" problems as symbolic and numerical computations and even at beating chess grandmasters, but many fairly simple problems remain unsolved. These include classifying "unknown" images without prior training with the use of pre-classified samples (such as recognizing images of apples for an algorithm that is only aware of cherries and pears), motility and rational reasoning. This phenomenon is known as Moravec's paradox.[27] To a great extent it reflects human perception: the abilities commanded by any adult human being, those which are based on millions of years of evolution, are taken for granted, whereas mathematical problems like finding the shortest route on a map appear unnatural, and their solutions nontrivial.

AI Myths

The 50 years of research into AI have given birth to a multitude of myths and misconceptions, related to both the technology's capabilities and to its shortcomings. For example, one statement perpetuated in the field of machine translation is that a low-quality translation tool from the 1960s turned the phrase "The spirit is strong but the flesh is weak" into "The vodka is good but the meat has gone rotten" when translated into Russian and then back into English. In fact, this particular example was originally used to illustrate the incorrect translation made by a human who was only armed with a dictionary, a basic knowledge of grammar and a runaway imagination; there is no hard evidence that machine translation software has ever performed in this manner.[28]

Let us enumerate several myths which distort the public perception of AI.

AI is Magic

"Any sufficiently advanced technology is indistinguishable from magic," Arthur C. Clarke once said. Fair enough, but contemporary AI research is based on mathematics, robotic technology, statistics and information technology. From what is mentioned earlier in this article, it follows that any software or hardware–

[24] Scheirer W.J., Jain L.P., Boult T.E. Probability Models for Open Set Recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI). 2014. № 11 (36).
[25] Sharif M. et al. Adversarial Generative Nets: Neural Network Attacks on State-of-the-Art Face Recognition // CoRR. 2017.
[26] Goodfellow I., Shlens J., Szegedy C. Explaining and Harnessing Adversarial Examples // International Conference on Learning Representations. 2015.
[27] Moravec H. Mind Children: The Future of Robot and Human Intelligence. MA: Harvard University Press. 1990.
[28] Hutchins J. «The whisky was invisible», or Persistent myths of MT // MT News International. 1995. (11). pp. 17–18.

software complex incorporating AI capability is primarily intended to solving several mathematical problems, often based on pre-collected and pre-classified data.

AI is a Computer Brain

The perception of AI as an anthropomorphic mechanism, one similar in many respects to the human brain, is firmly ingrained in contemporary culture. In reality, however, even neural networks are a mathematical abstraction that has very little in common with the biological prototype. Other models, like Markov chains, have no biological equivalents whatsoever. The existing algorithms are not aimed at modelling human consciousness; rather, they specialize in solving specific formalized problems.

There are a Limited Number of Applications for AI

According to the stereotype fed by popular culture, AI is rarely involved in anything but planning in the interests of major corporations or governments. In fact, the aforementioned capabilities allow AI technologies to be harnessed for the benefit of broad strata of the population: everyone can avail of the algorithms used in real life, and the range of problems addressed by machine learning includes fairly routine assignments like assessing the need for repairs across a taxi cab fleet or planning the optimal use of bank teller halls.

AI-Related Dangers

As a rule, apocalyptic examples of science fiction paint a gloomy picture of AI deciding, at some point, to destroy humankind, with scientists either being unable to prevent this scenario or remaining blissfully unaware of the danger. In real life, the scientific community has long been discussing the hazards associated with super-intelligent software[29]. The key identified risk is believed to consist in the incorrect tasking of a strong AI that would command considerable computational power and material resources while disregarding the interests of people. The aforementioned work by Nick Bostrom offers possible solutions to this problem.

The Reality of Strong AI

The excessively high or low expectations of scientific and technical progress result in conflicting forecasts of a strong AI becoming available any day now (next year at the latest) – or never. In reality, the creation of a strong AI is difficult to predict: this depends on whether or not several engineering and mathematical problems of unknown complexity will be solved successfully. As illustrated by the Fermat Last Theorem, the solution to which only emerged 350 years it was first proposed, no one can offer any exact date for the creation of a strong AI.

International Trends

The expansion of AI's range of applications has prompted an increased interest of military and security circles in the capabilities of autonomous systems. Research and development related to the aforementioned possible applications has

[29] Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford, UK: Oxford University Press, 2014.

provoked heated international debates on the expedience of restricting or even banning robotized systems altogether. The Campaign to Stop Killer Robots has been particularly prominent in this respect:[30] its supporters demand a full ban on the development of autonomous combat systems out of ethical considerations. In this light, it is worth mentioning not just combat robots, but also classification systems, which inform decisions to use force based exclusively on metadata, without any regard for even the content of suspects' messages.[31]

This heightened public attention on autonomous combat systems has resulted in talks in the framework of the *UN Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*. The document covers anti-personnel mines and blinding lasers. However, as of early 2018, the diplomatic process had not resulted in any mutual obligations.[32] This is to a great extent due to the difficulty in defining the notion of an autonomous combat system, the existence of systems (primarily air-defense and anti-missile systems) which meet the possible definition of such systems, and the unwillingness of governments to give up on the promising technology.[33] Nevertheless, it would be an oversimplification to restrict the arguments in favour of establishing control over autonomous weapons exclusively to the domain of morality and ethics. Strategic stability is also at risk:[34] first, the use of autonomous systems could lead to uncontrolled escalation of armed conflicts with unpredictable consequences; second, exports of such systems would be difficult to control. Indeed, the handover of a cruise missile is hard to conceal, and its range is limited by its physical properties, whereas control over software code is impossible: the use of AI algorithms is not limited to military applications, and restrictions on research into software with such a broad range of possible applications, possibly including a ban on scientific publications, would inevitably be resisted by the scientific community, which largely survives thanks to international cooperation and major conferences.

[30] Campaign to Stop Killer Robots. URL: https://www.stopkillerrobots.org/
[31] Former CIA, NSA Director: 'We Kill People Based on Metadata' // RT in Russian. 2014. URL: https://russian.rt.com/article/31734
[32] The United Nations Office at Geneva. CCW Meeting of High Contracting Parties. 2017. URL: https://www.unog.ch/80256EE600585943/httpPages)/A0A0A3470E40345CC12580CD003D7927?OpenDocument
[33] Boulanin V., Verbruggen M., Op. cit.
[34] Altmann J., Sauer F. Autonomous Weapon Systems and Strategic Stability // Survival. 2017. № 5 (59). pp. 117–142.

Artificial Intelligence: A Blessing or a Threat for Humanity?

A. Kolonin

The Main Problems and Potential Threats of AI Development

Despite the multitude of advances in neuroscience, we still do not know exactly how natural intelligence works. For this same reason, we do not know for sure how to create artificial intelligence (AI). There are a number of known problems that need to be resolved, as well as differing opinions as to how these problems should be prioritized. For example, Ben Goertzel, who heads the OpenCog and SingularityNET, open-source international projects to create artificial intelligence, believes that all the requisite technology for creating an artificial general intelligence has already been developed, and that the only thing necessary is to combine them in a way that would ensure the necessary synergy.[35] Other experts are more sceptical, pointing out that many of the problems that we will discuss below need to be resolved first. Moreover, expert estimates for when a strong AI may be created vary greatly, from ten or so years to several decades from now.

Even the emergence of autonomous or adaptive AI systems, let alone general or strong AI, is associated with several threats of varying degrees of severity that are relevant today. These include:

- the creation of so-called lethal autonomous weapons systems (LAWS), such as unmanned vehicles that could be used in targeted assassinations;
- a new round of the arms race, with countries improving the intelligence levels of autonomous weapons;
- the creation of an intelligent system (not necessarily a combat system, but one that could have industrial or domestic applications too) that would be capable not only of deliberate activity, but also of autonomous conscious target-setting, which could run counter to human goals;
- increasing automation, driving most of the population out of the material production sectors, which could result in even greater social stratification, the reduced effectiveness of "social elevators," and an increase in the numbers of people made redundant, with adverse social consequences.
- The problems of controlling artificial intelligence systems are currently associated, among other things, with the closed nature of the existing applications, which are based on "deep neural networks." Such applications do not make it possible to validate the correctness of decisions prior to implementa-

[35] Ben Goertzel, Cassio Pennachin, Nil Geisweller, Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy (Atlantis Thinking Machines), Atlantis Press. 2014.

tion, nor do they allow for an analysis of the solution provided by the machine after the fact. This phenomenon is being addressed by the new science, which explores explainable artificial intelligence (XAI). The process is aided by a renewed interest in integrating the associative (neural) and symbolic (logic-based) approaches to the problem.

In August 2018, Czech Technical University in Prague simultaneously hosted several conferences on AI-related topics: human-level AI,[36] artificial general intelligence,[37] biologically inspired cognitive architectures,[38] and neural-symbolic integration technology.[39] Reports were presented by prominent experts representing global leaders in artificial intelligence: Microsoft, Facebook, DARPA, MIT and Good AI. The reports described the current status of AI developments, identified the problems facing society that have yet to be resolved, and highlighted the threats arising from the further development of this technology. In this review, we will attempt to briefly identify the main problems and threats, as well as the possible ways to counter these threats.

To begin with, let us provide definitions for some of the terms that are commonly used in conjunction with AI in various contexts: *weak, or specialized, AI; autonomous AI; adaptive AI; artificial general intelligence (AGI); strong AI; human-level AI; and super-human AI*.

Weak, or specialized, AI is represented by all existing solutions without exception and implies the automated solution of one specific task, be it a game of Go or face recognition with CCTV footage. Such systems are incapable of independent learning for the purpose of solving other problems: they can only be reprogrammed by humans to do so.

Autonomous AI implies a system's ability to function for protracted periods of time without the intervention of a human operator. This could be a solar-powered UAV performing a multi-day flight from Champs-Elysees in Paris to Moscow's Red Square or back, independently selecting its route and recharging stops while avoiding all sorts of obstacles.

Adaptive AI implies the system's ability to adapt to new situations and obtain knowledge that it did not possess at the time of its creation. For example, a system originally tasked with conducting conversations in Russian could independently learn new languages and apply this knowledge in conversation if it found itself in a new language environment or if it deliberately studied educational materials on these new languages.

Artificial general intelligence implies adaptability of such a high level that the corresponding system could, given the appropriate training, be used in a wide variety of activities. New knowledge could either be self-taught or learned with

[36] Human Level AI. The Joint Multi-Conference on Human-Level Artificial Intelligence // HLAI. 2018. URL: https://www.hlai-conf.org/
[37] Conference Series on Artificial General Intelligence // AGI. 2018. URL: http://agi-conference.org/
[38] Biologically Inspired Cognitive Architectures Society // BICA Society. 2018. URL: http://bicasociety.org/
[39] Neural-Symbolic Integration. Workshop series on Neural-Symbolic Learning and Reasoning // Neural Symbolic Integration. 2018. URL: http://daselab.cs.wright.edu/nesy/

the help of an instructor. It is in this same sense that the notion of strong AI is often used in opposition to weak or specialized AI.

Human-level AI implies a level of adaptability comparable to that of a human being, meaning that the system is capable of mastering the same skills as a human and within comparable periods of time.

Super-human AI implies even greater adaptability and learning speeds, allowing the system to masker the knowledge and skills that humans would never be able to.

Fundamental Problems Associated with Creating a Strong AI

The emergence of a strong AI is logical in the framework of the general process of evolution as the emergence of molecules from atoms and cells from molecules, the creation of the central nervous system from specialized cells, the emergence of social structure, the development of speech and writing systems and, ultimately, the nascence of information technology. Valentin Turchin demonstrates the logic behind the increasing complexity of information structures and organizational mechanisms in the process of evolution.[40] Unless humanity perishes first, this evolution will be inevitable and will, in the long run, rescue humankind, as only non-biological lifeforms will be able to survive the inevitable end of the Solar System and preserve our civilization's information code in the Universe.

It is important to realize that the creation of a strong AI does not necessarily require the understanding of how the natural intelligence works, just as the development of a rocket does not necessarily require understanding how a bird flies. Such an AI will certainly be created, sooner or later, in one way or another, and perhaps even in several different ways.

Most experts identify the following fundamental problems that need to be solved before a general or strong AI can be created:[41]

- **Few-shot learning**: systems need to be developed that can learn with the use of a small amount of materials, in contrast to the current deep-learning systems, which require massive amounts of specifically prepared learning materials.
- **Strong generalization**: creating problem recognition technologies allowing for recognizing objects in situations that differ from those in which they were encountered in the learning materials.
- **Generative learning models**: developing learning technologies in which the objects to be memorized are not the features of the object to be recognised, but rather the principles of its formation. This would help in addressing the more profound characteristics of objects, providing for faster learning and stronger generalization.
- **Structured prediction and learning**: developing learning technologies based on the representation of learning objects as multi-layered hierarchical structures,

[40] Turchin V. The Phenomenon of Science: A Cybernetic Approach to Human Evolution. Moscow: ETS, 2000. URL: http://refal.net/turchin/phenomenon/

[41] Iklé M., Franz A., Rzepka R., Goertzel B. (Eds.), Artificial General Intelligence, 11th International Conference, AGI 2018, Prague, Czech Republic. 2018. URL: https://www.springer.com/us/book/9783319976754

with lower-level elements defining higher level ones. This could prove an alternative solution to the problems of fast learning and strong generalization.

- Solving the problem of **catastrophic forgetting**, which is pertinent to the majority of existing systems: a system originally trained with the use of one class of object and then additionally trained to recognize a new class of objects loses the ability to recognize objects of the original class.
- Achieving an **incremental learning ability**, which implies a system's ability to gradually accumulate knowledge and perfect its skills without losing the previously obtained knowledge, but rather obtaining new knowledge, with regard to systems intended for interaction in natural languages. Ideally, such a system should pass the so-called Baby Turing Test by demonstrating its ability to gradually master a language from the baby level to the adult level.[42]
- Solving the **consciousness** problem, i.e. coming up with a proven working model for conscious behaviour that ensures effective prediction and deliberate behaviour through the formation of an "internal worldview," which could be used for seeking optimum behavioural strategies to achieve goals without actually interacting with the real world. This would significantly improve security and the testing of hypotheses while increasing the speed and energy efficiency of such checks, thus enabling a live or artificial system to learn independently within the "virtual reality" of its own consciousness. There are two applied sides to solving the consciousness problem. On the one hand, creating conscious AI systems would increase their efficiency dramatically. On the other hand, such systems would come with both additional risks and ethical problems, seeing as they could, at some point, be equated to the level of self-awareness of human beings, with the ensuing legal consequences.

Potential AI-Related Threats

Even the emergence of autonomous or adaptive AI systems, let alone general or strong AI, is associated with several threats of varying degrees of severity that are relevant today.

The first threat to humans may not necessarily be presented by a strong, general, human-level or super-human AI, as it would be enough to have an autonomous system capable of processing massive amounts of data at high speeds. Such a system could be used as the basis for so-called lethal autonomous weapons systems (LAWS), the simplest example being drone assassins (3D-printed in large batches or in small numbers).[43]

[42] Partee B.H., Peters S., Thomason R. (Eds.), Report of Workshop on Information and Representation. Washington. D.C.. March 30 through April 1. 1985. URL: https://files.eric.ed.gov/fulltext/ED261533.pdf

[43] Metro. Killer robots are fast becoming a reality – we must stop this from happening if we want to stop a global AI arms race // Metro. 2018.
URL: https://metro.co.uk/2018/09/02/killer-robots-are-fast-becoming-a-reality-we-must-stop-this-fromhappening-if-we-want-to-stop-a-global-ai-arms-race-7903717/?ito=cbshare

Second, a threat could be posed by a state (a potential adversary) gaining access to weapons system based on more adaptive, autonomous and general AI with improved reaction times and better predictive ability.

Third, a threat for the entire world would be a situation based on the previous threat, in which several states would enter a new round of the arms race, perfecting the intelligence levels of autonomous weapon systems, as Stanislaw Lem predicted several decades ago.[44]

Fourth, a threat to any party would be presented by any intellectual system (not necessarily a combat system, but one that could have industrial or domestic applications too) with enough autonomy and adaptivity to be capable not only of deliberate activity, but also of autonomous conscious target-setting, which could run counter to the individual and collective goals of humans. Such a system would have far more opportunities to achieve these goals due to its higher operating speeds, greater information processing performance and better predictive ability. Unfortunately, humanity has not yet fully researched or even grasped the scale of this particular threat.

Fifth, society is facing a threat in the form of the transition to a new level in the development of production relations in the capitalist (or totalitarian) society, in which a minority comes to control material production and excludes an overwhelming majority of the population from this sector thanks to ever-growing automation. This may result in greater social stratification, the reduced effectiveness of "social elevators" and an increase in the numbers of people made redundant, with adverse social consequences.

Finally, another potential threat to humanity in general is the increasing autonomy of global data processing, information distribution and decision-making systems growing, since information distribution speeds within such systems, and the scale of their interactions, could result in social phenomena that cannot be predicted based on prior experience and the existing models. For example, the social credit system currently being introduced in China is a unique experiment of truly civilizational scale that could have unpredictable consequences.[45]

The problems of controlling artificial intelligence systems are currently associated, among other things, with the closed nature of the existing applications, which are based on "deep neural networks." Such applications do not make it possible to validate the correctness of decisions prior to implementation, nor do they allow for an analysis of the solution provided by the machine after the fact. This phenomenon is being addressed by the new science, which explores explainable artificial intelligence (XAI).[46] The process is aided by a renewed interest in integrating the associative (neural) and symbolic (logic-based) approaches to the problem.

[44] Lem S., Peace on Earth, Operation Eternity Almanac 1988, Moscow.: Mir, 1988.
AI Safety Research: A Road to Nowhere // Medium, 2016.
URL: https://medium.com/@petervoss/ai-safety-research-a-road-to-nowhere-f1c7c20e8875

[45] Kovachich L. Big Brother 2.0. How China is Building a Digital Dictatorship // Moscow Carnegie Center, 2017.
URL: https://carnegie.ru/commentary/71546

[46] Explainable Artificial Intelligence (XAI) // 2017. DARPA/I2O.
URL: https://www.darpa.mil/attachments/XAIProgramUpdate.pdf

Ways to Counter the Threats

It appears absolutely necessary to take the following measures in order to prevent catastrophic scenarios associated with the further development and application of AI technologies.

- An international ban on LAWS, as well as the development and introduction of international measures to enforce such a ban.
- Governmental backing for research into the aforementioned problems (into "explainable AI " in particular), the integration of different approaches, and studying the principles of creating target-setting mechanisms for the purpose of developing effective programming and control tools for intellectual systems. Such programming should be based on values rather than rules, and it is targets that need to be controlled, not actions.
- Democratizing access to AI technologies and methods, including through reinvesting profits from the introduction of intellectual systems into the mass teaching of computing and cognitive technologies, as well as creating open-source AI solutions and devising measures to stimulate existing "closed" AI systems to open their source codes. For example, the Aigents project is aimed at creating AI personal agents for mass users that would operate autonomously and be immune to centralized manipulations.[47]
- Intergovernmental regulation of the openness of AI algorithms, operating protocols for data processing and decision-making systems, including the possibility of independent audits by international structures, national agencies and individuals. One initiative in this sense is to create the SingularityNET open-source platform and ecosystem for AI applications.[48]

[47] Aigents // Aigents. URL: https://aigents.com/
[48] SingularityNET // SingularityNET. https://singularitynet.io/

AI Nationalism and AI Nationalization are Ahead

S. Karelov

AI as a Geopolitical Factor

Technological inequality has always had a major impact on global politics and the world economy. The most technologically advanced states became the most successful, gained undisputed military superiority and begin to impose their will onto less developed countries.

In the 20th century, two world wars and the extraordinary acceleration of scientific and technological progress made this impact even greater. With the third decade of the 21st century approaching, the avalanche-like growth of machine learning capabilities has forced us to talk about the spectrum of IT technologies, unified by the metaphorical term "Artificial Intelligence" (AI) as the key factor in economic, geopolitical and military power of the coming decades.

2018 marked a watershed. Previously, when the media, society and politicians discussed the future challenges and dangers of AI development, they focused on:

- cognitive AI agents (robots and software) supplanting people in many professions;
- the legal and ethical problems of autonomous AI devices (for instance, driverless cars);
- cybersecurity issues;
- the frighteningly alluring prospect of a "revolt of the machines" in the distant, or perhaps not so distant future.

However, things began to change in 2018. The above-listed threats and challenges did not exactly disappear, but rather faded into the background as politicians and the military became aware of the **global transformational trends** that can be provisionally termed

- AI nationalism
- AI nationalization

These trends were highlighted by the fact that, in drafting their national AI strategies, many developed countries simultaneously started to change their attitudes towards two basic principles that had previously seemed unshakeable:

- instead of comprehensive international cooperation, the global division of labour, the introduction of open platforms and the mutual overflow of talent, countries are now placing an emphasis on **AI nationalism**, which proclaims the priority of the economic and military interests of one's own country as the principal objective of its national AI strategy;

- instead of the separation of state and business that is traditional for many countries, the course has been set for
- **AI nationalization**, i.e. integrating governmental and private resources, aligning the pace of introducing AI innovations and refocusing strategic objectives on the state gaining economic, geopolitical and military advantages in the international arena.

The strengthening of these trends promotes the shift of state priorities in developed countries **away from the economy and towards geopolitics**. If this trend continues, the world will undergo major changes in the near future.

A Brief Technological and Geopolitical Forecast

First, the **geopolitical and military doctrines** of most developed countries **are going to change**. Many economic and political processes brought together under the term **globalization** will undergo radical transformations. It is possible that AI technologies will put an end to globalization, a trend that has been picking up pace following the fast technological development that came after World War II. And then **global Balkanization** will replace globalization.

Second, **the change in the model of co-existence of the state and business** will have an equally profound impact. The integration of the goals and resources of the state and private businesses in order to achieve military superiority will most likely result in the global triumph of the **authoritarian-democratic model** that China has already added to its armoury, both literally and figuratively.

As a result, international relations will be determined by the following key factors:

AI neo-colonialism in relations between AI leaders and outsiders;

An AI arms race between the leading countries that will guide and determine subsequent development of AI technologies.

Such race could have two possible outcomes:

A. **"World War III,"** after which World War IV will be fought with sticks and stones.

B. **The "AI Singularity,"** where the development of AI technologies, spurred on by the arms race, will end up generating a "**strong AI**" that will remove the opposing parties as unnecessary.

The first option appears to be the most probable, with the second option also being a possibility.

It is important to note that, at some point, the AI nationalism and AI nationalization trends will be driving their own development. As with any other arms race, the pace and strength of these trends will no longer depend on the degree of AI progress. And even if AI progress turns out to be more modest than expected, the gigantic inertia of both trends will require much time and effort to be overcome.

2018: The Year of the Great AI Watershed

After the end of the Cold War, the United States held a virtually unsurpassed superpower status. The crucial factor was its unrivalled military and technological superiority. Nonetheless, technologies that had previously formed the foundation of the country's military defense, such as high-precision weapons, have spread throughout the world due to globalization and technology transfer. As a result, the rivals of the United States have developed their own capabilities that offer a progressively greater challenge to U.S. military superiority.

To retain and expand its military edge in future, the U.S. Department of Defense banked on AI technologies, the potential military use of which is highly diversified: from improving efficiency of logistical systems to more sensitive tasks, such as automated control and monitoring in modern weapons systems. The 2018 National Defense Strategy proceeds on the assumption that AI will mostly likely change the nature of warfare. That is why U.S. Deputy Secretary of Defense Patrick Shanahan states that the United States "must pursue AI applications with boldness and alacrity."[49]

There are grounds to believe that the U.S. government has finally heard Eric Schmidt, former head of Google and Alphabet, who called upon it last year to realize that the "Sputnik Moment" in AI had already arrived.[50]

Secretary of Defense **James Mattis** called upon President Trump to draft a national artificial intelligence development strategy both for the U.S. government and for the entire country. Mattis' letter to the President contained a suggestion on the establishment of a presidential committee capable of "inspiring a whole of country effort that will ensure the U.S. is a leader not just in matters of defense but in the broader 'transformation of the human condition."

In response to those statements,

- The "Summary of the 2018 White House Summit on Artificial Intelligence for American Industry" states that "the Trump Administration **has prioritized funding** for fundamental AI research and computing infrastructure, machine learning, and autonomous systems;"
- on July 31, 2018, the White House released Memorandum No. M-18-22: FY 2020 Administration Research and Development Budget Priorities[51] naming AI **the top of three national technological priorities** (followed by quantum informatics and supercomputing, respectively) and mandated that the Office of Management and Budget (OMB) jointly with the Office of Science and Technology Policy (OSTP) ensure **top budgeting priority** for these areas in all federal agencies in 2019–2020.

[49] Memorandum: Establishment of the Joint Artificial Intelligence Center (JAIC). 2018. URL: https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence_center_osd008412-18_r....pdf

[50] Our Artificial Intelligence 'Sputnik Moment' Is Now: Eric Schmidt & Bob Work // Breaking Defense. 2017. URL: https://breakingdefense.com/2017/11/our-artificial-intelligence-sputnik-moment-is-now-eric-schmidt-bob-work/

[51] Memorandum: FY 2020 Administration Research and Development Budget Priorities // The White House. 2018. URL: https://www.whitehouse.gov/wp-content/uploads/2018/07/M-18-22.pdf

It so happened that it was not just the U.S. government that experienced a "Sputnik Moment" in 2018.

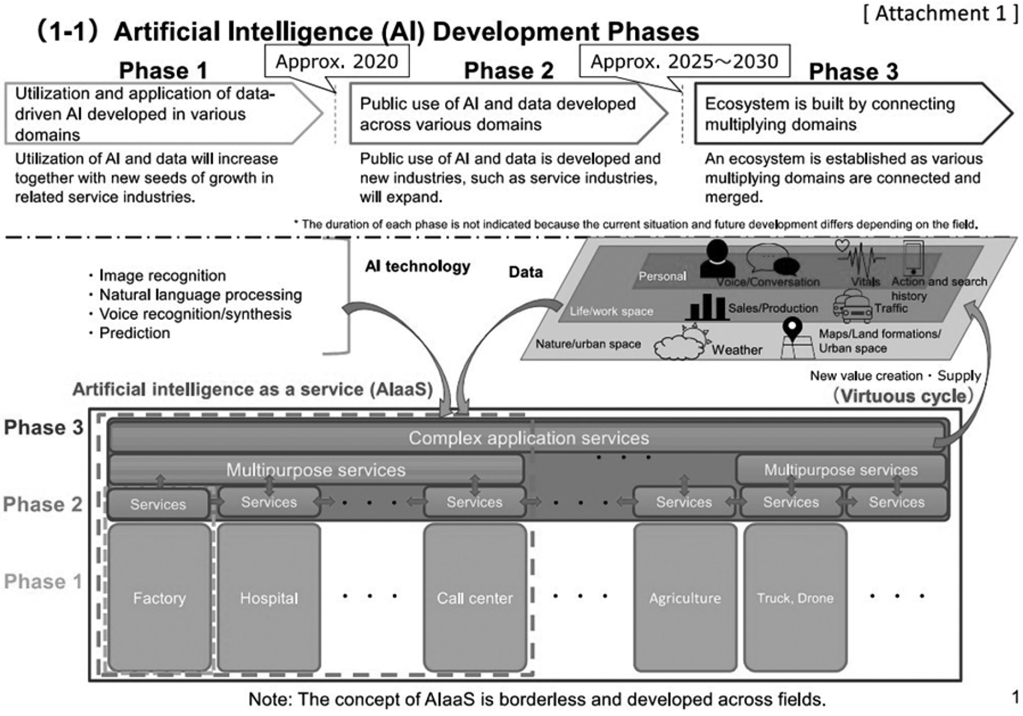
The number of countries announcing national AI development strategies more than tripled compared to 2017.

In 2017, AI enthusiasts (Japan, Canada and Singapore) inaugurated the marathon of signing national AI strategies. The trio was followed by China, one of the two claimants to global AI leadership, which deployed its main financial weapon.

Now, in 2018, it is much easier to pinpoint those who have not yet publicized their national AI strategy, since the United Kingdom, Germany, France and a dozen more countries have already presented theirs. As it should have been expected, different countries have set forth different goals and approaches in their national AI strategies.

The strategies that have been drafted upon the instructions of national governments are also different in length and the degree of their elaboration and detail.

Japan's national strategy is the most laconic (25 pages) and specific[52] and includes a basic description of the three stages in establishing a national AI as a service by 2030.



Three phases of establishing a national AI as a service in Japan by 2030

[52] Strategic Council for AI Technology. Artificial Intelligence Technology Strategy. 2017. URL: http://www.nedo.go.jp/content/100865202.pdf

A set of roadmaps for the three priority areas (productivity; health, medical care and welfare; and mobility) is attached to the basic outline.

In addition, Japan has also developed a plan for integrating AI technologies with technologies in principal economic sectors in the three priority areas. The three phases of this integration demonstrate the level of planned technological progress and social changes.

The UK approach appears to be most detailed and elaborated, with the following documents being drafted:

- The 180-page "AI in the UK: Ready, Willing and Able?" report[53] published by the Authority of House of Lords in April 2018, which is largely based on the 77-page report "Growing the Artificial Intelligence Industry in the UK"[54] by Dame Wendy Hall, Professor of Computer Science at the University of Southampton, and Dr. Jerome Pesenti, CEO of BenevolentAI
- Two volumes of written and oral evidence by professionals and experts confirming the contents of the report (Volume 1 contains 223 pieces of written evidence on 1581 pages;[55] Volume 2 contains 57 pieces of oral evidence obtained at 22 sessions and set forth on 423 pages[56])
- A 40-page Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK[57] presented to Parliament by the Secretary of State for Business, Energy and Industrial Strategy by Command of Her Majesty

What are the common features of the 20-plus national visions of the future AI, visions that are already set forth by states in a variety of forms, from plans and roadmaps (Japan) to "court rulings" giving specific commands to the national government (the United Kingdom)?

Having carefully studied the entire set of documents published as of end of August 2018, the author has distinguished two key trends that are clearly manifested in all "national strategies":

1. Setting a course for AI nationalism
2. Setting a course for AI nationalization

Now we need to explain what these trends mean.

---

[53] Strategic Council for AI Technology. AI in the UK: ready, willing and able? // UK Parliament. 2018. URL: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[54] Hall W., Pesenti J. Growing the Artificial Intelligence Industry in the UK URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

[55] Strategic Council for AI Technology. Collated Written Evidence Volume // UK Parliament. 2018. URL: https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Written-Evidence-Volume.pdf

[56] The Secretary of State for Business, Energy and Industrial Strategy by Command of Her Majesty. Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK: Ready, Willing and Able? // UK Parliament. 2018. URL: https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response2.pdf

[57] Strategic Council for AI Technology. Collated Oral Evidence Volume // UK Parliament. 2018. URL: https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Oral-Evidence-Volume.pdf

AI Nationalism

The First Source of AI nationalism: Ultra-High Expectations

The term AI nationalism became popular immediately after Ian Hogarth published an essay of the same title.[58] The essay defines AI nationalism as *a new kind of geopolitics driven by continued rapid progress in machine learning in developed countries.*

AI generates a **new kind of instability** on national and international levels, forcing the governments of developed countries to act so as not to find themselves among the losers in the new competition for AI superiority.

This competition is unique and unlike anything in the past, including the race for the nuclear bomb and intercontinental ballistic missiles.

AI is unique due to three factors, one economic and two military:

1. AI tools are universal as a means for increasing efficiency in virtually all post-industrial sectors and activities (the closest example of such universality is introducing electricity everywhere).
2. Based on prior military experience, AI's is assumed to have tremendous potential for revolutionary breakthroughs
   - both in developing radically new classes of comprehensive military technological solutions (such as stealth technologies),
   - and in building radically more advanced AI-based military intelligence and managing military logistics and combat throughout the theatre of operations (including the change of customary paradigms for specific branches of the Armed Forces, as happened with air carriers transformed from a transport for intelligence aircraft equipped with deck guns into a floating airfield that is super-efficient in handling independent military tasks).
3. The assumed capability (not yet proven, but taken seriously by many military officials) of resolving the nuclear containment problem favouring a particular side (following the old cowboy wisdom of outdrawing your enemy).

These factors are largely hypothetical. They do not reflect real capabilities of the current AI technologies, but are rather extrapolations into the near future, provided that AI technologies continue to develop at the current pace.

In other words, all three factors that transform AI into a hypothetical means of gaining superiority in the international arena are merely expectations of the military and politicians.

But that does not prevent them from claiming that the world is on the threshold of **a new singularity – namely, a military singularity**. The Center for a New American Security (CNAS) report "Battlefield Singularity"[59] demonstrates how seriously the United States and China, leaders in the AI race, treat this matter. What

[58] Hogarth I. AI Nationalism // Ian Hogarth. 2018. URL: https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism

is striking about this report is not only its logic and analytics, but the gigantic list of 334 U.S. and Chinese sources it cites.

The Second Source of AI Nationalism: Technological Entanglement

The unprecedented importance of AI can make the policies in this area the key element of government policy. However, a very unpleasant problem of **technological entanglement**, unique to AI technologies,[60] gets in the way.

Technological entanglement means that globalization drives intertwining the interests and resources of international companies in developing dual-use technologies (peaceful and military). This intertwining is emerging and strengthening as a tight-knit and multi-faceted phenomenon.

Technological entanglement results in technological sovereignty virtually disappearing. Even the United States, the undisputed technological leader in AI, finds itself in a very difficult situation due to technological entanglement.

How can the United States retain its leading positions in AI and prevent their technologies from leaking to foreign rivals if:

- private business is the technological engine of the country;
- under globalization, international private giants, such as Alphabet, site their AI research labs in their chief rival (China) and hire thousands of professionals from that country;
- Chinese investors, including China's largest AI companies, own significant shares in many promising AI companies in Silicon Valley; and
- thousands of Chinese students and graduate students study in the best United States universities?

The latest and most decisive turn in technological entanglement was China announcing its **National Strategy for Military-Civil Fusion**.[61]

President of the Information Technology and Innovation Foundation **Robert Atkinson** summarized[62] the situation that resulted from technological entanglement in his article published in *National Review* on July 26, 2018.

- China has deployed a vast panoply of "innovation mercantilist" practices that seek to unfairly advantage Chinese producers, including:
  - requiring foreign companies to transfer their technologies to Chinese firms in order to access the Chinese market;

[59] Kania E.B. Battlefield Singularity // CNAS. 2017.
URL: https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power

[60] Technological entanglement. Cooperation, competition and the dual-use dilemma in artificial intelligence. The Australian Strategic Policy Institute // ASPI. 2018.
URL: https://www.aspi.org.au/report/technological-entanglement

[61] Levesque G., Stokes M. Blurred Lines: Military-Civil Fusion and the "Going Out" of China's Defense Industry // Pointe Bello. 2016.
URL: https://static1.squarespace.com/static/569925bfe0327c837e2e9a94/t/593dad0320099e64e1ca92a5/1497214574912/062017_Pointe+Bello_Military+Civil+Fusion+Report.pdf

[62] Atkinson R.D. Who Lost China? // National Review. 2018.
URL: https://www.nationalreview.com/magazine/2018/08/13/us-china-relations-who-lost-them/

  - theft of foreign intellectual property;
  - manipulation of technology standards;
  - massive government subsidies; and
  - government-backed acquisitions of foreign enterprises.
- The U.S. government should have one and only major trade-policy goal at this moment: rolling back China's innovation-mercantilist agenda, which threatens the United States' national and economic security.

The AI is expected to bring incredible returns ensuring a country's superiority on the international arena. This and the exacerbating technological entanglement were the principal sources of technological and nationalist agenda of several countries, and their number is growing.

The potential consequences of this include: various protectionist measures by states in order to support "national AI champions;" restrictions (and the possible prohibition) on transferring patents, open research publications, and exporting AI technologies; and also restrictions (and the possible prohibition) on M&A transactions, the free overflow of investment and, of course, talent both in training and in professional activities.

The Avant-Garde of AI Nationalism

As of the end of August 2018, China, the United States, India, the United Kingdom and France announced their intention to pursue some form of AI nationalism. These five countries have also announced the development of AI-based **national military programs**.

- China: within the framework of the 13th Five-Year Plan,[63] the "Artificial Intelligence 2.0" program of the Plan of Innovative Scientific and Technical Development of the 13th Five-Year Plan,[64] and the Three-Year Action Plan for Promoting Development of Next Generation Artificial Intelligence Industry (2018–2020)[65] – under the umbrella of the national Military-Civil Fusion strategy.[66]
- The United States: within the framework of the National Security Strategy,[67] the National Defense Strategy[68] and the Memorandum on the Establishment of the Joint Artificial Intelligence Center (JAIC) of the Department of Defense (JAIC).[69]

[63] 13th Five-Year Plan. State Council of the People's Republic of China, 2016. URL: http://www.gov.cn/xinwen/2016-03/17/content_5054992.htm

[64] Plan of Innovative Scientific and Technical Development of the 13th Five-Year Plan State. Council of the People's Republic of China, 2016. URL: http://www.gov.cn/zhengce/content/2016-08/08/content_5098072.htm

[65] Three Year Action Plan Focuses On Next generation Artificial Intelligence // USITO. 2018. URL: http://www.usito.org/news/three-year-action-plan-focuses-next-generation-artificial-intelligence

[66] Levesque G., Stokes M. Blurred Lines: Military-Civil Fusion and the "Going Out" of China's Defense Industry // Pointe Bello. 2016. URL: https://static1.squarespace.com/static/569925bfe0327c837e2e9a94/t/593dad0320099e64e1ca92a5/1497214574912/062017_Pointe+Bello_Military+Civil+Fusion+Report.pdf

[67] National Security Strategy of the United States of America // The White House. 2017. URL: https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf

[68] A Summary of the 2018 National Defense Strategy of The United States of America. US Department of Defense. 2018. URL: https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

[69] Memorandum: Establishment of the Joint Artificial Intelligence Center (JAIC). Deputy Secretary of Defense. 2018. URL: https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence_center_osd008412-18_r....pdf

- India: within the framework of the National Strategy for Artificial Intelligence drafted for the Government by the National Institution for Transforming India (NITI Aayog)[70] and the "AI Development Roadmap for Ensuring India's Defense and Security"[71] drafted by Tata Sons at the instruction of the Ministry of Defense.
- France: as part of (1) the sharp increase in the country's spending on AI to develop future weapons systems as announced[72] by France's Minister of Armed Forces Florence Parly (USD 1.83 billion); (2) the Parliamentary "Villani Report"[73] stating that AI is now the central political principle of ensuring security, maintaining superiority over the country's potential enemies, and supporting its stance towards its allies; (3) a roadmap of AI capabilities for weapons and its first stage, the Man-Machine Teaming (MMT)[74] project.
- When to comes to AI, the United Kingdom hopes to become part of U.S. programs and, within the framework of this cooperation, opened an AI Lab[75] in May 2018, the centre for AI applied research at the Defense Science and Technology Laboratory (Dstl) following the relevant decision of the Secretary of State for Defense of the United Kingdom.

All the countries mentioned declared the following **three policies** as their principal vectors in steering the course of AI nationalism.

1. A policy guaranteeing the preservation of economic and military AI first mover advantages, as these countries consider themselves, and not without reason, to be trailblazers in AI.
2. A policy of preventing others from copying AI technologies, primarily those that are easily reproduced by any country with a similar technological level.
3. A policy of weakening the stimuli for international trade that automatically result in the global spread of AI technologies.

What is particularly noted is the "readiness to act" using the entire arsenal of state regulation by rapidly and decisively developing new rules and preventing any attempts to undermine national technological AI sovereignty.

AI Nationalization

The fear of falling behind in the global race for AI superiority has spawned the trend of AI nationalism. However, things go beyond AI nationalism.

When it turned out that China's Military-Civil Fusion policy allowed the country, in the course of just a few years, to effectively catch up with the United States,

[70] Strategy for Artificial Intelligence // The National Institution for Transforming India (NITI Aayog). 2018. URL: http://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf
[71] AI Task Force Hands Over Final Report to RM // PIB. 2018. URL: http://pib.nic.in/newsite/PrintRelease.aspx?relid=180322
[72] Turner J. Intelligent design: inside France's €1.5bn AI strategy // Army Technology. 2018. URL: https://www.army-technology.com/features/intelligent-design-inside-frances-e1-5bn-ai-strategy/
[73] Villani C. For a Meaningful Artificial Intelligence // AI For Humanity. 2018. URL: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf
[74] Man-Machine Teaming (MMT). 2018. URL: https://man-machine-teaming.com/
[75] UK launches new artificial intelligence lab at Dstl // Army Technology. 2018. URL: https://www.army-technology.com/news/uk-launches-new-artificial-intelligence-lab-dstl/

the previously undisputed global AI leader, it became clear that other countries have no other option but to go down the same path.

Instead of the separation of state and business that is customary for developed countries, now everyone talks of the advantages of **AI nationalization**, i.e. integrating the resources of public and private companies, aligning the pace of introducing AI innovations, and refocusing strategic objectives on the state gaining economic, geopolitical and military advantages in the international arena.

The AI nationalization agenda is divided into:

- the agenda of the leaders of the AI race,
- the agenda of the outsiders of the AI race.

The Leaders' Agenda: The United States and China

Some experts argue that China has barely covered half the ground of the United States in terms of AI.[76]

Main Driver in AI	Proxy Measure[s]	China	USA
Hardware	Int'l market share of semiconductor prod. (2015)	4% of world	50% of world
	Financing for FPGA chip-makers (2017)	USD 34.4 million (7.6% of world)	USD 192.5 million (42.4% of world)
Data[a]	Mobile users (2016)[b]	1.4 billion (20.0% of world)	416.7 million (5.5% of world)
Research and Algorithms	Number of AI experts	39, 200 (13.1% of world)	78,700 (26.2% of world)
	Percentage of AAAI Conference Presentations (2015)[c]	20.5% of world	48.4% of world
Commercial AI Sector	Proportion of world's AI companies (2017)	23%	42%
	Total investments in AI companies (2012-2016)	USD 2.6 billion (6.6% of world)	USD 17.2 billion (43.4%)
	Total global equity funding to AI startups (2017)	48% of world	38% of world
AI Potential Index[d]	Avg. of the four avg. proxy measures[e]	(5.8 + 20 + 16.8 + 25.9)/4 = 17	(46.2 + 5.5 + 37.3 +41.1)/4 = 33

Calculating the AI-potential index from the methodology of "Deciphering China's AI Dream" research[77] demonstrates that China has covered just over half the ground of the United States in terms of AI.

[76] Deciphering China's AI Dream. The context, components, capabilities, and consequences of China's strategy to lead the world in AI // Future of Humanity Institute, University of Oxford. 2018. URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

[77] Deciphering China's AI Dream. The context, components, capabilities, and consequences of China's strategy to lead the world in AI // Future of Humanity Institute, University of Oxford. 2018. URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

However, a comprehensive analysis of China's civil and military breakthrough technologies,[78] the results of hearings in the United States–China Economic and Security Review Commission[79] and its final report to the U.S. Congress,[80] as well as the latest report on the in-depth analysis of China's weapons[81] recorded an approximate parity in the development of AI technologies in the United States and China.

TABLE 5: CURRENT STATE OF U.S. TECHNOLOGICAL COMPETITION WITH CHINA IN NINE SECTORS

U.S. Leads	Close Competition	China Leads
• Biotechnology • Nanotechnology • Cloud computing • Collaborative robots	• Artificial intelligence • Quantum information science • High performance computing	• Exascale computing • Commercial drones

Note: Factors that determine the state of technological leadership include: the number of firms, global market share, amount of R&D funding provided, the number of patent applications, the number of articles published in high-ranking journals, and the number of citations per publication. The status of technological leadership may shift due to changes in government policies or breakthroughs in R&D.

Source: This assessment is based on testimony received at the Commission's March 2017 hearing on China's pursuit of next-generation, dual-use technologies; contracted research; consultations with government officials, academics, and industry experts; and open source research and analysis.

When examining the nine sectors of dual-use technologies that can be considered the most important in terms of breakthroughs, experts from the U.S. Congress classified AI as a group of three classes of technologies where the United States and China have approximate parity.[82]

[78] Planning for Innovation: Understanding China's Plans for Technological, Energy, Industrial, and Defense Development // US-China Economic and Security Review Commission. 2016.
URL: https://www.uscc.gov/sites/default/files/Research/Planning for Innovation-Understanding China%27s Plans for Tech Energy Industrial and Defense Development072816.pdf

[79] China's Advanced Weapons. Hearing Before the U.S.-China Economic and Security Review Comission // US-China Economic and Security Review Comission. 2016.
URL: https://www.uscc.gov/sites/default/files/transcripts/China%27s Advanced Weapons.pdf

[80] 2017 Report to Congress of the U.S.-China Economic and Security Review Commission» Executive Summary and Recommendations // US-China Economic and Security Review Commission. 2017.
URL: https://www.uscc.gov/sites/default/files/annual_reports/2017 Executive Summary and Recommendations_1.pdf

[81] China's Advanced Weapons Systems. Jane's by IHS Markit for U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.
URL: https://www.uscc.gov/sites/default/files/Research/Jane%27s%20by%20IHS%20Markit_China%27s%20Advanced%20Weapons%20Systems.pdf

[82] China's Advanced Weapons. Hearing Before the U.S.-China Economic and Security Review Comission // US-China Economic and Security Review Comission. 2016.
URL: https://www.uscc.gov/sites/default/files/transcripts/China%27sAdvancedWeapons.pdf

Unlike the high-quality, but civil report,[83] these experts take three special factors into account in their conclusions:[84]

1. In China's long-term development strategy, autonomous unmanned systems and AI will have the main development priority.[85]

2. The specifics of China's AI development roadmap, which prioritizes developing "intellectual weapons,"[86] including it on the list of four critical technological "strategic frontier" areas that set the objective of surpassing the U.S. military by using the "leapfrog development" strategy: a strategy of jumping over several development stages.

3. China's intention to use AI to "accelerate the development of *shashoujian* armaments," which ironically, is translated into English as "Trump Card".[87]

This is not a reference to President Trump, however, but rather to the so-called "shashoujian armaments." The term "shashoujian" can be translated differently: in English, it is a "trump card" or "assassin's mace"; in Russian, it translates as "hitman." The word refers to the Chinese legend where *shashoujian* was used to unexpectedly disable a stronger enemy by using a clever trick (something like the sling David used to defeat Goliath). In his time, Jiang Zemin, former General Secretary of the Central Committee of the Communist Party of China, used the term "*shashoujian*" in the following manner: "Whatever the enemy fears most, that is what we should develop." Since then, China has established the priority development of its "Trump Card," the "*shashoujian* armament" that strikes at the enemy's most vulnerable places. AI has proved very useful for such an approach.

The United States Armed Forces is perfectly aware that China is banking on its "Trump Card." Robert Work,[88] U.S. Deputy Secretary of Defense in 2014–2017, said that, "The whole 'Chinese theory of victory' is known (in translation) as 'systems destruction warfare' because it focuses on electronically paralyzing command-and-control rather than physically destroying tanks, ships, and planes."

The simplest and most obvious example is using the swarm intelligence of micro drones to disable aircraft carriers.[89]

---

[83] Deciphering China's AI Dream. The context, components, capabilities, and consequences of China's strategy to lead the world in AI // Future of Humanity Institute, University of Oxford. 2018.
URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

[84] China's Advanced Weapons. Hearing Before the U.S.-China Economic and Security Review Comission // US-China Economic and Security Review Comission. 2016.
URL: https://www.uscc.gov/sites/default/files/transcripts/China%27sAdvancedWeapons.pdf

[85] Press Release. U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.
URL: https://www.uscc.gov/sites/default/files/China%27s%20Advanced%20Weapons_PR.pdf

[86] Kania E.B. Quest for an IA Revolution in Warfare // The Strategy Bridge. 2017.
URL: https://thestrategybridge.org/the-bridge/2017/6/8/-chinas-quest-for-an-ai-revolution-in-warfare

[87] Kania E. China's trump Card // Prospect Magazine. 2017.
URL: https://prospectmagazine.co.uk/other/chinas-trump-card

[88] Freedberg Jr.S.J. US Must Hustle On Hypersonics, EW, AI: VCJCS Selva & Work // Breaking Defense. 2018.
URL: https://breakingdefense.com/2018/06/us-must-hustle-on-hypersonics-ew-ai-vcjcs-selva-work/

[89] Kania E.B. 杀手锏 and 跨越发展 : Trump Cards and Leapfrogging. // The Strategy Bridge. 2017.
URL: https://thestrategybridge.org/the-bridge/2017/9/5/-and-trump-cards-and-leapfrogging

Another example is the intellectualization of missiles. Here AI is set the objective of enabling "missiles to have advanced capabilities in sensing, decision-making, and execution of missions, including through gaining a degree of 'cognition' and the ability to learn."[90]

A high-ranking Pentagon official recently said[91] that "the first nation to deploy an electromagnetic pulse weapon on the battlefield to disable enemy systems would reshape the face of warfare. Once again, it is far from obvious that is a race the United States will win." He refers precisely to integrating the capabilities of electromagnetic pulse weapons and AI.

A detailed description of the entire landscape of potential threats stemming from China's priority introduction of AI technologies into cutting edge weapons systems can be found at the following link.[92]

Thus far, the United States and China are head to head in AI. Many experts, however, believe that, in no more than ten years,[93] China's authoritarianism will summarily defeat[94] U.S. democracy due to its absolute dominance in terms of the volume of data collected. China already has four times the volume of data that the United States posesses. And the gap is growing.

As Kai-Fu Lee, one of the most knowledgeable professionals in the U.S.–China AI confrontation and a former colleague of mine at Silicon Graphics, once said, "In fact, there is only one fundamental innovation in AI, and that is deep learning. Everything that is done now in AI is just fine-tuning deep learning to meet the needs of specific applied areas."

Deep learning requires as much data as possible. And the one in possession of greater volume of big data has already most likely won the competition. Although, as Jack London's Smoke Bellew said, "The race was not lost until one or the other won," and the United States will not just give up in the **AI arms race**.

A duopoly looms in the near future. Thus far, the two global AI leaders, the United States and China, adhere to radically opposed strategies:

- China: continuing to do THAT regardless
- The United States: doing everything to hamper China in doing THAT

[90] Chinese Advances in Unmanned Systems and the Military Applications of Artificial Intelligence – the PLA's Trajectory towards Unmanned, "Intelligentized" Warfare. Testimony before the U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2017.
URL: https://www.uscc.gov/sites/default/files/Kania_Testimony.pdf

[91] Apps P. Commentary: Western armies are losing their high-tech edge // Reuters. 2018.
URL: https://www.reuters.com/article/us-apps-military-commentary/commentary-western-armies-are-losingtheir-high-tech-edge-idUSKBN1JV2LN

[92] Implications of China's Military Modernization. Testimony before the U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.
URL: https://www.uscc.gov/sites/default/files/Nurkin_Written%20Testimony.pdf

[93] Press Release. U.S.-China Economic and Security Review Commission // U.S.-China Economic and Security Review Commission. 2018.
URL: https://www.uscc.gov/sites/default/files/China%27s%20Advanced%20Weapons_PR.pdf

[94] Linn T.C. Race to develop artificial intelligence is one between Chinese authoritarianism and U.S. democracy // San-Francisco Chronicle. 2018.
URL: https://www.sfchronicle.com/opinion/openforum/article/Race-to-develop-artificial-intelligence-is-one-13189380.php

For a detailed description of THAT, see the following link[95]. In short, it boils down to the following:

A. Using protectionist measures to protect its own market from import and competition in AI.

B. State-sponsored IP theft through physical theft, cyber-enabled espionage and theft, evasion of U.S. export control laws, and counterfeiting and piracy.

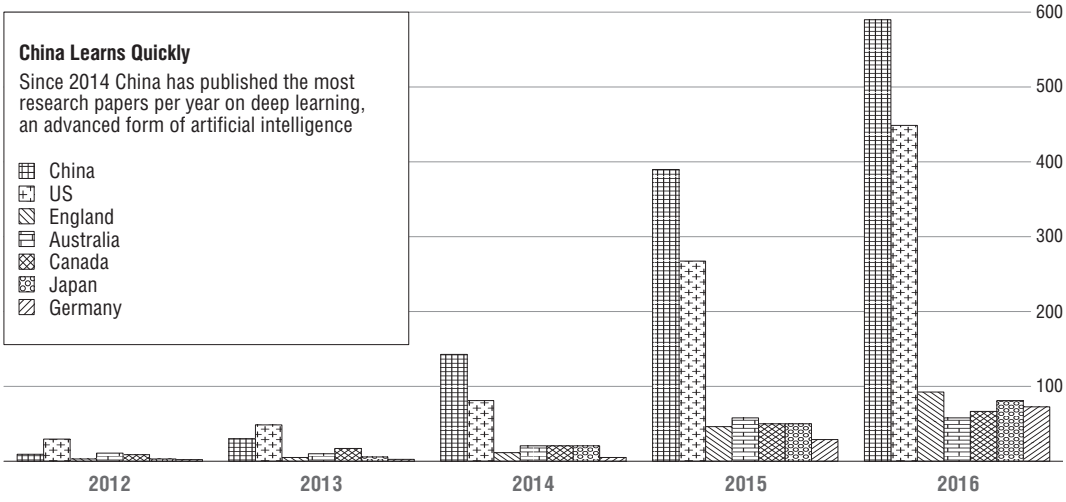
C. Coercive and intrusive regulatory gambits to force technology transfer from foreign companies, typically in exchange for limited access to the Chinese market.

D. Methods of information harvesting that include open source collection; placement of non-traditional information collectors at U.S. universities, national laboratories, and other centres of innovation; and talent recruitment of business, finance, science, and technology experts.

E. State-backed, technology-seeking Chinese investment.

The Agenda of the Outsiders in the AI Race

The real situation is such that everyone is the outsider in the AI race except for its leaders, the United States and China. Although technologically developed countries, such as France, Germany, India and South Korea appear to be incomparable to third-world countries in the level of AI technologies, the fate of becoming **AI colonies** of the leading countries awaits them all.



The absolute leadership of the United States and China in academic publications on deep learning. Source: Digital Transformation Monitor

[95] How China's Economic Aggression Threatens the Technologies and Intellectual Property of the United States and the World // White House Office of Trade and Manufacturing Policy. 2018. URL: https://www.whitehouse.gov/wp-content/uploads/2018/06/FINAL-China-Technology-Report-6.18.18-PDF.pdf

In **AI neo-colonialism,** 21st century colonizers receive Big Data for their AI technologies, the new equivalent of gold and silver, while the only thing colonial countries can do is hope for the civilizational spirit of AI neo-colonialists and for their financial assistance.

However, in this situation, as in the race between the two leaders, "the race was not lost until one or the other won."

For instance, European countries are trying to do at least something to avoid the fate of becoming colonies that supply the United States with big data.

Recognizing their loss to the United States and China in infrastructure and scale of their AI programs, European specialists describe their AI capabilities as "blobs of brilliance" and dream of[96] ways to "take those blobs of brilliance and bring them all together."

"A Special Outsider": Russia's Stance in the AI Race

The question of Russia's potential and prospects is somewhat more complicated than simply filing it under the outsiders of the race.

First, there is a theoretical scenario where outsiders could gain an advantage over the leaders in the field of AI for military purposes. This scenario has remained outside the scope of our article, since it does not aim to consider all possible scenarios, but only the most probable ones. However, it is necessary to mention a scenario that favours outsiders, since it is being considered and analysed by very influential experts. This scenario was discussed in the Future of War issue of *Foreign Policy's Fall 2018*.[97] Michael C. Horowitz's article "The Algorithms of August"[98] analyses the prospects of this scenario and sets Russia apart as a "special outsider" that is, like China, potentially capable of competing for leadership.

Second, the combination of Russia's traditionally asymmetric responses to geopolitical challenges, with its still considerable scientific and technological Soviet legacy, can have major consequences at the junctions of[99] AI technologies and new classes of weapons (from hypersonic to electronic warfare). And there is also the junction of AI with quantum computing[100] and some other things that might radically change the balance of military power in the near future.

So nothing is clear-cut with Russia, and this fact should be studied separately, since the best works on the subject (for instance, "The Impact of Technological Factors on the Parameters of Threats to National and International Security, Military Conflicts, and Strategic Stability"[101]) barely touch on these matters.

[96] Helmer J. The geopolitics of AI raises trust issues for Learning and HR // Lumesse Learning. 2018. URL: https://www.lumesselearning.com/ai-trust-issues/
[97] Tepperman J. The Future Of War // Foreign Policy. Fall 2018. URL: https://foreignpolicy.com/2018/09/12/the-future-of-war-editors-note-fall-2018/
[98] Horowitz M.C. The Algorithms of August // Foreign Policy. 2018. URL: https://foreignpolicy.com/2018/09/12/will-the-united-states-lose-the-artificial-intelligence-arms-race/
[99] Freedberg Jr.S.J. US Must Hustle On Hypersonics, EW, AI: VCJCS Selva & Work // Breaking Defense. 2018. URL: https://breakingdefense.com/2018/06/us-must-hustle-on-hypersonics-ew-ai-vcjcs-selva-work/
[100] Thomas-Noone B. US playing catch-up as technology advantage erodes // ASPI. The Strategist. 2018. URL: https://www.aspistrategist.org.au/us-playing-catch-up-as-technology-advantage-erodes/

Where is AI Nationalization Taking Us?

However different the agendas of the leaders and outsiders of the AI race are, they presuppose the same steps essentially leading to Chinese-style AI nationalization.

First, it means establishing combined national civil military complexes for developing AI technologies.

Of course, the giants of AI business will resist this for all they are worth. For instance, on June 1, 2018, Google announced that it would not renew its contract to support a U.S. military initiative called Project Maven. This project is the military's first operationally deployed "deep-learning" AI system used to classify images collected by military drones. The company's decision to withdraw came after roughly 4000 of Google's 85,000 employees signed a petition to ban Google from building "warfare technology."

The response followed immediately.

As early as June 6, 2018, Gregory Allen, an expert of the Center for New American Security, published an article[102] entitled "AI Researchers Should Help with Some Military Work." The article formulates the new "ethical imperative" for commercial AI companies:

"The ethical choice for Google's artificial-intelligence engineers is to engage in select national-security projects, not to shun them all."

Allen writes about Google's proposals to recuse themselves from participating in Project Maven:

"Such recusals create a great moral hazard. Incorporating advanced AI technology into the military is as inevitable as incorporating electricity once was, and this transition is fraught with ethical and technological risks. It will take input from talented AI researchers, including those at companies such as Google, to help the military to stay on the right side of ethical lines."

The AI industry community rushed to the aid of Google, which had essentially been accused of its employees placing their ethical principles above national security. In July, **Elon Musk**, the founders of DeepMind, a founder of Skype and several renowned IT professionals called upon their colleagues[103] to sign the pledge not to develop AI-based "lethal autonomous weapons," aka "killer robots" (for more on this, see the following link[104]).

[101] The Impact of Technological Factors on the Parameters of Threats to National and International Security, Military Conflicts, and Strategic Stability. Moscow: MSU Press, 2017.

[102] Allen G.C. AI researchers should help with some military work // Nature. 2018.
URL: https://www.nature.com/articles/d41586-018-05364-x

[103] Vincent J. Elon Musk, DeepMind founders, and others sign pledge to not develop lethal AI weapon systems // The Verge. 2018.
URL: https://www.theverge.com/2018/7/18/17582570/ai-weapons-pledge-elon-musk-deepmind-foundersfuture-of-life-institute

[104] Initial Reference Architecture of an Intelligent Autonomous Agent for Cyber Defense. ARL US Army Research Laboratory // Cornell University Library. 2018.
URL: https://arxiv.org/ftp/arxiv/papers/1803/1803.10664.pdf

In August, 116 renowned AI professionals and experts signed a petition[105] calling upon the United Nations[106] to prohibit lethal autonomous weapons. In their statement, the group states that developing such technologies will result in a "third revolution in warfare" equal to the invention of gunpowder and nuclear weapons.

While the United Nations remains silent, **Sandro Gaycken**, Senior Advisor to NATO, offered his response,[107] noting that such initiatives are supremely complacent and risk granting authoritarian states an asymmetric advantage.

"These naive hippy developers from Silicon Valley don't understand – the CIA should force them," Gaycken said.[108]

"Forcing to understand" applies not only to the giants of the AI industry, such as Google, but to a host of promising startups working on the cutting edge of AI development.

"Startups have to be embedded into large corporate structures, to have access to the kind of data they require, to build high-quality AI," Gaycken believes.[109]

The same logic applies to individual AI professionals: they should also work on national security tasks.

"There are also clear differences in how talent can be utilised in more authoritarian systems. The command and control economies of authoritarian countries can compel citizens, experts and scientists to work for the military. 'Where you require very good brains to understand what is going on and to find your niche, to find specific weaknesses and build specific strengths – in those countries they simply force the good guys to work for them," Gaycken explains.[110]

Such a picture is hard to imagine in the United States:

- AI startups working for giant AI corporations;
- Giant AI corporations working for the military;
- AI professionals working where the military tell them to (like "*sharashkas*" or R&D labs in the Soviet labour camp system under Stalin).

But that was the case of the USSR. And this is now the case of China…

Whether the United States will succeed in introducing such practices ultimately depends only on the level of threat. Any democracy ends where a high level of danger for national security begins.

Many influential people in the United States are convinced even now that the AI war with China is already on. The only thing remaining is to convince the majority of Americans of the same. After the election of Trump, it does not seem all that impossible.

---

[105] Killer robots: World's top AI and robotics companies urge United Nations to ban lethal autonomous weapons // Future Of Life. 2017.
URL: https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nationsban-lethal-autonomous-weapons/

[106] Vincent J. Elon Musk and AI leaders call for a ban on killer robots // The Verge. 2017.
URL: https://www.theverge.com/2017/8/21/16177828/killer-robots-ban-elon-musk-un-petition

[107] Upchurch T. How China could beat the West in the deadly race for AI weapons // Wired. 2018.
URL: https://www.wired.co.uk/article/artificial-intelligence-weapons-warfare-project-maven-google-china

[108] Ibid.

[109] Ibid.

[110] Ibid.

A Brave New World Without Work

N. Markotkin

As Strong Artificial Intelligence continues to develop, so does the range of jobs it can do

It is quite conceivable that the majority of the working people will find themselves back in school in the foreseeable future. The problem, however, is that no one really knows what to study. It has been estimated that as much as 85 per cent of the professions that will be in demand in 2030 do not yet exist.

The technological changes to come promise in the very least to transform the labour market in developed countries. Jobs associated with unskilled labour will be first in the line of fire. Following that, the development of artificial intelligence will also bring down hard times on lawyers, teachers, miners, middle management, and journalists among others. Work that requires a creative approach and is not limited to the performance of predictable combinations will be best prepared to deal with the new reality.

The Russian economy could hardly be called high-tech. A significant part of the working population is engaged in routine low-skilled labour, and productivity remains low as well. Another disaster waiting to happen to the Russian economy is related to outdated industry and the decline of domestic engineering. Furthermore, one cannot rule out that in the near future Russia will launch a massive programme to introduce robotic automation and artificial intelligence. Because of the lack of a strong trade union movement and the prevalence of hybrid and grey forms of employment, labour automation could lead to much more severe social consequences in Russia than in Western countries. Finally, it is entirely possible that the catch-me-if-you-can nature of such modernization will mean Russia introducing more primitive technologies than in more developed countries.

What's the first thing that comes to mind when you think about the soon-to-come widespread introduction of robots and artificial intelligence (AI)? Endless queues of people waiting to get unemployment benefits? Skynet drones ploughing the sky over burnt-out slums? Or the opposite: idleness and equality provided by the labour of mechanical slaves? In all likelihood the reality will be less flashy, though that doesn't mean we should ignore the social consequences of the technological changes taking place before our very eyes.

Revolution on the March

The Fourth Industrial Revolution with its robotics, bio and nanotechnologies, 3D printing, Internet of things, genetics, and artificial intelligence is rapidly spreading across the world[111]. The coming technological changes will have direct con-

[111] Marsh P. The New Industrial Revolution. Consumers, Globalization, and the End of Mass Production. Moscow: Gaidar Institute Press, 2015.

sequences for a number of existing professions and promise in the very least to transform the labour market in developed countries.

The high speed of change (suffice it to say that 10 of the most popular professions of 2010 did not exist[112] in 2004) makes it difficult to predict the impact on society. In this regard, the assessments of experts and international organizations range from optimistic[113] to alarmist.[114] However, even if we were to eliminate the most extreme case scenarios, we could still say with certainty that a fundamental restructuring of the global economy, comparable to the one that took place in the 18th–19th centuries during the First Industrial Revolution, awaits us in the foreseeable future.

According to the World Economic Forum (WEF) Future of Jobs[115] report, 65 per cent of today's primary school students will have hitherto unheard-of professions. *McKinsey* came to the same conclusion, highlighting in their report[116] that at the current level of technological development, 30 per cent of the functions of 60 per cent of professions can be automated. M. Osborne and C. Frey of Oxford University give an even more pessimistic forecast. According to their research,[117] 47 per cent of jobs in the US risk being automated within 20 years.

Who will robots replace?

What professions are at risk? First at risk is, of course, unskilled labour. The Osborne and Frey study found clerks, data entry workers, librarians, machine operators, plumbers, sales specialists, and equipment adjusters among others to be those most vulnerable.

According to WEF,[118] from 2015 to 2020, job reductions will have the greatest effect on office professions (4.91 per cent) and the manufacturing sector (1.63 per cent). Employment in areas such as design, entertainment, construction, and sales should also decline by 1 per cent. In turn, the most significant growth in jobs is predictably expected in the field of computer technology (3.21 per cent), architectural and engineering specialties (2.71 per cent), and management (just under 1 per cent).

Predictably, professions related to transport risk automation in the medium term. The development of self-driving vehicles could radically change both the passenger and freight traffic markets. In the US alone, 8.7 million people are employed[119] in the long-distance freight traffic industry. If you take into account

[112] Shift Happens // The Guardian. 2010.
URL: https://www.theguardian.com/science/punctuated-equilibrium/2010/oct/20/2
[113] Annunziata M. and Biller S. The Future of Work, GE Discussion Paper, General Electric. 2014.
[114] When machines can do any job, what will humans do? // EurekAlert. 2016.
URL: https://www.eurekalert.org/pub_releases/2016-02/ru-wmc021016.php
[115] The Future of Jobs // World Economic Forum. 2016.
URL: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf
[116] A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017.
URL: https://www.mckinsey.com/~/media/McKinsey/Global Themes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx
[117] New study shows nearly half of US jobs at risk of computerisation // University of Oxford. 2013.
URL: http://www.eng.ox.ac.uk/about/news/new-study-shows-nearly-half-of-us-jobs-at-risk-of-computerisation
[118] The Future of Jobs // World Economic Forum. 2016.
URL: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf
[119] Gerstein D.M. Will a Robot Take My Job? // The National Interest. 2018.
URL: http://nationalinterest.org/feature/will-robot-take-my-job-25444

all of the business operations connected to trucking (motels, roadside cafes, etc.), the number increases to 15 million or about 10 per cent of the country's labour force. Reductions in passenger transport and the public transport sector are likely to be even more significant. It is also probable that self-guiding technologies will be introduced into sea freight traffic[120] in the near future. The development of artificial intelligence should also bring down hard times[121] on lawyers, teachers, miners, middle management, and journalists among others.

It can be said that on the whole, employment will gradually move from services to other sectors of the economy, many of which have yet to be created. The possibility is a confirmation of the revolutionary nature of the changes that are taking place rather than something unique. Before the First Industrial Revolution, over 70 per cent of the population[122] was occupied with agriculture, whereas nowadays the number hovers around a few percent[123] in developed countries. The percentage of those employed in manufacturing continued to grow until the mid-twentieth century, though it has now fallen to 24 per cent in the EU and 19 per cent in the US (27 per cent in Russia)[124] as a result of the Digital Revolution. Meanwhile, although there are fewer workers, production volume continues to rise steadily.[125] It would now appear to be time to automate services.

The Golden Age of Engineers and Psychiatrists?

Professions associated with intellectual work or direct personal contact with clients are least likely to suffer in the short term. According to the study from Oxford University,[126] professions least susceptible to automation include various jobs in medicine and psychology, as well as coaches, social workers, programmers, engineers, representatives of higher management and creative professionals.

In other words, those whose work requires a creative approach and is not limited to the performance of predictable combinations will be best prepared to deal with the new reality. If we were to speak of engineers in this regard, it would have to be clarified that design engineers are generally safe, while operating engineers, on the contrary, are at risk.

Three key factors[127] are keeping automation away from the creative professions. To successfully perform their tasks, artificial intelligence must possess intuition

[120] Rolls-Royce: Autonomous Vessels Will Precede Autonomous Cars and Planes // Pro-Arctic.ru. 2018. URL: http://pro-arctic.ru/08/05/2018/technology/31900
[121] A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017. URL: https://www.mckinsey.com/~/media/McKinsey/Global Themes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx
[122] Clark G. The Agricultural Revolution and the Industrial Revolution: England, 1500–1912 // University of California. 2002. URL: http://faculty.econ.ucdavis.edu/faculty/gclark/papers/prod2002.pdf
[123] Gerstein D.M. Will a Robot Take My Job? // The National Interest. 2018. URL: http://nationalinterest.org/feature/will-robot-take-my-job-25444
[124] Employment in industry (% of total employment) (modeled ILO estimate). International Labour Organization, ILOSTAT database // The World Bank. 2018. URL: https://data.worldbank.org/indicator/SL.IND.EMPL.ZS
[125] Gerstein D.M. Will a Robot Take My Job? // The National Interest. 2018. URL: http://nationalinterest.org/feature/will-robot-take-my-job-2544
[126] New study shows nearly half of US jobs at risk of computerisation // University of Oxford, 2013. URL: http://www.eng.ox.ac.uk/about/news/new-study-shows-nearly-half-of-us-jobs-at-risk-of-computerisation
[127] Lucky R.W. Are Engineers Designing Their Robotic Replacements? // IEEE Spectrum. 2016. URL: https://spectrum.ieee.org/at-work/tech-careers/are-engineers-designing-their-robotic-replacements

and an ability to manipulate material objects (touch) and make use of creative and social intelligence. Technology at its current level of development does not actually allow for the resolution of these problems. However, as strong AI[128] continues to develop, the range of jobs available to it will invariably increase as well. It will expand the limits of automation[129] that have already been achieved with existing technologies and will make it possible for computers to make managerial decisions and even, perhaps, engage in creative activity. Therefore, it cannot be ruled out that in the medium or long term, machines might successfully replace writers and artists along with engineers and managers. Furthermore, precedents do exist for AI's successfully composing literary texts.[130]

Thus, it is quite conceivable that the majority of the labour force will find itself back in school in the foreseeable future. The problem, however, is that no one really knows what to study. It has been estimated,[131] that as many as 85 per cent of the professions that will be in demand in 2030 do not yet exist. Even in developed countries, the education systems have yet to adapt[132] to the new reality.

What will become of our country and of us?

Today, most researchers have little doubt that developed countries will successfully adapt to the changes coming one way or another (which does not rule out the possibility of social tension and growth in income inequality[133]). New technologies could help create additional jobs[134] to replace those that have been lost, as it was not long ago[135] following the rapid development of the Internet. It is assumed that the new professions will be more creative and better paid.

A new balance will gradually be established in the labour market. The nature of manufacturing will also change. The development of automation and 3D printing will make it possible to create efficient local production facilities[136] focused on the specific needs of consumers. This will facilitate the return of a part of production from developing countries to developed (so-called reshoring).

In turn, the consequences of automation could be much more negative for countries of the third world. The percentage of non-skilled jobs in developing coun-

[128] Scheftelowitsch D. Artificial Intelligence: Time of the Weak // Working Paper No. 44 "International and Social Impacts of Artificial Intelligence Technologies". Moscow: RIAC. 2018.
[129] The Future of Jobs // World Economic Forum. 2016.
URL: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf
[130] Shoemaker N. Japanese AI Writes a Novel, Nearly Wins Literary Award // BigThink.com. 2016.
URL: http://bigthink.com/natalie-shoemaker/a-japanese-ai-wrote-a-novel-almost-wins-literary-award
[131] Emerging Technologies' Impact on Society & Work in 2030 // Dell Technologies. 2017.
URL: https://www.delltechnologies.com/content/dam/delltechnologies/assets/perspectives/2030/pdf/SR1940_IFTFforDellTechnologies_Human-Machine_070517_readerhigh-res.pdf
[132] Shewan D. Robots will destroy our jobs – and we're not ready for it // The Guardian. 2017.
URL: https://www.theguardian.com/technology/2017/jan/11/robots-jobs-employees-artificial-intelligence
[133] A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017.
URL: https://www.mckinsey.com/~/media/McKinsey/GlobalThemes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx
[134] Annunziata M., Biller S. The Future Of Work // General Electric. 2014.
URL: http://files.publicaffairs.geblogs.com/files/2014/04/AM_II_FOW_WhitePaper_FINAL-1.pdf
[135] A Future that Works: Automation, Employment, and Productivity // McKinsey&Company. 2017.
URL: https://www.mckinsey.com/~/media/McKinsey/Global Themes/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works_Full-report.ashx
[136] Fratocchi L. Is 3D Printing an Enabling Technology? // Springer. 2017.
URL: https://link.springer.com/chapter/10.1007%2F978-3-319-58883-4_5

tries decreased[137] by 8 per cent between 1995 and 2012. Reshoring could significantly accelerate this process in the short term. Since the proportion of people engaged in low-skilled work in low and middle-income countries is much higher, the growth of unemployment would threaten to become a major global problem. The situation would be further aggravated by the underdevelopment of labour protection institutions in these countries.

It must be noted that risks of this sort are endemic to Russia as well. Despite the significantly higher level of education of its citizens in comparison to that in developing countries, the Russian economy could hardly be called high-tech. A significant part of the working population is engaged in routine low-skilled labour, and productivity remains low as well. At the present time, Russia lags significantly behind[138] other developed countries in regards to this indicator (and behind the US by more than 100 per cent), and according to some estimates falls below the world average. What's more, factory jobs are not the only ones at stake – an army of many millions of bureaucrats and clerks is also under threat of redundancy as a result of digitalization.

Another disaster waiting to happen to the Russian economy is related to outdated industry and the decline of domestic engineering. At present, institutions of higher education produce mainly operational engineers trained to maintain tools and machines. What's more, even the limited innovative potential of Russian engineers is not needed[139] by Russian industry.

Furthermore, it cannot be ruled out that in the near future Russia will launch a massive programme to introduce robotic automation and artificial intelligence. All the more since it fits in perfectly with the desire to modernize and digitalize[140] the national economy repeatedly spoken of by the Russian leadership. Because of the lack of a strong trade union movement and the prevalence of hybrid and grey forms of employment, labour automation could lead to much more severe social consequences in Russia than in Western countries. Finally, it is entirely possible that the catch-me-if-you-can nature of such modernization will result in Russia introducing more primitive technologies than in more developed countries. Editor-in-Chief of Russia in Global Affairs magazine and RIAC Member Fyodor Lukyanov cleverly described a similar scenario in his article.[141]

Saving the Rank and File

Ways to reduce the social consequences of labour automation have long been at the heart discussions surrounding the Fourth Industrial Revolution and the development of AI. The Robot Tax is one measure being considered. Microsoft

[137] Frey C.B., Rahbari E. Do labor-saving technologies spell the death of jobs in the developing world? // Brookings.edu. 2016.
URL: https://www.brookings.edu/wp-content/uploads/2016/07/Global_20160720_Blum_FreyRahbari.pdf

[138] Labor Productivity in Russian Federation // Analytical Center for the Government of the Russian Federation. 2017.
URL: http://ac.gov.ru/files/publication/a/13612.pdf

[139] Training of Engineers: Challenges of the Modern Era // UrFU. 2015.
URL: http://elar.urfu.ru/bitstream/10995/32709/1/978-5-7996-0000_2015.pdf

[140] President Signed the Decree "On national objectives and strategic tasks of Russian Federation's development in the period up to 2024" // President of Russian Federation. 2018.
URL: http://kremlin.ru/events/president/news/57425

[141] Lukyanov F. Russia 2035: Boundaries of the Possible // Kommersant. 2017.
URL: https://www.kommersant.ru/doc/3384073

Founder Bill Gates supports the idea and has proposed[142] collecting income tax and social payments on robot labour to slow down the pace of automation. "Right now, the human worker who does, say, $50,000 worth of work in a factory, that income is taxed and you get income tax, social security tax, all those things. If a robot comes in to do the same thing, you'd think that we'd tax the robot at a similar level," he declared in an interview for the Internet publication Quartz. It is his opinion that the funds received from payments of this sort should be used by governments to create social security systems for those who have lost their jobs as a result of automation.

The first country to resort to this measure is South Korea, which introduced[143] an indirect tax on robots in August 2017. The European Union also discussed the introduction of a similar tax, though the clause proposed by Progressive Alliance of Socialists and Democrats Representative Mady Delvaux was rejected by the European Parliament[144] was rejected by the European Parliament because it could slow the development of innovations. At the same time, the parliament approved the resolution itself, which calls for granting robots the status of legal entities.

A universal basic income could also soften the effect of rising unemployment and inequality. Elon Musk[145] supports the initiative together with numerous other businessmen and experts.[146] At the same time, a lack of work to afford one the opportunity to fulfil one's potential poses a significant social risk. Significant unemployment, even in the absence of poverty, can contribute to the marginalization of the population and the growth of crime – the first jobs to go are those of low-skilled employees, who are unlikely to spend all of their permanent free time engaged in yoga and self-improvement activities.

Possible ways of mitigating the consequences of the upcoming restructuring of the world economy include a change in the nature of employment.[147] Technological changes and expanding access to the Internet allow more and more people to work remotely. Thus, some of those who lose their jobs will be able to find themselves a place in the new economy without having to change their place of residence.

Some believe[148] that automation will increase and not reduce the total number of jobs by accelerating the pace of economic development over the long term. *Amazon* is one example of how automation has not resulted in staff reduction.

---

142 Delaney K.J. The robot that takes your job should pay taxes, says Bill Gates // QUARTZ. 2017.
URL: https://qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/

143 Vigliarolo B. South Korea 'robot tax' is no tax at all; it's a warning of looming automation crisis // TechRepublic. 2017.
URL: https://www.techrepublic.com/article/south-korea-robot-tax-is-no-tax-at-all-its-a-warning-of-loomingautomation-crisis/

144 European parliament calls for robot law, rejects robot tax // Reuters. 2017.
URL: https://www.reuters.com/article/us-europe-robots-lawmaking-idUSKBN15V2KM

145 Weller C. Elon Musk doubles down on universal basic income: 'It's going to be necessary' // Business Insider. 2017. URL: http://www.businessinsider.com/elon-musk-universal-basic-income-2017-2

146 Weller C. Giving people free money could be the only solution when robots finally take our jobs // Business Insider. 2016.
URL: http://www.businessinsider.com/basic-income-could-be-the-only-solution-in-a-robot-economy-2016-4

147 Gerstein D.M. Will a Robot Take My Job? // The National Interest. 2018.
URL: http://nationalinterest.org/feature/will-robot-take-my-job-25444

148 Annunziata M., Biller S. The Future Of Work // General Electric. 2014.
URL: http://files.publicaffairs.geblogs.com/files/2014/04/AM_II_FOW_WhitePaper_FINAL-1.pdf

While increasing the number of robots[149] employed in its warehouses from 1,400 to 45,000, it has managed to retain the same number of jobs. It has also been noted that automation is becoming increasingly necessary due to a decrease in the working-age population[150] (primarily in developed countries).

It should be noted that these measures are all limited in nature and hardly correspond to the scale of changes that stand to be swept in by the Fourth Industrial Revolution. To avoid mass unemployment and social instability, governments must develop comprehensive short-term strategies for adapting the population to the new reality. It is very likely that new programs will be needed to retrain citizens en masse for new professions.

Russia is no exception here; on the contrary, it is of vital importance that our country reform its education system in the near future, especially as regards technical education. It is equally important to develop targeted support programs for those parts of the population that are most vulnerable to automation and digitalization. Moreover, it would seem advisable to make use of existing experience to mitigate the social consequences of factory closures in Russian single-industry towns. If we continue to move as sluggishly as we are moving at present, we risk turning into a kind of reserve for yesterday's technologies with a population becoming ever more rapidly marginalized.

[149] Kessler S. The optimist's guide to the robot apocalypse // QUARTZ. 2017. URL: https://qz.com/904285/the-optimists-guide-to-the-robot-apocalypse/

[150] Global growth: Can productivity save the day in an aging world? // McKinsey&Company. 2015. URL: https://www.mckinsey.com/~/media/McKinsey/Global Themes/Employment and Growth/Can long term global growth be saved/MGI_Global_growth_Full_report_February_2015pdf.ashx

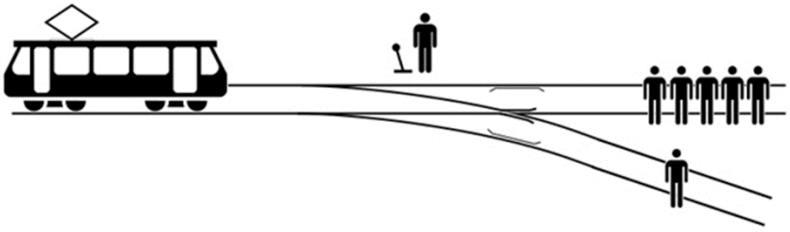
Ethical and Legal Issues in Artificial Intelligence

M. Karliuk

Ethics and law are inextricably linked in modern society, and many legal decisions arise from the interpretation of various ethical issues. Artificial Intelligence (AI) adds a new dimension to these issues. Systems that use artificial intelligence technologies are becoming increasingly autonomous in terms of the complexity of the tasks they can perform, their potential impact on the world and the diminishing ability of humans to understand, predict and control their functioning. Most people underestimate the real level of autonomy of these systems. They can learn from their own experience and perform actions their creators did not intend them to perform. That generates several ethical and legal difficulties that will be addressed in this article.

Ethics and AI

There is a famous thought experiment known as the trolley problem. It raises a series of important ethical issues that have a direct bearing on AI. Imagine a runaway trolley hurtling along its tracks, and there are five people tied to those tracks. You are standing next to a lever that turns a switch, and the trolley will be redirected to another track where there is one person tied to the tracks. Do you pull the lever?



Trolley problem. Source: Wikimedia.org

There is no unequivocal answer to the question. Moreover, there are many variations of the situation when such a decision has to be made.[151] And different social groups tend to give different answers. For instance, Buddhist monks are overwhelmingly prepared[152] to sacrifice the life of one person to save five others, even in a more complicated version of the trolley problem.

[151] See, for instance, Edmonds D. Would You Kill the Fat Man? The Trolley Problem and What your Answer Tells Us about Right and Wrong. Pinceton: Princeton University Press, 2013.

[152] Madrigal A.C. If Buddhist Monks Trained AI // The Atlantic. 2017. URL: https://www.theatlantic.com/technology/archive/2017/06/how-do-buddhist-monks-think-about-the-trolleyproblem/532092/

As for AI, such a situation can arise, for instance, with self-driving vehicles when an accident on the road is unavoidable. The question thus arises as to whose lives should take priority – those of the passengers, the pedestrians or neither. The Massachusetts Institute of Technology (MIT) even launched a special website[153] on the problem where users can try various scenarios and choose what to do in a particular situation.

The question arises: What actions are allowable from the legal point of view? What are the criteria for a particular choice? And who should ultimately be held responsible? The problem has already been addressed by companies and regulators. For instance, Mercedes announced that their cars will prioritize the lives of passengers. The Federal Ministry of Transport and Digital Infrastructure of Germany responded[154] to this immediately, in anticipation of future regulations, stating that making such a choice based on a set of criteria would be illegal, and that the car manufacturer be held responsible for any injury or loss of life.

Other countries could choose a different approach. Take China's Social credit system,[155] for instance, which rates every citizen based on how law-abiding and how useful to society they are, etc. Those with low ratings will face sanctions. What is stopping the Chinese government from introducing a law that forces manufacturers of driverless vehicles to sacrifice the lives of lower-rated citizens in the event of an unavoidable accident? Facial recognition technologies and access to relevant databases will make it possible to determine and compare the ratings of potential victims.

The Main Problems Facing the Law

The legal problems run even deeper, particularly when it comes to robots. A system that learns from the information it receives from the outside world can act in ways that its creators could not have predicted,[156] and predictability is crucial to modern legal approaches. Moreover, such systems can operate independently from their creators or operators, which complicates the task of determining responsibility. These characteristics posit problems related to predictability and the ability to act independent while at the same time not being legally responsible.[157]

The are many options in terms of regulation, including regulation that is based on existing norms. For instance, technologies that use artificial intelligence can be regulated as items subject to copyright or as property. However, complications arise again, especially if we take into account the ability of such technolo-

[153] Moral Machine // URL: http://moralmachine.mit.edu/
[154] Brown M. Mercedes's Self-Driving Cars Will Kill Pedestrians Over Drivers // Inverse. 2016. URL: https://www.inverse.com/article/22204-mercedes-benz-self-driving-cars-ai-ethics
[155] Gordeev A. Digital Dictatorship: How China is Introducing Social Credit System // RBC. 2016. URL: https://www.rbc.ru/business/11/12/2016/584953bb9a79477c8a7c08a7
[156] Asaro P. From Mehcanisms of Adaptation to Intelligence Amplifiers: The Philosophy of W. Ross Ahby // Wheeler M., Husbands P., and Holland O. (eds.) The Mechanical Mind in History, Cambridge, MA: MIT Press. pp. 149–184
[157] Asaro P. The Liability Problem for Autonomous Artificial Agents // AAAI Symposium on Ethical and Moral Considerations in Non-Human. Agents, Stanford University, Stanford, CA. p. 191.

gies to act autonomously, against the will of their creators, owners or proprietors. In this regard, it is possible to apply the rules that regulate a special kind of ownership, namely animals, since the latter are also capable of autonomous actions. In Russian Law, the general rules of ownership are applied to animals (Article 137 of the Civil Code of the Russian Federation), therefore, liability is incurred under Article 1064 of the Civil Code of the Russian Federation: the issue of responsibility, therefore, comes under Article 1064 of the Civil Code of the Russian Federation: injury inflicted on the persona or property of an individual shall be subject to full compensation by the person who inflicted the damage, i.e. the owner of the autonomous agent.

Proposals on the application of the law on animals have been made,[158] although they are somewhat limited. First, the application of legislation on the basis of analogy is unacceptable within the framework of criminal law. Second, these laws have been created primarily for domestic animals, which we can reasonably expect will not cause harm under normal circumstances. There have been calls in more developed legal systems to apply similar rules to those that regulate the keeping of wild animals, since the rules governing wild animals are more stringent.[159] The question arises here, however, of how to draw a distinction with regard to the specific features of artificial intelligence mentioned above. Moreover, stringent rules may actually slow down the introduction of artificial intelligence technologies due to the unexpected risks of liability for creators and inventors.

A widespread suggestion is to apply the norms for legal entities to AI systems.[160] Since a legal entity is an artificially constructed subject of the law,[161] robots can be given similar status. The law could be sufficiently flexible and grant rights virtually to anyone or anything. It could also restrict rights. For instance, historically, slaves had virtually no rights and were mere property. The opposite situation can also be observed, in which entities that do not demonstrate any explicit capacity for action are granted rights. Even today, there are examples of unusual objects recognized as legal persons both in developed and developing countries. In 2017, a law was passed[162] in New Zealand recognizing the status of the Whanganui River as a legal entity. The law states that the river is a legal entity and, as such, has all the rights, powers and obligations of a legal entity. The law has thus transformed the river from property into a legal entity, thereby expanding the understanding of what can and what cannot be considered property. In 2000, the Supreme Court of India recognized the main sacred text[163] of the Sikhs, the Guru Granth Sahib, as a "juristic person."

[158] Arikhipov V., Naumov V. On Certain Issues in the Theoretical Foundations for Developing Legislation on Robotics: Aspects of Will and Legal Personality // Zakon. 2017, No. 5, p. 167.
[159] Asaro P. The Liability Problem for Autonomous Artificial Agents // AAAI Symposium on Ethical and Moral Considerations in Non-Human. Agents, Stanford University, Stanford, CA. p. 193.
[160] Arikhipov V., Naumov V. On Certain Issues in the Theoretical Foundations for Developing Legislation on Robotics: Aspects of Will and Legal Personality // Zakon. 2017. No. 5, p. 164.
[161] See, for instance, Winkler A. We the Corporations: How American Business Won Their Civil Rights. NY: Liverlight, 2018. See a description here: https://www.nytimes.com/2018/03/05/books/review/adam-winkler-we-the-corporations.html
[162] Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 // Parliamentary Council Office. New Zealand. 2017. URL: http://www.legislation.govt.nz/act/public/2017/0007/latest/whole.html
[163] Padmanabha T. Rao. Guru Granth Sahib, a juristic person: SC // The Hindu. 2000. URL: https://www.thehindu.com/2000/04/03/stories/01030005.htm

However, even if we leave aside the most extreme cases and look at ordinary regular companies, legal systems make legal entities liable under civil and, in certain cases, criminal law.[164] Without determining whether a company (or state) has free will or intent, or whether they can act in a premeditated manner or knowingly, they can be recognized as legally responsible for certain actions. Similarly, it is not necessary to ascribe intent or free will to robots to recognize them as responsible for their actions.

However, the analogy of legal entities is problematic since the concept of legal entity is necessary in order to carry out justice in a speedy and effective manner. But the actions of legal entities always go back to those of a single person or group of people, even if it is impossible to determine exactly who they are.[165] In other words, the legal liability of companies and similar entities is linked to the actions performed by their representatives or employees. Moreover, legal entities are only deemed to be criminally liable if it is determined that the individual who performed the illegal action was acting on behalf of the legal entity.[166] The actions of AI-based systems will not necessarily be traced directly back to the actions of an individual.

Finally, legal norms on sources of increased danger may be applied to such systems. Under Paragraph 1 of Article 1079 of the Civil Code of the Russian Federation, legal entities and individuals whose activities are associated with increased danger for the surrounding population (the use of vehicles, mechanisms, etc.) must redress damage caused by a source of increased danger unless they prove that the damage was a result of force majeure circumstances or at the intent of the injured person. The problem is identifying which artificial intelligence systems can be deemed sources of increased danger. The issue is similar to the one mentioned above regarding domestic and wild animals.

National and International Regulation

Many countries are now actively creating the legal conditions for the development of technologies that use artificial intelligence. For example, the "Intelligent Robots Development Distribution Promotion Act"[167] has been in place in South Korea since 2008. The act is aimed at improving the quality of life and developing the economy through the creation and promotion of a strategy for the sustainable development of the smart robot industry. Every five years, the government works out a basic plan to ensure that these goals are achieved.

Two recent examples merit particular consideration here: which has declared its ambitions to become a European and world leader inartificial intelligence; and the European Union, which has put forward advanced rules for the regulation of smart robots.

[164] In countries that use the Anglo-Saxon legal system, the European Union and some Middle Eastern countries. This kind of liability also exists in certain former Soviet countries: Georgia, Kazakhstan, Moldova and Ukraine. It does not exist in Russia, although it is being discussed.

[165] B. Brożek, M. Jakubiec. On the legal responsibility of autonomous machines // Artificial Intelligence Law. 2017. № 25(3). pp. 293–304.

[166] Khanna V.S. Corporate criminal liability: what purpose does it serve? // Harvard Law Review. 1996. No. 109. pp. 1477–1534.

[167] Intelligent Robots Development Distribution Promotion Act // KLRI, 2017.
URL: https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=39153&type=part&key=18

France

In late March 2018, President of France Emmanuel Macron presented[168] the national AI strategy. Over five years, France plans to invest 1.5 billion euros in supporting research and innovation in the field. The strategy is based on recommendations made in the report[169] prepared under the supervision of French mathematician and National Assembly member Cédric Villani. The strategy is geared towards four specific sectors: healthcare, transportation, environment and defense. The objective is to use the potential of comparable advantages and competence areas with a particular focus on sectors where companies could play the key global role; and because these technologies are important for the public interest, etc.

There are seven key proposals, and one is particularly interesting for this article, namely, making AI more open. Indeed, algorithms are mostly proprietary and classified as a commercial secret in many cases. However, algorithms can be biased. For example, in the process of self-learning, they may absorb and adopt the stereotypes that exist in society or which are transferred to them by developers and make decisions based on them. There is already legal precedent for this. In the United States, a defendant was sentenced[170] to a lengthy prison termon the basis of information obtained from an algorithm predicting the likelihood of repeat offences.[171] The defendant's appeal against the use of an algorithm in the sentencing process was rejected because the criteria used to evaluate the possibility of repeat offences were a trade secret and therefore not presented. The French strategy proposes developing transparent algorithms that can be tested and verified, determining the ethical liability of persons working in AI, establishing consulting ethics committee, etc.

The European Union

The creation of the European Parliament Resolution of 2017[172] on the Civil Law Rules on Robotics marked the first step towards the regulation of artificial intelligence. A working group on legal issues related to the development of robotics and artificial intelligence in the European Union was established back in 2015. The resolution is not binding, but it offers the European Commission a series of recommendations on possible actions in the area of artificial intelligence, not only with regard to civil law, but also to the ethical aspects of robotics.

The resolution defines a "smart" robot as a robot that acquires "autonomy through sensors and/or by exchanging data with its environment"; the robot has

[168] AI For Humanity. URL: https://www.aiforhumanity.fr/

[169] Villani C. For a Meaningful Artificial Intelligence // AI For Humanity. 2018. URL: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

[170] Smith M. In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures // The New York Times. 2016. URL: https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretelldefendants-futures.html

[171] Ibid.

[172] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) // European Parliament, 2017. URL: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN

"at least a minor physical support"; adapts "its behaviour and actions to the environment"; does not have "life in the biological sense." According to the document, "a comprehensive Union system of registration of advanced robots should be introduced" managed by "a designated EU Agency for Robotics and Artificial Intelligence." The same Agency would provide technical, ethical and regulatory robotics assessment. As for liability, two options are proposed: either strict liability (without fault) or risk management approach (the liability of the person who could have minimized risks). Liability, according to the document, "should be proportional to the actual level of instructions given to the robot and of its degree of autonomy." Rules on liability could be complemented by a compulsory insurance scheme for robot users, and a compensation fund to pay out compensation in case no insurance policy covered the risk.

The resolution proposes two codes of conduct for dealing with ethical issues: a Code of Ethical Conduct for Robotics Engineers and a Code of Conduct for Research Ethics Commissions. The first code proposes four ethical principles: 1) Beneficence – robots should act in the best interests of humans; 2) Non-maleficence – the doctrine of "first, do no harm," whereby robots should not harm a human; 3) Autonomy – the capacity to make an informed, un-coerced decision about the terms of interaction with robots; 4) Justice – fair distribution of the benefits associated with robotics and affordability of homecare and healthcare robots in particular.

* * *

The examples provided in the article thus demonstrate, among other things, how social values influence the attitude towards artificial intelligence and its legal implementation. Therefore, our attitude to autonomous systems, be it robots or something else, and our reinterpretation of their role in society and their place among us could have a transformational effect. Legal personality determines what is important for society and allows the decision to be made as to whether "something" is a valuable reasonable entity for the purposes of possessing rights and obligations

Due to the specific features of artificial intelligence, suggestions have been put forward regarding the direct responsibility of certain systems.[173] According to this line of thought, there are no fundamental reasons why autonomous systems should not be legally liable for their actions. The question remains, however, about the necessity or desirability of introducing this kind of liability (at least at the present stage).This is due, among other reasons, to ethical problems listed above. Perhaps making programmers or users of autonomous systems liable for the actions of those systems would be more effective. But this could slow down innovation. This is why we need to continue to search for the perfect balance.

[173] Hage J. Theoretical foundations for the responsibility of autonomous agents // Artificial Intelligence Law. 2017. No. 25(3). pp. 255–271.

In order to find this balance, we need to address a number of issues. For instance: What goals are we pursuing in the development of artificial intelligence? Whose interests should prevail? How effective will it be? The answers to these questions will help us to prevent situations like the one that appeared in Russia in the 17th century, when an animal (specifically a goat) was exiled to Siberia for its actions.[174]

[174] Pagallo U. The Laws of Robots. Cries, Contracts, and Torts. Springer. 2013. p. 36.

About Authors

Sergey Karelov, Founder and Chief Technology Officer, Witology, Chairman of Board of League of Independent Experts, RIAC Expert

Anton Kolonin, PhD in Technical Sciences, Founder, Aigents, Architect, SingularityNET, RIAC Expert

Maksim Karliuk, Research Fellow, HSE – Skolkovo Institute for Law and Development, National Research University Higher School for Economics, RIAC Expert

Nikolai Markotkin, Media and Government Relations Manager at the Russian International Affairs Council, PhD Candidate at the Federal Center of Theoretical and Applied Sociology of the Russian Academy Of Sciences

Dimitri Scheftelowitsch, PhD in Natural Sciences (Doktor der Naturwissenschaften), Research fellow at TU Dortmund University, specialist in autonomous decision-making systems, RIAC Expert

Russian International Affairs Council

Russian International Affairs Council (RIAC) is a non-profit international relations think-tank working to provide policy recommendations for all of the Russian organizations involved in external affairs.

RIAC engages experts, government officials and entrepreneurs in public discussions with an end to increase the efficiency of Russian foreign policy.

Along with research and analysis, the Russian Council is involved in educational activities to create a solid network of young global affairs and diplomacy experts.

RIAC is a player on the second-track and public diplomacy arena, contributing the Russia's view to international debate on the pending issues of global development.

RIAC members are the thought leaders of Russia's foreign affairs community – among them diplomats, businessmen, scholars, public leaders and journalists.

RIAC President Igor Ivanov, Corresponding Member of the Russian Academy of Sciences, served as Minister of Foreign Affairs of the Russian Federation from 1998 to 2004 and Secretary of the Security Council from 2004 to 2007.

Director General of RIAC is Andrey Kortunov. From 1995 to 1997, Dr. Kortunov was Deputy Director of the Institute for US and Canadian Studies.

Notes

RUSSIAN INTERNATIONAL AFFAIRS COUNCIL (RIAC)
1, B. Yakimanka street, 119180, Moscow, Russia
Tel.: +7 (495) 225 6283
Fax: +7 (495) 225 6284
E–mail: welcome@russiancouncil.ru
www.russiancouncil.ru