Legal Issues in the Digital Age. 2024. Vol. 5. No. 2.
Вопросы права в цифровую эпоху. 2024. Том 5. № 2.

Research article
УДК: 340, 347
DOI:10.17323/2713-2749.2024.2.4.35

Legal Horizons of the New Artificial Intelligence Paradigm

Aleksandr Amiranovich Kartskhiya

National Oil and Gas Gubkin University, 65 Lenin Avenue, Moscow 119991, Russian Federation, arhz50@mail.ru, Web of Science Researcher ID AAZ-1083-2020, ORCID 0000-0002-8041-0055, Scopus ID 57217114108, AuthorID 771380

Abstract

Modern society is undergoing a structural transformation of the world economy. This is as a result of the transition to a new technological base through the introduction of artificial intelligence, cutting-edge information and communication technology, energy technology, biotechnology and nanotechnology. Artificial intelligence has the ability to significantly change the economy and social relations in society, and its newly discovered capabilities are transformational and global in nature. At the same time, the extraordinary capabilities of artificial intelligence technologies involve risks that can threaten stability and undermine human values. In order to eliminate possible threats and risks and mitigate potential dangers, it is crucial to develop systemic legal measures and ways to regulate AI technologies and models on a national and international scale and to define the legal status of AI, which must include protection of humans from the uncontrolled influence of AI and the inviolability of guarantees of human rights and freedoms. With this in mind, and in order to mitigate potential dangers and ensure the controllability and sustainability of AI technologies based on the concept of trusted (responsible) AI, it is necessary to agree on universal international guidelines for the development and application of AI technologies and models. Furthermore, it is necessary to create a universal code of conduct for AI developers, who together can create a basis for a uniform framework of legal regulation within the national legislation of each country on the principles of human rights protection, privacy and data protection, transparency and explainability, fairness, accountability and safety of artificial intelligence, adequate human oversight and ethical standards for the creation and application of AI models.

Keywords

intellectual property; cybersecurity; trusted intelligence; neural networks; liability; artificial intelligence security; technological sovereignty; threats and risks of AI technology; sustainable development.

For citation: Kartskhiya A.A. (2024) Legal Horizons of the New Artificial Intelligence Paradigm. *Legal Issues in the Digital Age,* vol. 5, no. 2, pp. 4–35. DOI:10.17323/2713-2749.2024.2.4.35

AI has hacked the operating system of human civilization
Yuval Noah Harari. The Economist, April 23th 2023

For humanity's sake, regulation is needed to tame market forces
Helen Toner and Tasha McCauley, former OpenAI Board member.
The Economist, May 26th 2024

Introduction

Over the past decades, scholars have dissected the manifold ways in which artificial intelligence (AI) systems and digital technologies impact pillars of the law in fields such as human rights law, constitutional law, criminal law, tortious liability and contracts, administrative law, international humanitarian law, and more [Barfield W., Pagallo U., 2020: 25]. According to the European Commission High-Level Expert Group (2018)[1], the challenges brought forth by AI in the legal domain depend on the complexity, opacity, openness, autonomy, predictability, data-drivenness, and vulnerability of computers that mimic human intelligence.

A recent survey showed that the most common AI technologies are ChatGPT, Microsoft CoPilot, Character AI for text and code; Midjourney, Stablefusion, and Dalle3 for image generation; and Parker AI, Runway, and Google Gemini for multi-models (which can combine text, images, and video)[2].

Such technologies require significant financial outlays and technical development. According to the Economist, as an example, Elon Musk's

[1] High-Level Expert Group on Artificial Intelligence Draft Ethics Guidelines for Trustworthy AI, European Commission. 2018. Available at: https://www.euractiv.com/wp-content/uploads/sites/2/2018/12/ AIHLEGDraftAIEthicsGuidelinespdf.pdf (accessed: 10.04.2024)

[2] Available at: https://bristolcreativeindustries.com/(accessed: 10.04.2024)

start-up had raised $6bn. The investors, such Silicon Valley stalwarts as Sequoia Capital and Andreessen Horowitz, two venture-capital giants, and an investment fund with ties to the Saudi royal family put AI's financial firepower in the big league, alongside model-builders such as OpenAI, the creator of ChatGPT, and Anthropic (see Fig. 1)[3].

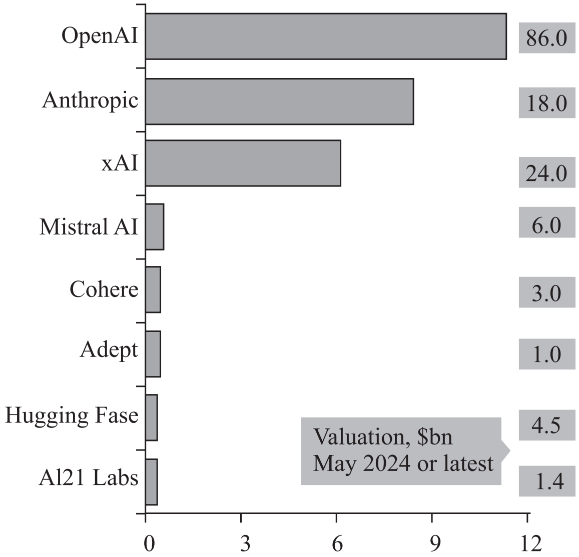

Fig. 1. The money isn't artificial. Al startups, cumulative capital raised, $bn

Source: The Economist. May 30th, 2024.

Moreover, the rumoured Apple-OpenAI deal represents a significant collaboration between two tech giants, promising to integrate OpenAI's advanced generative AI technology into Apple's software ecosystem. Apple is poised to enter the AI landscape in June 2024, and people think the announcement of the Apple OpenAI deal will be made that day alongside new iOS[4].

The fact the development and regulation of artificial intelligence is relevant is also evident on the international agenda. Thus, in November 2023 several countries, including the United States, China, the European

[3] Can Elon Musk's x AI take on Open AI? The Economist. May 29, 2024.Available at: https://www.economist.com/ business/2024/ 05/29/can-elon-musks-xai-take-on-openai (accessed: 11.04.2024)

[4] Available at: https://dataconomy.com/2024/05/31/chatgpt-apple-openai-deal/ (accessed: 11.04.2024)

Union, the United Kingdom, France, Italy, India, Brazil, Japan, Saudi Arabia, United Arab Emirates (Russia did not participate) held the first international summit and have approved a Declaration on the Artificial Intelligence Safety (The Bletchley Declaration on AI Safety)[5]. The declaration expresses a shared understanding of the opportunities and risks associated with artificial generative intelligence and states the urgent need to recognise and collectively manage the potential risks of AI through a new collaborative global effort to ensure the safe and responsible development and deployment of advanced AI. The participating countries agreed that significant risks could arise from potential intentional misuse or unintentional difficulties with control over advanced AI. Cyber security, biotechnology and disinformation risks are of particular concern in this connection. The Declaration notes the potential for serious, even catastrophic harm, intentional or unintentional, arising from the most significant capabilities of AI technologies and models. Among the main risks that the Declaration highlights are bias and breach of confidentiality in the application of AI.

The so-called Hiroshima Process organized by a number of Western countries was another important international event in the world of AI in recent times. On 30 October 2023 in Hiroshima, Japan, the G7 group of countries has approved a joint G7 Leaders' Statement on the Hiroshima AI Process, which proclaimed the International Guiding Principles on Artificial Intelligence and recommended a Code of Conduct for AI developers containing a set of rules AI developers are encouraged to follow on a voluntary basis to mitigate risks throughout the AI lifecycle.

By signing the Declaration, the parties have agreed that the risks posed by AI are inherently international and can be best addressed through international co-operation. The signatories agreed to co-operate in an inclusive manner to ensure the creation of a human-centred, trustworthy and responsible artificial intelligence.

The International AI Safety Summit and Declaration mentioned focused on "Frontier Artificial Intelligence" (Frontier AI) — highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced model." Frontier AI is a subset of AI focused on highly advanced general purpose AI models, including foundation models that may have capabilities equal to

[5] Available at: https://www.gov.uk/government/news/countries-agree-to-safe-and-responsible-development-of-frontier-ai-in-landmark-bletchley-declaration (accessed: 11.04.2024)

or greater than the most sophisticated modern systems (e.g., narrower than the scope of the EU AI Act). Today, the most advanced general-purpose language models for large languages are, e.g., OpenAI GPT-4 and Google PaLM 2.

It has been declared that advanced AI (Frontier AI) systems pose significant security risks, especially in areas such as cybersecurity and biotechnology. Concerns arise from the potential for misuse, control issues and increased risks such as misinformation. However, the crucial difference between narrow models and general purpose models is that the latter are often made available through "broad deployment" via sector-agnostic platforms such as APIs, chatbots or open sourcing, and as such "can be integrated into a large number of diverse downstream applications possibly including safety critical sectors."

The Bletchley Declaration on AI Safety is not legally binding and is more a symbol than a detailed roadmap. Yet, the conference participants have agreed that it is necessary to:

Identify the security risks associated with AI, develop a common, evidence-based understanding of those risks, and maintain that understanding as opportunities arise in the context of a broader global approach to understanding the influence of AI on society, and

Based on the risks identified, develop appropriate policies in their countries to ensure secure countering of such risks: increased transparency accompanied by the adoption by private companies of advanced AI capabilities, appropriate assessment indicators, security testing tools, and the development of appropriate public sector capacity and research.

The proposed Code of Conduct contains a non-exhaustive list of recommendations for entities developing the most advanced artificial intelligence systems. These entities must operate on the basis of risk assessment at all stages of the lifecycle, including the design, development, deployment, and use of advanced AI systems. The AI development process consists set of actions, namely:

(a) identify, assess and mitigate risks throughout the AI lifecycle;

(b) develop and implement an AI and risk management policy based on a risk-based approach;

(c) develop and implement robust mechanisms for authenticating content and its origin, including watermarks or other methods that allow users to identify content created by AI;

(d) prioritise the development of advanced AI systems to address the world's most important challenges, in particular the global climate agenda, health, education, and others;

(e) implement appropriate measures to protect intellectual property and personal data.

Along with these international acts, the Resolution on Artificial Intelligence "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development"[6] (hereinafter "the Resolution") passed on 21 March 2024 by the UN General Assembly is of key significance. Supported by more than 120 member states, the Resolution aims to encourage countries to protect human rights, safeguard personal data and monitor AI for risks on a non-legally binding basis. Though the UN does not have the ability to pass laws or regulations regarding AI or its implementation, the UN Charter gives to the General Assembly the power to initiate studies and make recommendations to promote the development and codification of international law. The main purpose of the document is to ensure "safe, secure and trustworthy AI systems" on a global level. It encourages all 193 Member States and multi-stakeholders from all regions and countries (private sector, international and regional organizations, civil society, the media, academia and research institutions and technical communities and individuals) to develop and support regulatory and governance frameworks.

The Resolution claims improper or malicious design, development, deployment and use of artificial intelligence systems, e.g., without adequate safeguards or in a manner inconsistent with international law, pose risks that could hinder progress towards the achievement of the 2030 Agenda for Sustainable Development and its Sustainable Development Goals and undermine sustainable development in its three dimensions — economic, social and environmental; widen digital divides between and within countries; reinforce structural inequalities and biases; lead to discrimination; undermine information integrity and access to information; undercut the protection, promotion and enjoyment of human rights and fundamental freedoms, including the right not to be subject to unlawful or arbitrary interference with one's privacy; and increase the potential risk for accidents and compound threats from malicious actors.

[6] Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. UN General Assembly. March 2024. Available at: https://ai.gov.ru/ (accessed: 11.04.2024)

At the same time, while the Resolution does not define "Artificial Intelligence," it does set out provisions of secure and "trustworthy artificial intelligence systems" which refers to artificial intelligence systems in the non-military domain, whose life cycle includes the stages: pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning. The systems are referred to as human-centric, reliable, explainable, ethical, inclusive, in full respect, promotion and protection of human rights and international law, privacy preserving, sustainable development oriented, and responsible. According to the Resolution, such AI systems have the potential to accelerate and enable progress towards the achievement of all 17 Sustainable Development Goals and sustainable development in its three dimensions — economic, social and environmental — in a balanced and integrated manner; promote digital transformation; promote peace; overcome digital divides between and within countries; and promote and protect the enjoyment of human rights and fundamental freedoms for all, while keeping the human person at the centre.

However, as the Resolution has no enforcement powers on its Member States, there are no regulators under the Resolution, nor does the Resolution stipulate how the Member States should regulate AI systems in their own jurisdictions. As the Resolution is not legally binding, it does not confer enforcement powers or give rise to any penalties for non-compliance.

At the same time, as the Foreign Policy Concept of the Russian Federation[7] notes, mankind is going through an era of revolutionary change. This is primarily due to a structural transformation of the world economy as a result of the transition to a new technological base through the introduction of AI, cutting-edge information and communication technology, energy technology, biotechnology and nanotechnology. Other reasons include the growth of national identity, cultural and civilizational diversity and other objective factors that accelerate the redistribution of development potential to new centres of economic growth and geopolitical influence, and contribute to the democratisation of international relations.

It seems reasonable to agree with the view that the advent of Generative AI marks a paradigm shift in the AI landscape, the complexity and emergent autonomy of AI models introduce challenges in predictability and legal compliance [Novelli C., Casolari F., 2024: 1–2].

[7] Decree of the President of the Russian Federation of 31 March 2023 No. 229 "On Approval of the Concept of the Foreign Policy of the Russian Federation" // Collection of Laws of the Russian Federation. 03 April 2023. No. 14. P. 2406.

However, as AI capabilities become more powerful, the growing use of AI systems, as analysts believe [Brundage M. et al., 2018: 5–6], could lead to changes in the threat landscape, which can be categorised as follows: the scalable application of AI systems to perform tasks previously performed by humans — as a result, we see an expansion of existing threats; new threats posed by evolving technologies and AI models; the increasing use of artificial intelligence systems for malicious purposes significantly expands the range of AI applications, types of threats and risks. Three areas of security for AI systems can be distinguished here:

Digital security. The use of AI offers the potential to significantly increase the efficiency of cyber-attacks, which will create new threats by exploiting human vulnerabilities in the form of phishing, speech and image synthesis (deep fakes) or data leakage.

Physical security. The use of unmanned aerial, surface and underwater vehicles and other automated systems (including autonomous weapon systems, microdrone swarms, etc.), as well as attacks on cyber-physical systems (in transportation and industry) or critical infrastructure.

Political security. The use of AI to collect and analyse data for targeted propaganda or manipulation of consciousness and public opinion by violating privacy or analysing and manipulating people's behaviour, attitudes and beliefs on the basis of available data.

It is noteworthy that the National Strategy of the Russian Federation for the Development of Artificial Intelligence for the period until 2030[8] proclaims that the goals of AI development are to ensure the growth of welfare and quality of life of the population, ensure national security and law and order, and achieve sustainable competitiveness of the Russian economy, which includes global leadership globally in the field of AI. According to the National Security Strategy of the Russian Federation,[9] in order to ensure and protect the national interests of Russia from external and internal threats, including unfriendly actions of foreign states, the Russian Federation should more efficiently use its achievements and competitive advantages with account for long-term global trends. In order to solve the tasks in

[8] Decree of the President of the Russian Federation of 10 October 2019 No. 490 On Development of Artificial Intelligence in the Russian Federation // Collection of Laws of the Russian Federation, 14 October 2019, No. 41. P. 5700.

[9] Decree of the President of the Russian Federation of 02 July 2021 No. 400 On the National Security Strategy of the Russian Federation // Collection of Laws of the Russian Federation, 05 July 2021, No. 27 (Part II). P. 5351.

the sphere of national security, AI is used as a tool to ensure information security based on the application of advanced technologies. This includes AI and quantum computing technologies as a means of upgrading industrial enterprises and infrastructure, digitalisation to improve labour productivity and boost development of Russia's scientific and technological base, nanotechnology, robotics, medical, biological, genetic engineering, information and communication, big data processing, energy, laser, additive, creation of new materials, cognitive, and nature-like technologies.

In this situation the importance of comprehensive research into the development of AI and its new paradigm, including legal issues of the application of AI technologies in the digital economy, increases [Naumov V.B. et al., 2023].

1. Modern Legal Aspects of Artificial Intelligence Technologies

The current understanding of artificial intelligence gains particular importance at this time. E.g., the OECD definition contained in the OECD AI Principles 2019 built on the conceptual view of AI detailed in paper "Artificial Intelligence: A Modern Approach" by S. Russell and P. Norvig [Russell S., Norvig H., 2009]. It reads: "An AI system is a machine-based system that can, for a given set of human defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy."

This is in line with the updated definition of AI given in the OECD Memorandum 2023[10], which was formulated with the aim to harmonise and provide legal certainty for universal application. The updated definition reads as follows: «an AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment».[11]

The text above is replaced with the following updated definition: An AI system is a machine-based system that can, for a given set of human-de-

[10] OECD. Explanatory memorandum on the updated OECD definition of an AI system. OECD Artificial Intelligence Papers. 2024. No. 8. Available at: https://doi.org/10.1787/623da898-en. (accessed: 11.04.2024)

[11] Explanatory memorandum on updated OECD definition of AI System. Paris, 2024. Available at: http://www.oecd.org/termsandconditions . (accessed: 10.04.2024)

fined explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical real or virtual environments. Different AI systems are designed to operate with varying in their levels of autonomy and adaptiveness after deployment.

The earliest example of generative AI is a much simpler model known as the Markov chain. The method was named in honour of Andrei Markov, a Russian mathematician who in 1906 introduced this statistical method for modelling the behaviour of random processes. In machine learning, Markov models have long been used to predict the subsequent word, similar to the autocomplete function in an email programme. In text prediction, the Markov model generates the next word in a sentence by looking at the previous word or several previous words. The current basic AI models underlying ChatGPT and similar systems work in much the same way as the Markov model. But ChatGPT is much bigger and more complex: it has billions of parameters and is trained on huge amounts of data, mostly publicly available content on the Internet. In this huge body of text, words and sentences appear in sequences with certain dependencies. This helps the AI model to understand how to break the text into statistical chunks that have some predictability. AI learns the patterns of such blocks of text using this knowledge to suggest a particular solution [Zewe A., 2023].

The concept of Artificial Intelligence or Artificial Intelligence Systems usually includes categories of methods such as machine learning, and knowledge-based approaches and applications such as computer vision, natural language processing, speech recognition, intelligent decision support systems, intelligent robotic systems, and the application of these tools in various domains. Artificial intelligence technologies are advancing at a rapid pace, and additional methods and applications may be created in the future.

Usually, Generative AI (“GAI”) uses neural networks and other algorithms to create, through machine learning, new data or content similar to the original data.

The Generative AI model refers to generative modelling that is instantiated with a machine learning architecture (deep neural network) and, therefore, can create new data samples based on learned patterns. A generative AI system encompasses the entire infrastructure, including the model, data processing, and user interface components. The model serves as the core component of the system, which facilitates interaction and application within a broader context. Deep neural networks are particularly well suited for the purpose of data generation, such as diffusion probabilistic models for text-to-image generation or the transformer architecture and (large) language models (LLMs) for text generation. Generative AI is a branch of AI that can create new content such as texts, images, or audio that increasingly often cannot be distinguished anymore from human craftsmanship [Feuerriegel S. et al., 2024: 112-113].

Large generative AI models that can model output in and across specific domains or specific data types in a comprehensive and versatile manner are oftentimes also called foundation models [Bommasani R. et al., 2021: 4-5].

Generative AI is of immense importance for various industries such as media, arts, entertainment, advertising, and education. That said, it may also pose certain threats due to copyright infringement, dissemination of false or discriminatory information, and loss of control over the content created. Generative AI will have significant economic implications across various industries and markets. Generative AI can increase efficiency and productivity by automating many tasks that were previously performed by humans, such as content creation, customer service, code generation, etc. This can reduce costs and open up new opportunities for growth and innovation [Eloundou T. et al., 2023: 5].

Unlike GAI, descriptive AI based on machine learning is used to analyse, classify and make predictions from raw data, and to identify the data structure, dependencies and trends without creating new data. Descriptive AI can be used for various purposes such as: (a) classification, i.e., dividing data into groups based on their characteristics or attributes (classification of electrocardiograms into normal and abnormal, diagnosis of diseases, etc.); (b) regression, i.e. predicting unknown values based on known data (weather forecast, stock quotes, etc.); (c) clustering, i.e. dividing data into groups based on similarities between elements (business process modelling, etc.); (d) trend analysis, i.e. identifying trends and dependencies in data to provide information about future events or changes. Descriptive AI is the basis for many modern technologies such as recommender systems, automatic sound and image processing systems, quality control systems, and risk management systems. Although descriptive AI does not generate new data, it can provide important information and knowledge that can be used for decision making, planning and strategic planning.

Particular attention is paid to the definition of a conceptual approach to trusted artificial intelligence. In particular, the OECD[12] documents

[12] OECD 2023. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-

outline the principles of responsible governance of trustworthy AI, which complement each other and should be considered as a whole. These include, inter alia:

inclusive growth, sustainable development and well-being that involve engaging in responsible governance of trustworthy artificial intelligence to enhance human capabilities and creativity, promote inclusion, reduce economic, social, gender and other inequalities, and protect the environment, thereby promoting inclusive growth, sustainable development and well-being;

respect for the rule of law and human rights (freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, integrity, social justice and internationally recognised labour rights) and, to this end, the implementation of appropriate mechanisms and safeguards that are relevant to the context and in line with the state of the art;

transparency and lucidity, i.e., responsible disclosure of meaningful information about AI systems that is relevant to the context and consistent with the prior art;

reliability and security of the AI throughout its life cycle so that it functions properly and does not pose an unreasonable risk to safety under conditions of normal use, foreseeable use or misuse, or other adverse conditions;

accountability: AI agents should be responsible for the proper functioning of AI systems and for complying with the above principles based on their roles, context, and in accordance with the prior art.

At the same time, a new kind of AI self-developing artificial intelligence has already been developed. According to researchers at the Massachusetts Institute of Technology and the University of California (Fox News)[13], AI subsystems can be created without human assistance. Larger AI models like those used by ChatGPT can build on the “parent” algorithm to create smaller, specific AI applications that can be used, for example, to improve hearing aids, control oil pipelines, or monitor endangered wildlife.

But artificial AI technology continues to improve, and we see agentic AI models emerge. In comparison with General AI, a new model of AI,

LEGAL-0449; OECD.2020. Digitalization and Responsible Business Conduct: Stocktaking of policies and initiatives. Available at: https://www.oecd.org/daf/inv/mne/ publicationsdocuments/ reports/2/ (accessed: 11.04.2024)

[13] Available at: https://vfokuse.mail.ru/article/uchenye-zayavili-o-vozmozhnosti-ii-vosproizvoditsya-bez-uchastiya-cheloveka-59040575/ (accessed: 11.04.2024)

agentic AI is a more flexible system that could enable increased automation and worker productivity in certain types of industries and assist those who lack digital literacy. Large Action Model (LAM) adopts a learning-by-demonstration approach, observing human interactions with interfaces and replicating these actions reliably. AI systems that understand digital interfaces typically designed for humans and learn to execute human actions autonomously within these digital environments. AI agent might be able to interact with apps or websites, add items to a shopping cart and check out in accordance with pre-registered preferences and payment options, fill out and submit a form, or RSVP to an event. As an example, the recently released Humane AI Pin is attached to the user's shirt and acts as an AI-based digital assistant that responds to touch and voice and shows a laser projection on the user's palm; various smartphones and other hi-tech equipment are now equipped with an AI assistant [Pathirannehelage H. et al., 2022: 2]; [Aggarwal R., Singh H., 2024: 3].

Among other things, artificial intelligence, offering innovative solutions and analytical insights, has a great potential to shape the sustainable development model, revolutionise environmental and social processes, and scale the ESG model of corporate governance. There are many ways to realise the potential of AI to advance ESG-based sustainable development, offer innovative solutions to complex economic and governance challenges, and apply socially responsible practices. AI helps in developing strategies and planning scenarios for risk assessment and mitigation and customised risk management solutions tailored to specific industries and their unique challenges, including ESG risk mitigation. AI's ability to process complex data, predict trends and offer useful analytics is key to improving productivity and creating new business models for corporate governance. By harnessing the power of AI, companies can not only comply with regulations, but also introduce innovations, be competitive and comply with ethical business practices. However, AI should be seen as a complement to, not a replacement for, humans in their decision-making process. Successful integration of AI into management practices depends on a synergistic interrelation between technology and human understanding where AI acts as an enabler or catalyst for more informed, ethical and sustainable business decisions.

Generative Artificial Intelligence (GAI) has experienced dramatic growth recently and is accompanied, among other things, by growing challenges to the protectability of AI results in the intellectual property realm. Currently, a legal regime of artificial intelligence authorship and patent protection for AI technologies is being actively developed in various countries [Ivliev G.P., Egorova M.A., 2022: 32–46]; [Tikhomirov Yu.A. et al., 2019]; [Rozhkova M., 2021: 14–22]; [Morhat P., 2018: 1–8]; [Kharitonova Y., Savina V., 2020: 524–549].

The rapid development of artificial intelligence, and generative AI in particular, has created a whole maze of new copyright issues. These questions are primarily related to the way in which AI models are trained and whether the results of the development of these models constitute independently protectable subject matter so that they would be eligible for copyright protection.

The main question is whether works created by AI possess enough creativity to qualify for copyright protection. There is an opinion that works that are created with textual prompts and do not require any additional creative input from a human user, as in the case of generative AI tools, are not protected by copyright because these prompts are more like instructions for the commissioned artist.

The judicial practice in this area is not yet extensive, but is also interesting. E.g., in 2023, the court in Washington in *THALER vs. US Copyright Office* has ruled that only works with human authors can receive copyrights as human authorship is a "bedrock requirement of copyright" based on "centuries of settled understanding." According to the judgement copyright has never stretched far enough "to protect works generated by new forms of technology operating absent any guiding human hand, as plaintiff urges here. Human authorship is a bedrock requirement of copyright." The USA Copyright Act of 1909 explicitly provided that only a 'person' could 'secure copyright for his work' under the Act. Similarly, 9th Circuit appeals court ruling in 2018 that a monkey who took a selfie "could not sue under the Copyright Act for the alleged infringement of photographs this monkey had taken of himself, for 'all animals, since they are not human' lacked statutory standing under the Act."[14] Thaler was not able to point to any case "in which a court has recognized copyright in a work originating with a non-human".[15]

Likewise, in India it was decided that a work must involve a minimum degree of creativity and not be a product of only skill and labour. There-

[14] U.S. Copyright Office. Compendium of U.S. Copyright Office Practices § 101. 2021. Available at: https://copyright.gov/ comp3/ (accessed: 09.04.2024)

[15] Available at: https://www.digitalmusicnews.com/wp-content/uploads/2023/08/thaler-perlmutter-copyright-generative-AI-aug-2023.pdf (accessed: 11.04.2024)

fore, output produced by AI may not satisfy the requirement of ''creativity'' required for copyright protection, if viewed as a collection of data compiled from already existing sources without any infusion of creativity. In this sense, Indian and US copyright law agree that a certain class of AI-generated works would not qualify for copyright.

Interestingly, the Beijing Internet Court's decision in *Li v. Liu* (China) makes a distinction between ''straightforward'' AI generated output where the human author simply takes and uses the output "as is" without any creative involvement and AI generated output where the human author keeps experimenting and adding various prompts, including negative prompts and tech parameters, until they receive a satisfactory result. In the later scenario, the Beijing court determined that such "AI-assisted work" (meaning output where aesthetic choices were exercised and there was personal judgement in the final rendition) would be eligible for copyright protection [Hill M., Hackworth A., 2023].

Another problem of artificial intelligence machine learning is related to algorithmic fairness that aims to address and rectify biases often embedded in machine learning systems. These biases can lead to discrimination in automated decision-making processes. Certain principles such as transparency, explainability and accountability are fundamental to developing artificial intelligence applications if the aim is to turn existing risks of discrimination into an opportunity for increased equality and these principles are respected along the entire algorithmic design chain [Xenidis R., Senden L., 2020: 160].

At the same time, in the process of building AI technologies, developers train the models by providing a huge amount of content to improve the model's predictive abilities. But much of this content is copyrighted, and training a model on copyrighted material is itself a copyright infringement, even if the model does not reproduce the exact text as part of its output.

GAI raises copyright infringement concerns in several ways. Firstly, there is the problem of content created by artificial intelligence, or GAI itself, which possibly violates copyright on licensed use. Granting copyright to works created by AI has been widely debated because copyright laws traditionally protect only human-created works. Some experts believe that the content created by AI lacks human creativity and therefore does not fulfil the criteria of copyright. According to another viewpoint, copyright can be granted for GAI model creators generating such content. Another problem arises from the use of copyrighted data to train GAI models (so-called training data). The information sources that AI models use for training are copyrighted: text, images, and music. Arguments in defence of this practice are that using copyrighted data to train GAI models is fair use, while others argue that it constitutes infringement [Ivliev G., Egorova M., 2022: 46]; [Kirsanova E., 2023: 36-46].

The creation of new content and branding by AI based on compiled datasets or data stores, including visual elements such as logos, illustrations and textual elements such as image tags are assessed for legitimacy of using this data in new content. Previously, AI developers and vendors have disclaimed liability for any disruption resulting from their AI-based platforms. The key issue here is to determine who is liable for the content created by artificial intelligence: the AI user or the AI owner (provider). Generative AI companies usually publish disclaimers for the results of their AI platforms.

Recently, however, there has been a positive development: large AI vendors in some cases allow liability in the form of compensation for AI-generated content. But even those companies that have begun to offer compensation limit the protection by granting such rights generally to high-paying subscription tiers to the relevant AI applications. Amid growing scepticism about the use of AI, key industry players have formulated policies to ensure copyright protection for their users. E.g., Microsoft has introduced the CoPilot Copyright Commitment[16] where the company assumes liability for potential consequences arising from Microsoft's use of AI, the services of the second pilot and their outcomes. In addition, Microsoft commits to protecting its users from any third-party claims arising from such use.

Legal protection and defence of developments and technical patentable results created using GAI models is a problem in its own right. E.g., at the EPO, inventions involving AI are considered "computer-implemented inventions." Computer-implemented inventions are treated differently by patent offices in different jurisdictions, and in Europe, computer programs "as such" are excluded from patent protection. Nonetheless, software-related inventions remain eligible for patentability provided they exhibit a discernible technical character.

Over the years, the European case law has established a stable and predictable framework for the patentability of computer-implemented inventions, including inventions related to AI [Voller K., 2024].

[16] Available at: https://www.microsoft.com/en-us/licensing/news/microsoft-copilot-copyright-commitment (accessed: 12.04.2024)

An example is the case of Designation of inventor/DABUS (case J 0008/20) concerned two patent applications filed at the EPO (namely EP18275163 and EP18275174) where the applicant, Stephen Thaler, the inventor was noted to be "DABUS" — an AI created by Thaler himself. The EPO had rejected both applications on the grounds that the designated inventor, DABUS, did not meet the requirements for an inventor, that being a need for them to be a 'natural person'. Thaler subsequently appealed both decisions to the Board of Appeal with the opposition — whether an AI can be an inventor of a patent. The Board firmly rejected this point, as "under the European Patent Convention the designated inventor has to be a person with legal capacity". Further, Article 61 of the EPC notes that "[t]he right to a European patent shall belong to the inventor or his successor in title" (the latter being a legal successor in the title of the rights), and the rights of any employee, if they are the inventor, will be determined by the national legislation where they are employed. The Board clearly set out that "designating a machine without legal capacity can serve neither of these purposes" [17].

The UK Supreme Court has also firmly rejected the idea that a machine with AI can be recognised as an inventor under the UK Patents Act 1977. Addressing the ownership of inventions generated by DABUS, the court concluded that Dr. Thaler failed to establish a legal basis for claiming patent rights based on his ownership of the AI machine. It affirmed that Dr. Thaler had no independent right to obtain a patent for technical advances made by DABUS. The court judgement stipulated that, "it is not and has never been Dr. Thaler's case that he was the inventor and used DABUS as a highly sophisticated tool. Had he done so, the outcome of these proceedings might well have been different." The ownership of AI generated inventions is thus likely not an issue, provided a human inventor is identified, per the formal requirements[18].

Usually, the application of AI (including machine learning ("ML") and specific technical implementations of AI can be patented in Europe. However, fundamental algorithmic or mathematical level AI innovations typically fall outside the scope of patentability.

[17] The EPO Are Not 'Board' of AI Yet — EPO Board of Appeal Weighs in on Whether Artificial Intelligence Can Be an Inventor. 2022. Available at: https://www.ipiustitia.com/2022/08/the-epo-are-not-board-of-ai-yet-epo.html (accessed: 11.04.2024)

[18] The UK Supreme Court Judgement, December 20, 2023, Thaler vs Comptroller-General of Patents, Designs and Trademarks UK. Available at: https://www.supremecourt.uk/cases/uksc-2021-0201.html (accessed: 11.04.2024)

For AI to qualify for patent protection it must leave the abstract realm. This can be achieved in two ways. Firstly, the AI serves a technical purpose by addressing a technical challenge within a particular technology field, demonstrating its application in solving a specific technical problem. Secondly, the invention is directed to a specific technical implementation of AI motivated by technical considerations of the internal functioning of a computer, for example a specific technical implementation of neural networks by GPUs.

Generally, AI inventions are sensitive to the choice of network architecture, input representation, and training data. Since a specific technical purpose or implementation of the AI must be demonstrated, fundamental AI/ML improvements are generally not patentable. General purpose AI or generic AI with algorithmic efficiency are also not patentable.

The leading countries in the field of AI development relying on the active government support are rapidly developing national AI technologies. After the development of Deep Mind and the launch of the US-based Open AI ChatGPT in November 2022, public launches of similar LLM-based technologies in other countries followed. In November 2023, a government-backed AI company AI71 was launched in Abu Dhabi, UAE, to commercialise the LLM Falcon AI model. In December of the same year, the massive funding for the French AI Mistral was announced. India has been developing the models LLM Krutrim and Sarvam. States and private companies in the US, China, UK, France, Germany, India, Saudi Arabia and the UAE have massively funded AI development and expanded national production of graphics processing units and other elements necessary for AI development[19]. Russia has worked in a similar area and has certain achievements in neural networks, e.g., SBER (RuGPT-3).

These days, artificial intelligence finds more and more applications. E.g., AI arbitration is a relatively new concept that involves the use of artificial intelligence (AI) in the process of resolving disputes that exploits algorithms to analyse data related to the dispute and make recommendations on how it should be resolved. The use of AI can help to speed up the process

[19] Welcome to the era of AI nationalism. The Economist. January 1, 2024. Available at: https://www.economist.com/ business/2024/01/01/welcome-to-the-era-of-ai-nationalism?utm_content=article-link-2&etear=nl_ today_2&utm_ campaign=a.the-economist-today&utm_medium=email.internal-newsletter.np&utm_source=-salesforce-marketing-cloud&utm_term=1/1/2024&utm_id=1840347 (accessed: 01.04.2024)

of resolving disputes therefore, as the algorithms can analyse large amounts of data quickly and make recommendations in a timely manner that can be done through the use of smart contracts wherein the terms of the agreement and dispute resolution written directly into lines of code. However, there are also potential challenges to using AI in arbitration. One concern is that the algorithms may not be able to fully account for all of the nuances and complexities that can arise in legal situations. Additionally, there may be legal and regulatory issues that need to be addressed before AI arbitration can be widely adopted. For example, there may be concerns about the accountability and transparency of the algorithms used, and how breaches or damages would be handled. AI, which refers to the ability of machines to perform tasks that would normally require human intelligence, can be used to analyse data, make decisions, and optimise processes, as well as, secure a wide range of transactions, including those related to supply chain management, financial instruments, and identity verification.

The problem of AI risks and threats in the field of cyber security takes a special place.

2. Legal Aspects of Current Regulation of Artificial Intelligence

2.1. Legal AI Regulating in Russia

In Russia, the National Strategy for the Development of Artificial Intelligence for the period until 2030 (the “Strategy”), approved in 2019 and substantially extended in 2024, stipulates the following goals of AI development: ensure the growth of welfare and quality of life of the country’s population; ensure national security, law and order; achieve sustainable competitiveness of the Russian economy, including its leading positions in the world in the AI area.

The concept of artificial intelligence has been clarified in the new version of the Strategy, where AI is defined as a set of technical solutions that allow imitating human cognitive functions (including search for solutions without a predetermined algorithm) and obtaining results comparable to or exceeding the results of human intellectual activity when performing specific tasks. The set of technical solutions includes information and communication infrastructure, software (including software that uses machine learning methods), and processes and services for data processing and solution search. The Strategy defines the artificial intelligence model. It is a computer programme (a component of such a programme), which is designed to perform intellectual tasks at a level comparable to or exceeding the results of human intellectual activities and uses algorithms and data sets to deduce patterns, make decisions or predict results.

The Strategy contains new concepts, including:

large generative models of AI that are capable of interpreting (providing information based on queries, e.g., about objects in an image or about a text) and creating multimodal data (texts, images, videos and the like) at a level comparable to or superior to the results of human intellectual activity;

large fundamental models, i.e., AI models that (1) are the basis for creating and refining various types of software, (2) have been trained to recognise certain types of patterns, (3) contain at least 1 billion parameters, and (4) are used to perform a large number of different tasks;

promising AI methods, i.e. methods aimed at creating fundamentally new scientific and technical products, including the development of universal (strong) AI (ability to solve various problems independently, automatic design of physical objects, automatic machine learning, algorithms for solving problems based on data with partial partitioning and (or) insignificant amounts of data, information processing based on new types of computing systems, interpreted data processing, and other methods);

trustworthy AI technologies that meet safety standards, are developed with due regard for the principles of objectivity, non-discrimination, ethics, and rule out any possibility of harm to human beings and violation of their fundamental rights and freedoms, or damage to the interests of society and the state.

The Strategy notes that artificial intelligence is one of the most important technologies available to man today: thanks to AI, the world economy is growing already, innovation in all fields of science is accelerating, the quality of life of the population, availability and quality of medical care, quality of education, labour productivity and quality of recreation are improving. AI technologies are an area of international competition. Technological leadership in AI can enable states to attain meaningful results in key areas of social and economic development. In the late 2010s governments in developed countries began to focus on the development of AI technologies. To date, more than 60 countries have developed and approved their own national strategies for the development of artificial intelligence.

As the new version of the Strategy states, between 2022 and 2023, the world saw a new leap in the development of AI technologies owing to the improvement of large generative models in the fields of language, images (including video images) and sound. Large fundamental models are already capable of writing software codes according to technical tasks, composing poems on a given topic, giving precise and clear answers to test questions of various levels of complexity, including those from educational programmes. AI models create images on any topic in a matter of seconds based on a given text description or sketch. This poses a threat of the dissemination of prohibited information, copyright infringement and the generation of erroneous information.

AI will significantly impact the global economic growth. According to expert estimates, further development of large generative models can bring about a surge in labour productivity, which will lead to an annual increase in the global GDP by 1–2 percent and increase the remuneration of specialists in all sectors of the economy by increasing the volume of production (goods, works, services) and improving its quality.

At the same time, according to the National Security Strategy of the Russian Federation, in order to ensure and protect the country's national interests from external and internal threats, including unfriendly actions of foreign states, it is necessary to increase the efficiency of the use of the achievements and competitive advantages of the Russian Federation with account of long-term global trends. In order to solve the tasks set in the sphere of national security, AI is used as a tool to ensure information security based on the application of advanced technologies, including AI and quantum computing technologies as a means of upgrading industrial enterprises and infrastructure, digitalisation to improve labour productivity, and to boost the development of Russia's scientific and technological base, nanotechnology, robotics, medical, biological, genetic engineering, information and communication, big data processing, energy, laser, additive, creation of new materials, cognitive, and nature-like technologies.

At the same time, the Russian Federation Concept for the Development of Regulation of Relations in the Field of Artificial Intelligence and Robotics of 2020 (hereinafter the Concept[20]) developed in order to determine the main approaches to the transformation of the regulatory system in the Russian Federation so as to create conditions for creation and application of such technologies in various spheres of the economy while respecting the rights of citizens and ensuring the safety of individuals, society and the state,

[20] Collection of Laws of the Russian Federation. 31 August 2020. No. 35. P. 5593.

proceeds from the premise that the development of AI and robotics requires the creation of a regulatory environment comfortable for safe development and implementation of these technologies, based on a balance of interests of the individual, society, the state, companies developing AI and robotics systems, as well as consumers of their goods, works and services.

The Concept refers in Para 5 to technologies based on the use of AI: computer vision; natural language processing; speech recognition and synthesis; intelligent decision-making support; promising methods of AI. Promising AI methods include: ability to solve various problems independently, automatic design of physical objects, automatic machine learning, algorithms for solving problems based on data with partial partitioning and (or) insignificant amounts of data, information processing based on new types of computing systems, interpreted data processing, and other methods).

The Concept notes that the growing degree of AI and robotics systems autonomy, decreasing human control over the process of their application, and a not fully transparent decision-making process create a public demand for regulatory restrictions on the use of AI and robotics systems. At present, there are no unified approaches to regulating artificial intelligence and robotics technologies worldwide. This is due to the existence of a number of problems that have no clear solution.

The Concept outlines the Russian legal model for AI regulation in accordance with the National Strategy for the Development of Artificial Intelligence for the period up to 2030. It stipulates the following main areas for the creation of a comprehensive system for regulation of public relations arising in connection with the development and implementation of AI technologies:

ensuring a favourable legal environment (including the establishment of a pilot legal regime) for access to predominantly anonymised data, including data collected by public authorities and health care providers;

ensuring special conditions (regimes) for access to data, including personal data, for the purposes of academic research, creation of AI technologies and development of technological solutions based thereon;

creating legal conditions and establishing procedures for simplified testing and implementation of technological solutions developed on the basis of AI, as well as delegating to AI-powered information systems the ability to make certain decisions (except for decisions that may infringe upon the rights and legitimate interests of citizens). This includes the performance of state functions by state bodies (except for functions aimed at ensuring the security of the population and the state);

eliminating administrative barriers to the export of civilian products (works, services) created with AI;

creating unified systems of standardisation and conformity assessment of solutions developed on the AI basis, developing Russian Federation's international cooperation on standardisation issues and ensuring the possibility of certification of products (works, services) created on the AI basis;

encouraging investments by improving mechanisms for joint participation of investors and the state in projects related to the development of AI technologies, and providing targeted financial support to entities engaged in the development and implementation of AI technologies (provided that the introduction of such technologies will result in significant positive effects for the Russian economy);

developing ethical rules for human interaction with artificial intelligence.

The above areas must become the main landmarks in establishing a comprehensive system for regulation of public relations arising in connection with the development and implementation of AI technologies and robotics.

The Concept stipulates that, given the economic and social significance of AI and robotics technologies in various fields, their development and operation should not be confined to regulatory measures (except in cases involving a high risk of harm to human life and health). It is also unacceptable to use AI and robotics that pose a clear threat to the defence of the country and the state security.

For developing particular regulatory solutions it is necessary to use a risk-based approach based on an assessment of the amount of potential harm to these values, taking into account the likelihood of risk compared to the potential positive effect of the introduction of AI and robotics technologies, and the need to take measures to minimise the relevant risks.

The mere fact that AI systems and robotics are used should not be a basis for regulatory restrictions.

It is necessary to support the development of regulation developed and enforced by market participants (self-regulation), including the adoption and use of documents of the national standardisation system, ethical codes (sets of ethical rules) and other documents of self-regulatory organisations, as well as other instruments.

In view of the fundamental complexity of this sphere of legal relations, the development of a regulatory regime for artificial intelligence and robotics technologies requires the active involvement of representatives of corporate developers of AI and robotics systems and R&D organisations in the process of expert elaboration of the relevant laws and regulations.

In the future, some norms of law may also need to be clarified in order to provide normative legal regulation of new types of legal relations.

2.2. Legal AI Regulating in Europe

In March 2024, the European Parliament has passed a law regulating artificial intelligence that will come into force in June 2024 (EU Artificial Intelligence Act).[21] It applies to AI technology providers and users of AI-based technologies in the private and public sectors. The purpose of the Act is to improve the functioning of the internal market and the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence, while ensuring a high level of protection of health, safety, fundamental rights, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation.

As with other EU data-related legislation, the Act also applies extraterritorially to companies and organisations outside the EU. The AI Act applies to:

providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the EU, irrespective of their location or establishment;

deployers of AI systems that have their place of establishment or are located within the EU;

providers and deployers of AI systems that have their place of establishment or are located outside the EU, where the output produced by the AI system is used in the EU;

importers and distributors of AI systems;

product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark;

authorized representatives of providers which are not established in the EU.

[21] Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (accessed: 01.04.2024)

The AI Act establishes a legal framework for the application of AI based on the assessment of risks (as the combination of the probability of the occurrence of harm and the severity of that harm) associated with the use and placing on the market the following categories of artificial intelligence systems: prohibited artificial intelligence practices, high-risk artificial intelligence systems, systems with transparency requirements, and general purpose artificial intelligence models.

In the Act, an AI system' means a "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments» (Article 3). In addition to many other important definitions, the Law also contains a definition of "deep fake" that means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful".

According to the definition, a key characteristic that distinguishes "AI systems" from traditional software is that an AI system derives conclusions for the output from the input ("infers, from the input it receives, how to generate outputs"). This is intended to emphasise the ability of AI systems to derive models and/or algorithms from input data. By contrast, the EU wanted to exclude systems that are based on rules that are defined exclusively by natural persons in order to carry out automatic processes from the scope of the AI Act. By definition, the capabilities of AI systems should go beyond basic data processing operations and be understood more as learning, reasoning or modelling.

The definition in the AI Act also assumes that AI systems are "designed to operate with varying levels of autonomy". Accordingly, there must be a certain degree of independence of the system's actions from humans. In other words, the system must be able to operate without human intervention.

The characteristic of "adaptiveness" is intended to express the ability of an AI system to (continue to) learn itself and thus constantly change.

Prohibited Artificial Intelligence Practices are:

the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm;

the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;

the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following:

detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;

detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity;

the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless and in so far as such use is strictly necessary for one of the following objectives;

AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons;

the use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement.

The Act identifies high-risk AI systems that pose a potentially high risk to human rights and freedoms and differentiates them into two high-risk AI groups. The first group includes AI systems that pose a risk when the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation (e.g., Regulation (EU) 2017/745 on medical devices or Regulation (EU) No 167/2013 on agricultural and forestry vehicles); the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation.

It may apply to AI systems used in, among other things, cars, toys, lifts, equipment and safety components for use in medical devices and *in vitro* diagnostic medical devices, products related to civil aviation, marine equipment, products related to railway systems, and various types of vehicles.

General-purpose artificial intelligence models represent an independent type of AI systems. According to the EU AI Law, a general-purpose AI model is trained on large amounts of data using scalable self-monitoring that demonstrates significant generality, is capable of competently performing a wide range of individual tasks, and can be integrated into a variety of downstream systems or applications, including serving as the basis for general-purpose AI systems.

In addition, the AI Act also introduces a category of general purpose AI models with a systemic risk for more advanced general purpose AI models to be defined by the European Commission. General-purpose AI (GPAI) models with a systemic risk will be subject to additional obligations for model evaluation and testing, risk mitigation, security, and incident reporting.

GPAI models are subject to a range of obligations fostering technological deployment and ensuring adequate safeguards, including the provision of detailed technical documentation to the competent authorities, the provision of information to downstream providers, the implementation of policies to protect copyright and the publication of a summary of the content used for training the GPAI model. Providers that release GPAI models under a free and open-source licence are subject to certain exemptions of these obligations.

GPAI models are considered to have a systemic risk if they have high impact capabilities, e.g., if they have great computing power (currently when the computation used for its training is greater than 10^25 FLOPS and subject to future amendments by the Commission). Furthermore, GPAI models can be classified as having systemic risk in case of a decision of the

Commission (either ex officio or following a qualified alert from a scientific panel of independent experts). The provider of such GPAI model needs to:

perform model evaluation in accordance with standardised protocols;
conduct systemic risk assessments and mitigate systemic risks;
report incidents to authorities; and
ensure adequate cybersecurity protection, including the physical infrastructure of the model.

The EU AI Law follows a risk-based approach taking into account the risks of AI to natural persons. The AI Act therefore distinguishes between prohibited AI practices, high-risk AI systems, AI systems with transparency risk and GPAI models with/without a systemic risk. Before placing an AI system on the market, putting it into service, deploying, distributing, importing or otherwise using it, it must be carefully ruled out that such system does not entail an "unacceptable risk" within the meaning of the AI Act.

2.3. US Legal AI Regulating

On 30 October 2023, US President Biden has issued a new Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.[22] It sets new standards for AI safety, provides a set of measures and directs government agencies to implement specific policies to address areas of concern in national security, data protection, labour relations and social health. The Order stipulates an obligation for companies developing the most powerful AI systems to report the results of AI safety tests and other important information to the US. government. Under the Defense Production Act, the Order requires developers of AI foundation models that potentially pose a serious threat to national security, national economic security, or national public health to notify the federal government when training an AI model about the results of all pen-tests (red-team) to assess the cyber security of the AI model before companies make those results public.

The Order includes more than a hundred policy directives related to AI security to more than twenty federal agencies, tasking them with policies to address problem areas such as national security, data protection, workplace bias, and public health. It also imposes obligations on private companies developing powerful AI systems that could pose a threat to national security

[22] Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ (accessed: 11.04.2024)

or public health, requiring them to share safety test results and methods and other sensitive information with the U.S. government. Most of the directives issued by President Biden under this Order must be implemented within 2024.

In March 2023, the President has approved a new version of the National Cybersecurity Strategy[23] establishing protected US critical infrastructure has become one of the national security priorities. The initiative seeks to shift some of the burden of cyber security risk mitigation from end users and critical infrastructure operators to private sector enterprises that are best positioned to make meaningful progress on security and resilience. The Strategy also highlights the need to change incentives in favour of long-term private sector investment. The strategy is based on five pillars: protecting critical infrastructure; identification and destruction of threat actors; establishing market mechanisms to improve security and resilience; investing in a sustainable future; and building international partnerships to achieve common goals. Each pillar contains specific strategic objectives that build on previous programmes and guide the implementation efforts of government and private sector entities.

2.4. China AI Legal Regulating

China has achieved significant success in its efforts to become a technology superpower over the past few years, making continuous efforts to establish itself as the world's leading IP producer.

The country has transformed from a low-wage economy to a high-tech country. In fact, according to the World Intellectual Property Organisation (WIPO), China accounted for 47% of all patent applications worldwide in 2023. On 13 July 2023, the Chinese government has published regulations on generative artificial intelligence, Interim Measures for the Administration of Generative Artificial Intelligence Services (hereinafter Interim GAI Measures; Measures)[24] came into force on 15 August 2023. The Measures aim to regulate generative AI that is primarily intended for content creation. They are the latest addition to the emerging system of AI regulation in China, which already includes a number of AI-specific and local laws.

[23] Available at: https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf (accessed: 01.04.2024)

[24] Available at: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm (accessed: 11.04.2024)

The Chinese government has been supporting its AI industry on a national level since the beginning. In its 13th Five-Year Plan (2016-2020), China identifies AI as key to achieving economic growth goals. In 2017, the Chinese government presented its vision for AI development in its Next-Generation AI Development Plan. The Plan presents Beijing's comprehensive strategy to focus AI on the country's socio-economic development efforts (AI industry), which will make China a global AI leader by the year 2030.

GAI Interim Measures differ from other laws in that they specifically regulate the use of generative AI defined as "models and related technologies that have the ability to generate content such as text, images, audio and video," so as to provide content generation services to the Chinese public. Compared to the provisions of Deep Synthesis, the generative AI covered by GAI Interim Measures encompasses more than algorithm-based generative technologies, and includes rules-based models and systems.

The Measures apply to generative AI service providers defined as legal entities and individuals that use generative AI to provide generative AI services, including the provision of such services through application programming interfaces (APIs). Also, GAI Interim Measures cover the provision of generative AI services to the public indirectly through business arrangements. On the other hand, institutions that develop and apply generative AI technology but do not provide generative AI services to the public do not fall under this regulation.

In addition, the GAI Interim Measures establish an extraterritorial scope by specifying that they apply to the provision of services to the public in the PRC mainland, potentially extending their application to individuals and organisations outside of China that provide generative artificial intelligence services to individuals in the PRC. This nuance is complemented by another provision stating that, if generative AI vendors outside the PRC fail to comply with the Measures and other laws, this will entail notification to the relevant agencies to take technical measures and other necessary measures to deal with the perpetrators.

Conclusion

Exponential improvements in artificial intelligence and other advanced technologies in recent years have led to a surge in interest (academic, commercial, military, etc.) and financial investment in artificial intelligence.

It is obvious that the rapid progress in the development and practical application of AI technologies is driven by their expected potential to increase productivity, encourage innovation and entrepreneurship, provide solutions to global problems, including social problems such as improving healthcare and helping to solve the climate crisis, as well as to achieve the Sustainable Development Goals. At the same time, this process also generates new threats and challenges for the human civilisation, which is a special factor that must be taken into account when promoting the development of artificial intelligence.

References

1. Aggarwal R., Singh H. (2024) Overcoming Limitations of Ai Agents: Integrating Tacit Knowledge Through Inferred Latent Themes. Available at: SSRN: https://ssrn.com/abstract=4843878 (accessed: 10.04.2024)

2. Amelin R.V., Channov S.E. (2023) *Evolution of Law under Influence of Digital Technologies.* Moscow: Norma, 280 p. (in Russ.)

3. Antonova N.V. et al. (2019) *The Legal Concept of Robotization.* Moscow: Prospekt, 240 pp. (in Russ.)

4. Barfield W., Pagallo U. (2020) Advanced Introduction to Law and Artificial Intelligence. *Law in Context*, vol. 37, no.1. Available at: https://books.google.ru/books?id=7MgBEAAAQBAJ&printsec=frontcover&hl=ru#v=onepage&q&f=false (accessed: 11.04.2024)

5. Bommasani R. et al. (2021) Opportunities and Risks of Foundation Models. Available at: https://doi.org/10.48550/arXiv.2108.07258 (accessed: 11.04.2024)

6. Brundage M. et al. (2018) The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Available at: https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf (accessed: 01.04.2024)

7. Cowgill B., Tucker C. (2020) Algorithmic Fairness and Economics. Columbia Business School Research Paper. Available at: SSRN: https://ssrn.com/abstract=3361280 (accessed: 11.04.2024)

8. Eloundou T., Manning S. et al. (2023) GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. Available via license: Creative Commons Attribution-ShareAlike 4.0 International (accessed: 11.04.2024)

9. Feuerriegel S. et al. (2023) Generative AI. *Business & Information Systems Engineering,* vol. 66, no. 2, pp. 111–126.

10. Hill M., Hackworth A. (2023) Copyright in the Age of AI. Available at: https://www.charlesrussellspeechlys.com/en/insights/quick-reads/102j7d1-copyright-in-the-age-of-ai/#page=1 (accessed: 11.04.2024)

11. Ivliev G.P., Egorova M.A. (2022) On Legal Status of Artificial Intelligence and Products Created by Artificial Intelligence Systems. *Zhurnal rossiyskogo prava*=Journal of Russian Law, no. 6, p. 32–46 (in Russ.)

12. Kharitonova Y. S., Savina V. S. (2020) Artificial Intelligence Technology and Law: Challenges of Modernity. *Vestnik Permskogo universiteta*=Bulletin of Perm University, no. 49, pp. 524–549 (in Russ.)

13. Kirsanova E.E. (2023) Review of the Main Theories of Determining the Legal Regime of Objects Created by Artificial Intelligence. *Zakon*=Law, no. 9, pp. 36–46 (in Russ.)

14. Morhat P.M. (2018) Legal Personality of an Electronic Person. *Pravovye issledovania*=Legal Studies, no. 4, pp. 1–8 (in Russ.)

15. Naumov V.B. et al. (2021) Legal Aspects of Using Artificial Intelligence. Available at: https://www.hse.ru/mirror/pubs/share/480106412.pdf (accessed: 11.04.2024) (in Russ.)

16. Novelli C., Casolari F. et al. (2024) Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cyber security. Available at: SSRN: https://ssrn.com/abstract=4694565 (accessed: 10.04.2024)

17. Pathirannehelage H. et al. (2022) Design Principles for AI-augmented Decision-Making: an action design study. Available at: SSRN: https://ssrn.com/abstract=4071519 (accessed: 11.04.2024)

18. Rozhkova M.A. (2021) Will Artificial Intelligence Become an Independent Subject of Law? *Khozyaistvo i pravo*=Economy and Law, no. 6, pp. 14–22 (in Russ.)

19. Russell S., Norvig P. (2009) Artificial Intelligence: A Modern Approach. Available at: http://aima.cs.berkeley.edu/ (accessed: 11.04.2024)

20. Voller K. (2024) Generative AI — not so Great at Generating European patents. Available at: https://www.gje.com/resources/generative-ai-not-so-great-at-generating-european-patents/ (accessed: 11.04.2024)

21. Xenidis R., Senden L. (2020) EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping Challenges of Algorithmic Discrimination. In: U. Bernitz et al. (eds.) General Principles of EU Law and the EU Digital Order. Available at: SSRN: https://ssrn.com/abstract=3529524 (accessed: 10.04.2024)

22. Zewe A. (2023) Explained: Generative AI. Massachusetts Institute of Technology News. Available at: https://news.mit.edu/2023/explained-generative-ai-1109 (accessed: 11.04.2024)

Information about the author:

A.A. Kartskhiya — Doctor of Sciences (Law), Professor.

The article was submitted to editorial office 23.06.2024; approved after reviewing 28.06.2024; accepted for publication 28.06.2024.