



Governing AI Foundation Models: Regulatory Approaches and Global Perspectives

Osman A. Izhaev, Dmitry L. Kuteynikov

University of Tyumen, Tyumen, Russian Federation



Corresponding Author — Osman A. Izhaev

© O.A. Izhaev, D.L. Kuteynikov, 2026

Abstract: This article analyses the approaches of the European Union, the United States, and the Russian Federation to the legal regulation of the development and deployment of AI foundation models. Drawing on regulatory approaches reflected in the EU Artificial Intelligence Act, U.S. Executive Order 14110, the NIST AI Risk Management Framework, and relevant legal acts of the Russian Federation, it shows that legal regulation is primarily based on the classification of models according to their capabilities and associated risks.

Based on the findings derived from an analysis of the legal and ethical regulations in the Russian Federation, the study concludes that there are no specific rules dedicated to the development and deployment of foundation models.

To address this regulatory gap, the article formulates recommendations for policymakers in the Russian Federation. In particular, it proposes that foundation models be classified into two groups based on the computational resources required for their training. The first group includes models that do not pose significant risks, while the second group includes the most advanced models with potentially hazardous capabilities.

The proposed approach may contribute to the safe and secure development and deployment of foundation models in the Russian Federation by enabling a differentiated regulatory framework that takes into account their potential impact on human rights.

Keywords: artificial intelligence; AI systems; foundation models; general-purpose AI models; frontier models; large language models; training compute; AI governance; transparency; EU AI Act

Acknowledgments: The study was supported by a grant from the Russian Science Foundation No. 25-28-01582. Available at: <https://rscf.ru/project/25-28-01582/>.

Cite as: Izhaev, O.A. and Kuteynikov, D.L., (2026). Governing AI Foundation Models: Regulatory Approaches and Global Perspectives. *Kutafin Law Review*, 13(2), pp. 269–295, doi: 10.17803/2713-0533.2026.2.36.269-295

Contents

I. Introduction	270
II. Foundation Models: Key Concepts and Terminology in Legal Regulation	271
III. Understanding the Regulatory Approaches to the Development and Deployment of Foundation Models in the European Union, the United States and Russia	278
IV. Conclusion	290
References	291

I. Introduction

In recent years, the development of artificial intelligence (AI) has shifted towards increasingly general-purpose systems. Alongside specialised AI models designed for narrow tasks, a new class of large, adaptable systems has emerged. In 2021, researchers from Stanford University’s Centre for Research on Foundation Models (CRFM) introduced the term “foundation models” to describe models trained on broad data (typically using self-supervision at scale) that can be adapted, such as through fine-tuning, to a wide range of downstream applications (Bommasani, 2021).

Foundation models possess the following essential characteristics: a) emergent capabilities — the ability to exhibit new, unpredictable capabilities; and b) homogenization — the possibility of using an AI model as the foundation for developing different AI products. The aforementioned issues have led to several legal challenges, in addition to those already existing in the field of AI regulation and governance.

Firstly, many foundation models are general-purpose software that provides broad opportunities for their application in various fields. Sec-

only, developers (or upstream providers) of foundation models may avoid legal liability because they may not put their models on the market themselves. Instead, developers can provide access to their models to downstream providers on various terms, who then create AI systems based on the models and put them on the market. Thirdly, this type of relationship within the AI value chain creates an inequality of rights between developers and downstream providers, as it shifts legal obligations exclusively to the latter (Brown, 2023). This also limits the interests of end users of products, as they cannot directly influence developers to protect their rights. Fourth, only developers have access to the infrastructure, source code, parameters, and training data of the models, which leaves other downstream providers vulnerable to unilateral model modifications by the developer or in the event of errors or emergencies. Finally, fifthly, all foundation models have diverse characteristics, and the most advanced ones are under-researched, potentially posing systemic risks to human rights and society as a whole.

The response of legislators to these issues has been uneven (Chun et al., 2024): some countries are adopting a wait-and-see approach, while others have already formulated their own strategies to regulate the development and deployment of AI foundation models (Aloncho et al., 2025, p. 914). Thus, it is timely to conduct research in order to analyse different regulatory approaches and identify their shortcomings.

This study employs doctrinal legal analysis and comparative legal analysis of regulatory frameworks in the European Union, the United States, and the Russian Federation, and develops a normative proposal concerning the classification of foundation models based on technical criteria.

II. Foundation Models: Key Concepts and Terminology in Legal Regulation

The research in the field of the legal regulation of the development and deployment of foundation models is complicated by the fact that the terminology is unclear (Winter and Bullock, 2024). Alongside foundation models, related terms such as “large language models,” “gen-

eral-purpose AI models,” “generative AI,” “dual-use foundation models,” and “frontier models” are also used (Fernández-Llorca et al., 2025, pp. 875–888). In this paper, the term “foundation model” is used as the primary analytical category. However, when analysing legal and policy documents, the original terminology employed in those documents (e.g., “general-purpose AI models” or “frontier models”) is retained to ensure the accuracy of legal interpretation. It is assumed that these categories share essential characteristics inherent in foundation models. This approach ensures terminological clarity and internal logical consistency, which is necessary to achieve the objectives of the study.

Accurate definitions in legal acts are crucial, especially in interdisciplinary fields such as AI (Koivisto et al., 2024, pp. 412–432). Vague or overlapping definitions can create legal uncertainty and complicate law enforcement. In this regard, it seems appropriate to study the approaches in different jurisdictions. The European Union, the United States of America and the Russian Federation were selected as the objects of analysis. This choice was motivated, on the one hand, by the desire to cover approaches from different regions of the world, and on the other hand, by the need to consider the limited scope of the study.

In the European Union, approaches to terminology have evolved significantly during the drafting of the EU AI Act (hereinafter referred to as the AIA). In the AIA draft presented by the European Commission in April 2021, the regulation only applied to operators of AI systems (Hacker et al., 2025). Later, in December 2022, references to general-purpose AI (GPAI) systems appeared in the EU Council’s General Approach, and the text approved by the European Parliament already provided for the regulation of both foundation models and generative AI. The discussion of this section of the AIA draft sparked some of the most heated debates in the trilogue (Casonato and Olivato, 2024, pp. 87–112). As a result of the negotiation process, a compromise was reached, consisting of separating general-purpose AI models from those with systemic risk into distinct groups.

Thus, according to Para. 63 Art. 3 of the AIA, “a GPAI model is an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of

distinct tasks... and that can be integrated into a variety of downstream systems or applications.”¹ The AIA does not provide a literal definition of general-purpose AI models with systemic risk, but it follows from its provisions that these are general-purpose AI models that are capable of performing a wide range of distinct tasks and can be integrated into a variety of downstream systems or applications. The AIA does not provide a literal definition of AI models with systemic risk, but it follows from the entirety of its provisions that these are general-purpose AI models that pose high-impact capabilities. In turn, high-impact capabilities mean capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models.

Based on the above definitions, it is impossible to determine in practice which group a particular model should be classified under. In this regard, European legislators in the AIA have established technical criteria for identifying GPAI models with systemic risk (GPAISR). According to Art. 51 of the AIA, a model should be recognized as a GPAISR model in the following two cases.

1) The model demonstrates high-impact capabilities, which are determined based on the application of appropriate technical metrics, including relevant criteria and benchmarks. Part 2 Art. 51 of the AIA clarifies the meaning of the previous provision and states that if the total amount of computations used to train the model, expressed in floating point operations (FLOP), is greater than 10^{25} , it is presumed to have high impact capabilities.

2) The model is recognized as posing a systemic risk by decision of the European Commission, either at its discretion or following notification from the Scientific Panel of Independent Experts.² Annex XIII to the AIA explains that the Commission may take such a decision based on the criteria such as the number of model parameters, the quality and size of the dataset, the amount of computing power spent on training the model, and many others.

¹ Artificial Intelligence Act. Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> [Accessed 15.08.2025].

² The Scientific Panel of Independent Experts was established in accordance with Art. 68 of the AIA and Implementing Regulation 2025/454 of 7 March 2025 of the European Commission. Available at: https://eur-lex.europa.eu/eli/reg_impl/2025/454/oj/eng [Accessed 15.08.2025].

However, the AIA does not contain similar criteria for defining a regular GPAI model (without systemic risk). This gap is filled by the European Commission's Guidelines on the scope of the obligations for general-purpose AI models.³ According to Para. 17 of the Guidelines, a GPAI model is one where the computational power spent on training it exceeds 10^{23} FLOP and it is capable of generating natural language (whether in the form of a text or an audio, text-to-image or text-to-video).

Thus, in the EU, GPAI models are divided into two separate groups: those with systemic risk (GPAISR models) and those without it (GPAI models). In both cases, the training compute is used as a proxy for identifying the capabilities and risks of the models.

The United States takes a different approach. In particular, in October 2023, the U.S. President issued Executive Order 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.⁴ The next U.S. President repealed this order, but it contains a number of essential concepts that reveal the U.S. approach to terminology that are still relevant today.

First, the decree formulated the definition of a “dual-use foundation model” means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by:

a) substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons;

b) enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks; or

³ Guidelines on the scope of obligations for providers of general-purpose AI models under the AI Act. Available at: <https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act> [Accessed 15.08.2025].

⁴ Executive Order 14110 of 30 October 2023. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Available at: <https://www.federalregister.gov/d/2023-24283> [Accessed 15.08.2025].

c) permitting the evasion of human control or oversight through means of deception or obfuscation.

Secondly, the decree also effectively referred to dual-use models as: a) any model that has been developed using more than 10^{26} FLOP of training compute; b) models that were trained primarily using biological data sequences and whose training required more than 10^{23} FLOP of training compute.

The approach implemented in California State Law SB-53 Transparency in Frontier Artificial Intelligence Act (TFAIA)⁵ deserves special attention. In particular, foundation models in the act are defined as models trained on large datasets, designed for the generality of their output, and adaptable to a wide range of distinct tasks. At the same time, the act formulated a definition of a frontier model, which refers to a foundation model trained using more than 10^{26} FLOP.

Thus, both in the European Union and in the United States, legal definitions of foundation models are based on dividing them into two separate groups. The first group includes models that possess the essential characteristics of foundation models — the ability to perform a wide range of tasks and serve as a basis for creating other AI products. At the same time, it is believed that they do not pose significant threats to society and the state. The second group, on the contrary, includes foundation models that generate high or systemic risks, which requires increased attention from regulators. Their approaches are based on using the amount of computing power spent on training the model as a proxy for its capabilities and risks.

Another approach to terminology can be found in the legal acts of the Russian Federation (Ivanova, 2022, p. 204). For instance, some key terms are stipulated in the Decree of the President of the Russian Federation No. 490 dated 10 October 2019, which approved the “National Strategy for the Development of Artificial Intelligence for the Period until 2030” (hereinafter referred to as the Russian AI Strategy).⁶ The

⁵ SB-53 Artificial intelligence models: large developers. Available at: https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=20250260SB53 [Accessed 15.08.2025].

⁶ Decree of the President of the Russian Federation No. 490 of 10 October 2019 “On the Development of Artificial Intelligence in the Russian Federation” (together

decree provides definitions for a number of terms: “AI model,”⁷ “large generative model,”⁸ and “large foundation model.”⁹ A common feature of these definitions is that they are human-centric. An AI model, and therefore a foundation model, is an entity that solves problems at a level comparable to human intellectual work.

The anthropomorphisation of AI is a common phenomenon in the science fiction literature, but not in modern regulatory contexts. It is debatable whether the definition of AI in the Russian Federation’s Strategy “fully corresponds to the generally accepted understanding of AI as the implementation of thought processes inherent in the human brain by a computing device” (Polyakova and Smirnov, 2024, pp. 89–106). A correct understanding of the foundation model is required, at the very least, to limit the scope of regulation. If we compare AI to a human being, the latter becomes the main criterion for AI, which is a gross simplification (Placani, 2024, pp. 691–698). Under current conditions, many devices can imitate human intellectual activity, from calculators to programmed chatbots. If we equate any model that performs tasks that a human could perform, then the subject of regulation in the field of AI will be extremely broad (Finocchiaro, 2024, pp. 1961–1968).

In addition, the Russian AI Strategy defines a “large foundation model” as one that contains at least 1.5 billion parameters. However, the threshold of 1.5 billion parameters allows models with relatively modest performance to be included in this group. For example, GPT-2, which was released back in 2019, contains 1.5 billion parameters, while

with the “National Strategy for the Development of Artificial Intelligence for the Period up to 2030”) // Official Internet Portal of Legal Information: <http://www.pravo.gov.ru> [Accessed 15.08.2025].

⁷ An AI model is a software (or its component) designed to perform intellectual tasks at a level comparable to or exceeding the results of human intellectual activity, using algorithms and datasets to derive patterns, make decisions or predict outcomes.

⁸ Large generative models are AI models capable of interpreting (providing information based on queries, e.g., about objects in an image or analysed text) and creating multimodal data (text, images, video material, etc.) at a level comparable to or exceeding the results of human intellectual activity.

⁹ Large foundation models are AI models that form the basis for the development and modification of various types of software. They are trained to recognize certain types of patterns, contain at least 1 billion parameters, and are used to perform a wide range of tasks.

one of the many modern foundation models, GPT-oss-120 b (an open-weight model similar to GPT-3.5), has 120 billion parameters.¹⁰ Thus, large foundation models in the sense of the Russian AI Strategy are not analogous to the most advanced foundation models and, therefore, pose a danger.

The total number of parameters is indeed related to model capabilities and associated risks, and this criterion is also easy to measure and simple to apply. However, it also has significant drawbacks. For example, some model architectures, such as Mixture of Experts (MoE), can significantly change the number of model parameters without necessarily increasing their capabilities and risks (Villalobos et al., 2022). At the same time, the number of model parameters can be manipulated through post-training techniques such as model pruning without necessarily decreasing the model's capabilities and risks (Cheng et al., 2023).

It follows from the above that in the Russian Federation, foundation models are not classified into separate groups depending on their technical characteristics. The chosen approach to defining foundation models by specifying the number of parameters has significant drawbacks, both in terms of the threshold value of 1.5 billion and the limited effectiveness of such a metric for assessing the real capabilities of models and the associated risks.

Among the available proxies for model capabilities and risks, the computational power spent on training the model represents a practical and measurable technical criterion (Raman et al., 2025). This criterion has a number of significant advantages: a) it is difficult to circumvent reducing the amount of computation required to train the model limits its capabilities, thereby reducing potential risks; b) it can be assessed in advance, which simplifies preventive regulation; c) it is subject to external verification: such assessments can be performed by third party organisations, including providers, without disclosing confidential information about the model (Heim and Koessler, 2024). In addition, training compute has characteristics that are important for regulating the field of AI, such as detectability, excludability, measurability, and a narrow circle of entities dominating the chip supply chain (Sastry et al., 2024).

¹⁰ Data on AI Models. Available at: <https://epoch.ai/data/ai-models> [Accessed 15.08.2025].

At the same time, science also highlights several limitations of using training compute as a proxy (Caputo, 2025). In particular, advances in algorithms and hardware efficiency may render existing thresholds obsolete, and model capabilities may also depend on factors such as architecture, dataset characteristics, and post-training modifications (Price, 2025). Therefore, training compute should not be regarded as a definitive or exclusive indicator of model risk. Rather, it represents a practical and measurable proxy suitable for the purposes of preliminary classification and the development of differentiated regulatory approaches.

Taking into account the current legal framework and the absence of specific classification criteria for foundation models in the Russian Federation, it appears appropriate to consider the use of training compute as a technical criterion for their preliminary classification. We believe that a more reasonable approach would be to divide foundation models into two separate groups based on the computing power required for their training. The first group would include foundation models that do not pose significant risks to society and the State, while the second group would comprise the most advanced models with potentially dangerous capabilities. For the purposes of this study, significant risks refer to the potential of foundation models to have a substantial impact on protected legal interests, including fundamental rights, information security, public safety, and other societal interests, due to their scale, capabilities, and potential deployment contexts.

This approach would enable the establishment of a differentiated framework for regulating the development and deployment of foundation models, based on an assessment of their predicted impact on human rights and societal interests.

III. Understanding the Regulatory Approaches to the Development and Deployment of Foundation Models in the European Union, the United States and Russia

The European Parliament proposed establishing several stringent requirements for providers of foundation models. For example, the development of a risk management system, the use of high-quality and

transparent datasets, as well as measures to ensure the reliability, security, interpretability and adjustability of the model at all life cycle stages, which had to be confirmed by an independent assessment.¹¹ In turn, the European Council *de facto* proposed equating foundation models with high-risk systems, thereby imposing increased obligations on their providers.¹² As noted in the scientific literature, this would turn the EU's risk-based approach on its head (Hacker, 2023).

In a joint non-paper, Italy, France and Germany proposed establishing self-regulation through codes of conduct in November 2023.¹³ The primary measure of this approach was the publication of model cards by the relevant suppliers, which would reflect data on the model, including information on its capabilities and associated risks. At the same time, failure to comply with this recommendation would not entail any sanctions (Datta, 2025, pp. 119–138).

It is believed that the agreement between European Union institutions on general-purpose AI models was the result of a significant political compromise (Gstrein et al., 2024). This compromise is reflected in the fact that providers of general-purpose AI models were not subject to requirements similar to those imposed on providers of high-risk AI systems. The requirements proposed by the European Parliament were also not included in the final version of the act.

In the current version of the Act, the regulatory approach is based on the distinction between two groups of models: GPAI models and GPAISR models. The AIA sets specific requirements for each model provider group.

In accordance with Art. 51(1) of the AIA, providers of GPAI models are subject to the following obligations. First, providers are required

¹¹ Amendments adopted by the European Parliament on 14 June 2023. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf [Accessed 15.08.2025].

¹² Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights. Available at: <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/> [Accessed 15.08.2025].

¹³ Joint Non-paper by Italy, France and Germany. Available at: https://logistic-natives.com/wp-content/uploads/2023/11/9aaf68af_2128_4aa6_b8d6_a61212e5fd46_France_German_231122_140916.pdf [Accessed 15.08.2025].

to maintain technical documentation on the development and testing of the model, which must be made available to regulators upon request. Second, they must make information about the model available to downstream providers. This must include information about the model that enables providers of AI systems to have a clear understanding of the capabilities and limitations of the general-purpose AI model. A specific list of the minimum information required is set out in Annex XII. Third, it is necessary to ensure compliance with the EU copyright legislation, including Art. 4(3) of Directive (EU) 2019/790 (machine-readable reservations). Fourth, providers must make publicly available a sufficiently detailed summary about the content on which the model has been trained.

The obligations outlined above also apply to GPAISR model providers. However, Art. 55 of the AIA imposes additional obligations on them. Specifically, they must perform model evaluation, assess and mitigate systemic risks; promptly report serious incidents and possible measures to eliminate them to the competent state authority, and ensure an adequate level of cybersecurity and the corresponding physical infrastructure.

Thus, in the EU, GPAI model providers are required to comply with different sets of requirements depending on whether they pose systemic risks. Regarding regular GPAI model providers, the primary requirement is transparency. In particular, intended purposes for which a model can be used, as well as acceptable use policies, model release date, distribution methods, and a range of other information must be specified (Annexe XII).

It should be noted that detailed information for complying with the requirements for providers of both types of GPAI models is specified in the voluntary Code of Practice,¹⁴ which was developed in collaboration with industry representatives for the period until harmonized standards were developed. In particular, it contains a model documentation form that specifies the procedure for disclosing relevant information about the model and the requirements for the content of a copyright policy.

¹⁴ General-Purpose AI Code of Practice. Available at: <https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice> [Accessed 15.08.2025].

Regarding GPAISR model providers, the Code stipulates that the signatories must develop a safety and security framework and model reports.

The EU approach has both supporters (Casonato and Olivato, 2024, pp. 87–112) and opponents (Smuha and Yeung, 2025, pp. 228–258; Finocchiaro, 2024, pp. 1961–1968). In the context of the issue discussed in the article, it is noteworthy that transparency obligations are insufficient. In particular, providers are required to disclose information about models to the competent public authority and other downstream providers. Moreover, it follows from the Code of Practice that the organizations that have signed it undertake to publish only information from the Safety and Security Framework that is relevant to the public's understanding of systemic risks and, where necessary, information on the assessment and mitigation of those risks. Thus, transparency measures do not fully cover the interests of end users of AI products, as the latter will only have access to limited information about the relevant models.

Following the election of Donald Trump as the new American President in 2016, the United States embarked on a course to remove regulatory barriers limiting the development of AI technologies. First, Trump issued Executive Order 14179, "Removing Barriers to American Leadership in Artificial Intelligence,"¹⁵ which repealed, among other things, the previous U.S. President's Executive Order 14110, "On the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."¹⁶ Following this measure, executive authorities were instructed to review all documents developed in accordance with Executive Order No. 14110. Also in July 2025, the U.S. Presidential Administration issued the AI Action Plan,¹⁷ which, in particular, contains a provision that the government should not allow federal funding for AI technology development to be directed to states with burdensome AI regulations. Experts note that

¹⁵ Removing Barriers to American Leadership In Artificial Intelligence. Available at: <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/> [Accessed 15.08.2025].

¹⁶ Executive Order 14110 of 30 October 2023. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Available at: <https://www.federalregister.gov/d/2023-24283> [Accessed 15.08.2025].

¹⁷ America's AI Action Plan. Available at: <https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf> [Accessed 15.08.2025].

the existence of such a funding principle, in the absence of comprehensive federal regulation, could negatively affect public confidence in AI technologies.¹⁸ Thus, the regulatory landscape in the United States in the field of AI is a combination of disparate mandatory rules, advisory documents (Siddiqui et al., 2025), and judicial practice (Bullock et al., 2024) that do not form a coherent system of legal regulation.

Regarding foundation models specifically, the approach to regulating their development can be found in several guidance documents prepared by the National Institute of Standards and Technology (NIST). First and foremost, it is the AI Risk Management Framework (AI RMF), which was established in 2023 to lay the groundwork for developing risk management systems among entities that develop and deploy AI systems. In response to the widespread use of foundation models, NIST has also developed the AI RMF: Generative AI Profile.¹⁹ The latter was developed to help organizations to determine the best way to manage risks at different stages of the GenAI lifecycle.

GenAI Profile provides over 200 recommendations for AI actors. Its undeniable advantage is the high level of detail in its recommendations. For example, organizations that use third party GenAI systems are advised to review their contracts with the relevant suppliers (GV-6.2-007). In particular, it is recommended to be wary of provisions that allow providers to terminate services arbitrarily and suddenly, or that allow them to unpredictably shift responsibility for incidents to the organizations themselves, as well as those that give providers or third parties the right to process data generated by the organization as it uses the GenAI system.

However, the advantages mentioned above also imply certain disadvantages of the framework. First and foremost, there is no clear distribution of rights and responsibilities among specific AI actors located throughout the AI ecosystem chain. It is unclear which recommendations apply to which group of AI actors. For example, recommendation GV-6.1-005 suggests that organizations implement a system for assess-

¹⁸ Federation of American Scientists Statement on the Preemption of State AI Regulation in the One Big Beautiful Bill Act. Available at: <https://fas.org/publication/fas-opposes-ai-regulation-preemption/> [Accessed 15.08.2025].

¹⁹ AI RMF: Generative Artificial Intelligence Profile. Available at: <https://doi.org/10.6028/NIST.AI.600-1> [Accessed 15.08.2025].

ing compliance by third parties involved in the development of GenAI with standards for ensuring the traceability of data and technologies in order to detect anomalies or unauthorized changes in the operation of the system. It is unclear whether only those entities that use third party AI systems in their activities should follow this recommendation, or whether it also applies, for example, to developers who train the model and operators who refine the model through fine-tuning. This also raises the issue of the complexity of applying the document in practice. On the one hand, there are numerous recommendations, but on the other hand, due to their quantity and quality, it becomes difficult for specific entities to implement the entire process within their organizations. This requires significant labor resources and, moreover, calls into question the correct understanding of the distribution of roles and responsibilities of organizations in the AI chain ecosystem.

Other researchers have also drawn attention to the above and several other shortcomings. For example, as part of the U.S. Open Loop program, an analysis was conducted of the framework provisions and the practice of its direct implementation in the activities of 15 companies specializing in NIST AI 600-1 or GenAI Profile.²⁰ The experts and company representatives surveyed identified the following as the main shortcomings of the framework: a) the lack of distributed roles and responsibilities for specific AI actors throughout the AI ecosystem chain; b) the lack of measures reflecting the specific impact of open-source models on risk management; c) difficulties in the interoperability of this framework with other similar documents.

In the United States, AI regulation is developing not only at the national institutional level, such as NIST, but also within the framework of industry regulators developing their own guidance documents. One of the key cross-cutting acts is the AI Accountability Policy Report (2024) prepared by the National Telecommunications and Information Administration (NTIA).²¹

²⁰ Generative AI Risk Management and the NIST Generative AI Profile (NIST AI 600-1). Available at: https://openloop.org/reports/2024/10/report-2-nist-generative-ai-profile.pdf?utm_source=chatgpt.com [Accessed 15.08.2025].

²¹ AI Accountability Policy Report, (2024). Available at: <https://www.ntia.gov/issues/artificial-intelligence/ai-accountability-policy-report> [Accessed 15.08.2025].

The report emphasizes that ensuring the proper exchange of information between all participants at all stages of the AI system lifecycle is critical to the development of accountable and secure technologies. In this regard, the NTIA proposes that executive authorities implement several measures aimed at enhancing the transparency and accountability of developers.

First, it is recommended to use existing or newly granted powers to mandate independent audits of high-risk AI systems. Such audits should confirm that models comply with established standards of security, ethics, and non-discrimination.

Second, the NTIA encourages government agencies to collaborate with stakeholders to enhance disclosure standards. The report notes the need to tailor disclosures to the target audience, which may require the creation of multiple reporting formats with varying levels of detail or the involvement of specialized information intermediaries. Recommended transparency tools include model/system cards, data sheets, and AI nutrition labels.

The report also emphasizes the need to strengthen interdepartmental cooperation and develop institutional capabilities for managing cross-cutting risks associated with AI. To this end, the government is encouraged to establish a system of measures to ensure transparency and control over the use of AI technologies.

In particular, it recommends to adopt the following measures: establishing a national register of high-risk AI use cases and a platform for reporting incidents; maintaining a public register of AI system audits; participating in the development of audit standards and auditor certification; conducting preliminary testing and certification of high-risk systems prior to their implementation; collecting regular confirmation of the characteristics of implemented solutions; coordinating accountability policies with other government agencies, including foreign partners.

In the United States, there is also the Centre for AI Standards and Innovation (CAISI) that plays an essential role in the chain of interaction between the government and representatives of the technology business. Its key functions include developing recommendations and best practices for assessing and improving the security of AI systems

in collaboration with NIST. For example, in January 2025, CAISI published the second draft of guidelines on Managing Misuse Risk for Dual-Use Foundation Models (NIST AI 800-1) that contain best practices for identifying, assessing, and mitigating risks to public safety and national security at different stages of the AI lifecycle.

Another critical area of CAISI's work is conducting independent assessments of foundation models, both American²² and those from foreign countries. In particular, CAISI recently conducted an assessment of models developed by the Chinese company DeepSeek. Based on the results of the assessment, CAISI made a number of conclusions, namely: AI models from developer DeepSeek were found to lag behind U.S. models in performance, cost, security and adoption; security shortcomings and censorship may pose risks to application developers, consumers and U.S. national security; DeepSeek models advance Chinese Communist Party narratives; DeepSeek's products are contributing to a rapid rise in the global use of models from China.

Thus, all the measures outlined above are designed to form a unified system of government oversight and accountability in the field of AI. However, given the U.S. policy of deregulation in this area, the political and, even more so, the legal significance of such recommendations remains highly questionable. At the same time, political initiatives in the United States in previous years²³ have led to leading companies in the field of AI developing and publishing their own security

²² CAISI, in collaboration with UK AISI, evaluated OpenAI's o1 and Anthropic's Claude 3.5 Sonnet foundation models. They tested these models to assess their capabilities in three areas: 1) cyber capabilities, 2) biology capabilities, and 3) software and AI development. Pre-Deployment Evaluation of OpenAI's o1 Model. Available at: <https://www.aisi.gov.uk/work/pre-deployment-evaluation-of-openais-o1-model> [Accessed 15.08.2025]; Pre-Deployment Evaluation of Anthropic's Upgraded Claude 3.5 Sonnet. Available at: <https://www.aisi.gov.uk/work/pre-deployment-evaluation-of-anthropics-upgraded-claude-3-5-sonnet> [Accessed 15.08.2025].

²³ Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI. Available at: <https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/> [Accessed 15.08.2025].

frameworks.²⁴ Although such frameworks are voluntary and government agencies do not monitor their compliance, they appear to contribute to greater transparency and security in the functioning of foundation models.

The regulation of AI in Russia is currently the subject of active scientific debate (Khabrieva, 2025, p. 610). Domestic doctrine has accumulated a substantial body of research addressing issues of terminology (D'yakonova, 2024, pp. 52–63), legal liability (Zenin et al., 2024, pp. 163–173; Bakhteev, 2024, pp. 9–23), data flow (Kovaleva and Zhirnova, 2024, pp. 109–121), and numerous other topics (Kiselev, 2024, pp. 140–151; Zykov, 2024, pp. 131–139). However, despite their diversity, the problems of legal regulation of the development of foundation models have not yet become an independent subject of study. In this regard, the study of the Russian Federation's approach will be based on an examination of existing legal initiatives and acts in the field of AI.

It seems appropriate to begin with the Russian AI Strategy. Although it defines large foundation models, it does not envisage any special regulation regarding their providers. The only relevant provision is Para. 38 of the Russian AI Strategy, which identifies “defining rules for working with large generative models and their use” as one of the primary directions for developing regulation in the field of AI. To date, such rules have not been developed, and it is unknown at what level and of what nature they are planned to be established.

It is likely that the Russian AI Strategy does not pay much attention to legal regulation because it was developed with the aim of stimulating innovation. At the same time, a systematic analysis of its other provisions leads to the conclusion that the Russian Federation intends to regulate the development and application of so-called “trusted AI technologies” first and foremost. These are understood to be “technologies that meet safety standards, developed in accordance with the principles of objectivity, non-discrimination and ethics, and which, when used, exclude the possibility of causing harm to humans and violating their

²⁴ Frontier AI Safety Policies. Available at: <https://metr.org/faisc> [Accessed 15.08.2025].

fundamental rights and freedoms, or damaging the interests of society and the State.”

The thesis outlined above is based on the fact that, according to the Russian AI Strategy, only AI products that are “trusted” will be allowed to be used in critical infrastructure (government administration, medicine, etc.). In particular, Para. 35 of the Russian AI Strategy provides for the creation of a register of trusted AI technologies for public authorities and organizations (including foundation models);²⁵ the introduction in federal government bodies of only those AI solutions that have been certified; ensuring the centralized development and distribution in government bodies and organizations of model solutions created on the basis of trusted AI technologies.

The Federal Service for Technical and Export Control (FSTEC) is responsible for developing the procedure for classifying specific AI technologies as trusted. Open-source reports indicate that by the end of 2025, the FSTEC will submit a draft standard for the secure development of AI systems for public discussion.²⁶ Based on the available data, it is not possible to conclude whether the proposed standard will address the issues associated with foundation models in any way.

At the legislative level, the draft federal law “On the Regulation of Artificial Intelligence Systems in the Russian Federation” (Balashova and Lisycina, 2025) deserves attention (the text of the draft is available to the authors). This initiative has clear advantages and addresses significant issues in the field of AI regulation, including content labeling, establishing legal liability, and determining the legal framework for results using AI-generated output. However, the initiative does not reflect the fact that many AI systems operate on the basis of foundation models. This necessitates a different approach to defining the range of entities involved in the development and deployment of AI systems, as

²⁵ Artificial intelligence will gain trust through a registry. Available at: <https://www.comnews.ru/content/239300/2025-05-21/2025-w21/1007/iskusstvennyy-intellekt-obretet-doverie-cherez-reestr> [Accessed 15.08.2025]. (In Russ.).

²⁶ FSTEC has taken up the development of a standard for the safe development of AI. Available at: <https://www.comnews.ru/content/241320/2025-09-19/2025-w38/1007/fstek-vzyalas-za-razrabotku-standarta-bezopasnoy-razrabotki-iskusstvennogo-intellekta> [Accessed 15.08.2025]. (In Russ.).

well as establishing specific requirements depending on whether the foundation models are frontier or not.

One can also consider the draft Concept for the Development of Regulation in the Field of Artificial Intelligence Technologies until 2030.²⁷ The document recognizes the critical role of foundation models and predicts an increase in their influence on the industry. However, it also contains some highly controversial provisions, in particular stating that large language models belong to the “category of narrow AI, as they are effective in specific areas, but do not have the ability to generalize knowledge and transfer experience to fundamentally new tasks.” This statement appears to be incorrect. After all, the architecture of large language models,²⁸ especially those based on transformers, allows the knowledge gained from training on general data to be used to solve new, previously encountered tasks (Bommasani et al., 2021). For example, a model trained on texts can successfully apply the accumulated knowledge to generate code, translate, summarize documents, or analyze legal texts — even though it was not directly trained to do so.

The Concept contains several significant shortcomings, which the expert community has critically analyzed.²⁹ Of particular concern is the lack of clear regulatory guidelines for foundation models. Despite their critical importance and potential risks, the document does not provide specific guidelines for either classifying foundation models or developing appropriate legal regimes. Instead, there is a clear tendency to delegate these issues to mechanisms of soft law.

This circumstance makes it relevant to analyze the soft law mechanisms existing in the Russian Federation, their effectiveness and readiness to address the relevant issues. The primary documents in Russia

²⁷ The Ministry has prepared a concept for regulating artificial intelligence. Available at: <https://www.vedomosti.ru/economics/articles/2025/08/20/1132830-mintsifri-podgotovilo-kontseptsiyu-regulirovaniya-iskusstvennogo-intellekta> [Accessed 15.08.2025]. (In Russ.).

²⁸ Large language models are a type of a foundation model.

²⁹ Draft concept for the development of AI regulation until 2030: analysis of key provisions and potential risks. Available at: <https://pravo.hse.ru/religionandlaw/news/1079313649.html> [Accessed 15.08.2025]. (In Russ.).

on ethical issues are the “AI Code of Ethics,”³⁰ the “Declaration on Responsible Generative AI,”³¹ and the “AI Code of Ethics in Medicine and Healthcare”³² developed by the AI Alliance. Among them, the most significant for the purposes of this study is the Declaration on Responsible Generative AI, which is a concise document containing recommendations for developers and users of generative AI systems.

The document uses the terms “foundation models” and “generative AI-based services,” but the relationship between these concepts is not explained, which creates some confusion. An analysis of the content of the recommendations to developers reveals that they are, in fact, addressed to developers of end AI products — entities involved in adapting and implementing applications. For example, Point 2 of the recommendations requires organizations to “thoroughly test generative AI-based services before launching them, as well as to develop and implement systematic monitoring procedures.” Such wording clearly refers to the stage of implementing specific applications, which leads to the conclusion that the recommendations are aimed at downstream providers rather than developers of foundation models. It follows from the above that ethical documents do not impose any obligations specific to developers of foundation models.

The next possible level of regulation is the internal documents of the companies, which develop foundation models. For example, Yandex has developed a family of large language models called YandexGPT 5,³³ and Sber has developed GigaChat.³⁴ This raises the logical question of whether these companies have policies in place to regulate the development of such models.

³⁰ Code of Ethics in the field of AI. Available at: <https://a-ai.ru/code-of-ethics/> [Accessed 15.08.2025]. (In Russ.).

³¹ Declaration of Responsible Generative AI. Available at: <https://ethics.a-ai.ru/genai-declaration/> [Accessed 15.08.2025]. (In Russ.).

³² Code of Ethics in the field of AI in medicine and healthcare. Available at: https://a-ai.ru/?page_id=3025 [Accessed 15.08.2025]. (In Russ.).

³³ YandexGPT 5.1 Pro. Available at: <https://ya.ru/ai/gpt> [Accessed 15.08.2025]. (In Russ.).

³⁴ GigaChat. Available at: <https://developers.sber.ru/dev> [Accessed 15.08.2025]. (In Russ.).

To date, it is known that Yandex has obtained ISO/IEC 42001³⁵ international standard certification, confirming that the company has established a responsible management system related to the development and provision of large language models through the Yandex Foundation Models³⁶ service. Sber is taking various steps to promote a responsible approach to the development of AI systems, particularly within the framework of the aforementioned AI Alliance. In particular, the company, like many others, is a signatory to the AI Ethics Code. However, as noted above, this and other ethical documents do not contain specific provisions relating to the development of foundation models. The authors were unable to find any other relevant internal policies or frameworks from open sources.

It follows from the above that in the Russian Federation, both at the legal and ethical levels of AI governance, there is a gap in ensuring accountability of foundation models towards developers. None of the existing documents regulating the responsible and ethical use of AI contains provisions directly addressed to entities that train such models.

IV. Conclusion

The analysis conducted in this article demonstrates that the Russian AI regulation lacks specific rules covering relations related to the development and deployment of foundation models. Strategic and conceptual documents only provide general guidelines and do not address issues such as the classification of foundation models or the need to develop specific regulations for their development. Even soft law mechanisms are primarily addressed to developers implementing AI solutions, leaving out a key link in the chain of AI system creation — the developers of foundation models.

Meanwhile, it is precisely at the stage of designing and training a foundation model that key risks arise in relation to data quality, architectural solutions, energy efficiency, security and the stability of sys-

³⁵ Responsible AI. Available at: <https://ya.ru/ai/responsible-ai> [Accessed 15.08.2025]. (In Russ.).

³⁶ Certificate. Available at: https://yastatic.net/s3/lpc-ext/static/maxmax/pp-browser/yandex_%D1%81_42001_25.pdf [Accessed 15.08.2025].

tems that operate on the basis of such models. The absence of principles and standards governing the activities of basic level developers effectively leaves a fundamental element of the entire AI ecosystem uncontrolled, rendering further attempts at regulation fragmented and futile.

In order to overcome the identified gaps, it seems appropriate to adhere to the following recommendations when formulating the Russian Federation's policy on AI regulation:

- to develop a classification of foundation models based on their level of impact on human rights, state interests and society. As a technical criterion for such a classification, it seems appropriate to use the computing power spent on training the relevant model;

- to establish separate legal statuses for foundation model developers, foundation model providers, and downstream providers;

- to consider the possibility of establishing special requirements for model developers and providers aimed at raising awareness among downstream providers about the capabilities and limitations of the model (e.g., encourage the publication of model or system cards);

- to develop soft law mechanisms, in particular, prepare a code of conduct, best practices, and ethical recommendations for developers and providers of foundation models;

- to consider the possibility of establishing links between the emerging regime for trusted AI technologies and the development of foundation models, in particular, when designing requirements for trusted technologies, take into account aspects related to the fact that such products may function on the basis of foundation models.

It appears that taking these recommendations into account will enhance the quality of artificial intelligence regulation in Russia, aligning it more closely with international standards and contemporary technological realities.

References

Aloncho, S. et al., (2025). Comparing Apples to Oranges: A Taxonomy for Navigating the Global Landscape of AI Regulation. In: *Proceedings of June 23–26, 2025 (FAccT '25)*, USA, doi: 10.1145/3715275.3732059.

Bakhteev, D.V., (2024). The Legal Status of Artificial Intelligence Systems and a Model for Differentiating Legal Liability for the Damage They Cause. *Lex Russica*, 4, pp. 9–23, doi: 10.17803/1729-5920.2024.209.4.009-023. (In Russ.).

Balashova, A. and Lisycina, M., (2025). Russia will discuss banning AI with an alarming level of risk. Available at: https://www.rbc.ru/technology_and_media/11/04/2025/67f7dc399a79477fdd97bf30?from=from_main_3 [Accessed 15.08.2025]. (In Russ.).

Bommasani, R. et al., (2021). On the Opportunities and Risks of Foundation Models. Available at: <https://arxiv.org/abs/2108.07258v3> [Accessed 15.08.2025].

Brown, I., (2023). Allocating accountability in AI supply chains: a UK-centered regulatory perspective. Available at: <https://www.adalovelace-institute.org/resource/ai-supply-chains/> [Accessed 15.08.2025].

Bullock, C., van Arsdale, S. et al., (2024). Legal Considerations for Defining “Frontier Model”. *Institute for Law & AI Working Paper*, 2, pp. 1–28, doi: 10.2139/ssrn.4973370.

Caputo, N., (2025). Governing AI Beyond the Pretraining Frontier. Available at: <https://arxiv.org/abs/2502.15719> [Accessed 15.08.2025].

Casonato, C. and Olivato, G., (2024). AI Regulation in Europe: Exploring the Artificial Intelligence Act. In: Fabris, A. and Belardinelli, S. (eds). *Digital Environments and Human Relations*. Springer, pp. 87–112, doi: 10.1007/978-3-031-76961-0_5.

Cheng, H. et al., (2023). A Survey on Deep Neural Network Pruning-Taxonomy, Comparison, Analysis, and Recommendations. Available at: <https://arxiv.org/abs/2308.06767> [Accessed 15.08.2025].

Chun, J. et al., (2024). Comparative Global AI Regulation: Policy Perspectives from the EU, China, and the U.S.. Available at: <https://arxiv.org/abs/2410.21279> [Accessed 15.08.2025].

Datta, A., (2025). The Regulation of GPAI Model Providers Under the EU AI Act. In: Denga, M. and Hornuf, L. (eds). *Regulatory Competition in the Digital Economy*, pp. 119–138, doi: 10.1007/978-3-031-81089-3_6.

D'yakonova, O.G., (2024). On the Concept of Artificial Intelligence Technologies. *Courier of the Kutafin University (MSAL)*, 3, pp. 52–63. (In Russ.).

Fernández-Llorca, D. et al., (2025). An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI. *Artificial Intelligence and Law*, 33, pp. 875–888, doi: 10.1007/s10506-024-09412-y.

Finocchiaro, G., (2024). The regulation of artificial intelligence. *AI & Society*, 39, pp. 1961–1968, doi: 10.1007/s00146-023-01650-z.

Gstrein, O.J., Haleem, N. and Zwitter, A., (2024). General-purpose AI regulation and the European Union AI Act. *Internet Policy Review*, 13(3), doi: 10.14763/2024.3.1790.

Hacker, P. et al., (2025). *The Oxford Handbook of the Foundations and Regulation of Generative AI*. Oxford University Press. DOI: 10.1093/oxfordhb/9780198940272.001.0001.

Hacker, P., (2023). AI Regulation in Europe: From the AI Act to Future Regulatory Challenges. Available at: <https://arxiv.org/abs/2310.04072> [Accessed 15.08.2025].

Heim, L. and Koessler, L., (2024). Training Compute Thresholds: Features and Functions in AI Regulation. Available at: <https://arxiv.org/abs/2405.10799> [Accessed 15.08.2025].

Ivanova, E.S., (2022). On the Legal Regulation of Artificial Intelligence in the Russian Federation (A Comparative Legal Analysis). *State and Law*, 11, pp. 202–206. (In Russ.).

Khabrieva, T.Ya., (2025). Legal Problems of Artificial Intelligence Identification. *Bulletin of the Russian Academy of Sciences*, 7, p. 610, doi: 10.31857/S0869587324070015. (In Russ.).

Kiselev, A.S., (2024). Problems and Prospects of Legal Regulation of Social Relations Related to the Use of Neural Networks. *Lex Russica*, 2, pp. 140–151, doi: 10.17803/1729-5920.2024.207.2.140-151. (In Russ.).

Koivisto, I., Koulu, R. and Larsson, S., (2024). User accounts: How technological concepts permeate public law through the EU's AI Act. *Maastricht Journal of European and Comparative Law*, 31(3), pp. 412–432, doi: 10.1177/1023263X241248469.

Kovaleva, N.N. and Zhirnova, N.A., (2024). Issues of Ensuring Personal Data Confidentiality in the Use of Artificial Intelligence Systems. *Journal of Russian Law*, 28(7), pp. 109–121, doi: 10.61205/S160565900027561-1. (In Russ.).

Placani, A., (2024). Anthropomorphism in AI: hype and fallacy. *AI and Ethics*, 4, pp. 691–698, doi: 10.1007/s43681-024-00419-4.

Polyakova, T.A. and Smirnov, A.A., (2024). Legal Aspects of Ensuring Information Security in the Use of Artificial Intelligence Technologies. *Moscow University Bulletin*, 26(3), pp. 89–106, doi: 10.55959/MSU2413-631X-27-15-3-07. (In Russ.).

Price, C., (2025). Algorithmic Advancement in Artificial Intelligence. Available at: https://www.rand.org/pubs/research_reports/RRA3485-1.html [Accessed 15.08.2025].

Raman, D. et al., (2025). Intolerable Risk Threshold Recommendations for Artificial Intelligence. Available at: <https://arxiv.org/abs/2503.05812> [Accessed 15.08.2025].

Sastry, G. et al., (2024). Computing Power and the Governance of Artificial Intelligence. Available at: <https://arxiv.org/abs/2402.08797> [Accessed 15.08.2025].

Siddiqui, S. et al., (2025). Promising Topics for U.S. — China Dialogues on AI Safety and Governance. Available at: <https://www.oxfordmartin.ox.ac.uk/publications/promising-topics-for-us-china-dialogues-on-ai-safety-and-governance> [Accessed 15.08.2025].

Smuha, N.A. and Yeung, K., (2025). The European Union’s AI Act: Beyond Motherhood and Apple Pie? In N.A. Smuha (ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*. Cambridge University Press, pp. 228–258, doi: 10.1017/9781009367783.015.

Villalobos, P. et al., (2022). Machine Learning Model Sizes and the Parameter Gap. Available at: <https://arxiv.org/abs/2207.02852> [Accessed 15.08.2025].

Winter, C. and Bullock, C., (2024). The Governance Misspecification Problem. *Institute for Law & AI Working Paper Series*, 3, pp. 1–29. Available at: <https://law-ai.org/the-governance-misspecification-problem> [Accessed 20.09.2025].

Zenin, S.S., Yapyrintsev, I.M. and Ivanova, L.V., (2024). The Category of “Legal Responsibility” in the Context of Its Application to Artificial Intelligence Systems. *State and Law*, 11, pp. 163–173. (In Russ.).

Zykov, S.V., (2024). The Problem of Artificial Intelligence Personhood. *Actual Problems of Russian Law*, 2, pp. 131–139, doi: 10.17803/1994-1471.2024.159.2.131-139. (In Russ.).

Information about the Authors

Osman A. Izhaev, Cand. Sci. (Law), Senior Research Fellow, School of Law and Administration, University of Tyumen, Tyumen, Russian Federation (Corresponding Author)

izhaev.osman@gmail.com

ORCID: 0000-0003-3777-8927

Dmitry L. Kuteynikov, Cand. Sci. (Law), Senior Research Fellow, School of Law and Administration, University of Tyumen, Tyumen, Russian Federation

kuteynikov@me.com

ORCID: 0000-0003-1448-3085

Received 14.10.2025

Revised 18.11.2025

Accepted 23.11.2025