Original article
UDC 343.3/.7:004
doi 10.46741/2686-9764.2024.68.4.001

Prospects of Legal Regulation and Algorithm for Marking Generative Content

NIKOLAI F. BODROV
Kutafin Moscow State Law University (MSAL), Moscow, Russia, bodrovnf@gmail.com, https://orcid.org/0000-0002-9005-3821

ANTONINA K. LEBEDEVA
Kutafin Moscow State Law University (MSAL), Moscow, Russia, tonya109@yandex.ru, https://orcid.org/0009-0004-9344-2103

Abstract

Introduction: the article analyzes mechanisms for legal regulation of generative content in the Russian Federation and considers an algorithm of marking generative content. The analysis of current legislation shows that the increasing role of artificial intelligence technologies in modern society necessitates the improvement of existing norms, as well as the development of new ones for the legal regulation of generative content turnover. The Russian legislator is trying to develop legislative measures to control the turnover of generative content, but currently there is not even a unified definition of a deepfake as a fundamental concept of this sphere. *Purpose*: on the basis of the existing classification of generative content types to describe the existing types of generative content labelling and to present the definition of a deepfake. *Methods*: comparative-legal, empirical methods of description, interpretation; theoretical methods of formal and dialectical logic, legal-dogmatic and method of interpretation of legal norms, are applied. *Results*: the analysis of existing types of generative content, the purposes of its distribution demonstrates the need to create effective legal and technological mechanisms for regulating generative content. Theoretical provisions on the structure of subjects of creation and distribution of generative content are developed and recommendations on establishing their legal duties and responsibilities are offered. *Conclusions*: the mechanisms of legal regulation of generative content at this stage are reduced to the following: establishing the obligation of technology companies and users who create generative content to use watermarks to mark the content in order to inform other persons about the generative nature of the content; establishing liability for the refusal of marking or removal of any type of marking; establishing liability for the misuse of biometric personal data for the creation of a deepfake.

Key words: deepfake; generative content; legal regulation; content marking; liability.

5.1.1. Theoretical and historical legal sciences.

5.1.2. Public law (state law) sciences.

5.1.4. Criminal law sciences.

Acknowledgements: This publication has been prepared as part of the work under the state assignment on the topic “The Russian legal system in the realities of digital transformation of society and the state: adaptation and prospects for responding to modern challenges and threats (FSMW-2023-0006)”. The EGISUNIOKTR registration number: 124012000079-6.

For citation: Bodrov N.F., Lebedeva A.K. Prospects of legal regulation and algorithm for marking generative content. *Penitentiary Science*, 2024, vol. 18, no. 4 (68), pp. 348–357. doi 10.46741/2686-9764.2024.68.4.001.

Introduction

Undoubtedly, technologies of artificial intelligence (AI), which help create various types of content, are a powerful tool for creative self-expression, in demand in various spheres of human activity, from art to, for example, marketing. However, uncontrolled distribution of various generative content created with the help of modern neural network algorithms, their ability to synthesize text, graphics and sounds already leads to serious consequences that deserve due attention from legislators and law enforcement officers.

The most important problem of our time is virtually uncontrolled access to biometric personal data of a person based on those media materials that in recent years have been distributed in open access on social networks, instant messaging systems, cloud storage services, file sharing sites, video conferencing software data storage systems and video hosting. These materials are already the basis for creating digital products in the form of text, graphics, sound or a combination of them, generated in whole or in part using neural network technologies for the user to break control systems and get access to data [1]. Such products are essentially the content of the term “deepfake”.

The problem under consideration is aggravated by the fact that the technologies of neural network synthesis of deepfake content can be used by a user with any level of technical training, who using specially configured applications or sites can follow instructions step by step and create a deepfake. The commercialization of AI technologies is undoubtedly a serious threat to the information security of society and the state.

So, according to the Hong Kong police, a financial officer paid 25 million dollars after a video call from a “financial director”, whose appearance and voice were synthesized by an attacker using neural network technologies [2].

Besides, there are risks, when biometric identification systems mistakenly refuse users, relying on false alarms of deepfake detectors. For example, in the Admitad system, the user was denied service as a result of video verification, because the system mistakenly classified his appearance as a deepfake [3].

Even preliminary results of the analysis of the statistics of offenses related to the distribution of deepfakes show a very negative trend. For example, according to the report of the Sumsub identity verification platform, various incidents with deepfakes in the financial sector increased by 700% in 2023 compared to the previous year [4].

If a year ago the legislator considered it inappropriate to criminalize the use of AI technologies for criminal purposes [5], now the Russian legal system clearly needs to create mechanisms for legal regulation of generative content [6], including as an aggravating element [7] of the crime. The draft law provides for the introduction of additional qualified composition into some articles of the Criminal Code of the Russian Federation, such as commission of a crime “using an image or voice (including falsified or artificially created) of the victim or another per-

son, as well as using biometric personal data of the victim or another person" [7].

It is planned to amend the legislation in terms of articles: "Libel", "Theft", "Fraud", "Extortion", "Causing property damage by deception or abuse of trust" (Part 2.1 of Article 128.1, Paragraph "d" of Part 3 of Article 158, Part 2.1 of Article 159, Paragraph "d" of Part 2 of Article 163, Paragraph "c" of Part 2 of Article 165).

The legislator does not give a clear definition of falsified or artificially created voices. However, based on the content of the explanatory note and feedback on the draft law, it can be assumed that they mean deepfakes.

However, the list of articles seems to us to be extremely limited [8], the list of real and potential threats associated with the use and distribution of deepfakes is much wider.

So, in our opinion, in the very near future, the spread of generative content synthesis systems will lead to a significant transformation of crimes (primarily the ways they are committed):

– against the person (articles 110, 128.1, 146 of the Criminal Code of the Russian Federation),

– in the sphere of economics (articles 159, 159.3, 159.6, 185.3 of the Criminal Code of the Russian Federation),

– against public safety and public order (articles 205.2., 207, 207.1, 207.3, 242, 242.1, 272, 273, 274 of the Criminal Code of the Russian Federation),

– against state power (articles 280, 280.1, 280.3, 280.4, 282, 282.4, 284.2, 303 Criminal Code of the Russian Federation),

– against the peace and security of mankind (Article 354 of the Criminal Code of the Russian Federation) [9].

This forecast is based on the fact that in the above-mentioned crimes, deepfakes can most likely be used as an instrument of committing a crime.

The legislator's approach to the determination of an extremely limited list of articles in the above-mentioned draft laws does not seem to be entirely correct. The emergence of new qualified compositions at this stage will not solve the problem of using and distributing deepfakes, and the primary task is to create new and improve existing mechanisms for regulating the turnover of generative content created with the help of artificial intelligence.

The draft law review report states that "the industry legislation does not regulate the use of identity substitution technologies (deepfake). Thus, the introduction of the proposed regulation into criminal legislation is not possible due to the lack of corresponding norms of substantive legislation, which entails significant risks of the formation of incorrect law enforcement practice" [10].

Thus, while offering protection to citizens from deepfakes, the draft law developers do not specify what exactly is meant by the term "deepfake". Correctness of the use of the term "artificial audio recording" from the standpoint of forensic expertise and forensic investigative practice is questionable due to the fact that it reflects a significantly broader scope of the concept.

In our opinion, the term "deepfake" should be defined as follows: a deepfake is a digital product in the form of text, graphics, sound or a combination of them, generated in whole or in part using neural network technologies for the user to break control systems and get access to data.

As for the normative consolidation, it seems to us correct to fix the definition of a deepfake in basic concepts of the Federal Law No. 149-FZ of July 27, 2006 "On Information, Information Technologies and Information Protection".

The implementation of the deepfake concept into Russian legislation alone is clearly not enough, since issues of legal regulation of the creation, use and distribution of generative content remain outside the attention of the legislator, and for law enforcement it is important to clearly distinguish deepfakes from other types of generative content.

Moreover, carrying out scientific research in the field of terminological support for the described problems, we come to the conclusion that at the moment there is not only a definition in the scientific and normative-technical literature, but also approaches to describing such a phenomenon as a fake self.

Summarizing results of the risk assessment of illegal distribution of generative content, we believe that a fake self should be understood as a deepfake generated by a user based on his/her own biometric data in order to commit illegal actions aimed at evading liability or misleading other persons regarding events presented

as having happened to the user him/herself. For example, in situations of insurance fraud [11], users use neural network synthesis algorithms to generate photos with property damage to receive insurance payments. In such a situation, employees of insurance organizations may collude with intruders and provide images generated by a neural network instead of photos, for example, a pre-insurance inspection.

Legal mechanisms to counteract the spread and use of fake content should provide an opportunity for both ordinary users and law enforcement employees to detect such content. If measures of forensic counteraction can be provided as a mechanism for detecting deepfakes [12], then the development of legal norms regulating the technology and processes of introducing mandatory marking of generative content are measures to prevent the use and distribution of deepfakes.

A deepfake is just a kind of generative content. By distributing generative content, creators do not hide the source of its origin, even if the result of generation reaches a high degree of realism. In the course of legitimate use of generative content, it is necessary to ensure the availability of information about the method of its creation. The spread of deepfakes is connected with the concealment of a way to simulate authentic content. To inform users about the fact of content generation by the means of AI technologies, it is necessary to develop norms providing for the mandatory effective marking of any generative content.

Nowadays, there are no effective measures to ban generative content creation. The National Strategy for the Development of Artificial Intelligence approved by the Decree of the President No. 124 of February 15, 2024 fixes the need to "consolidate favorable regulatory conditions for the development and implementation of artificial intelligence technologies", while a complete ban on generative AI technologies certainly does not implement such conditions. Potential and real threats come not from the fact of creating AI technologies to generate various content, but from the fact of unfair use of the results of such generation without specifying the method of obtaining a digital product. It is safe to say that there is a risk of using and distributing generative content that can easily be mistaken for authentic.

Ways to mark generative content

In order to streamline the turnover of generative content without compromising information security and preventing misinformation, it is necessary to provide effective marking solutions and limit its use without appropriate marking.

In a similar situation, the mechanisms of legal regulation actually boil down to the creation of norms on marking generative content, in order to maintain information awareness of users about such content and its origin. Technologies suitable for practical use are already being used in various fields (including the field of legal proceedings). Legislation, therefore, should limit the turnover of generative content and establish liability for breaking rules of informing about the generative nature of content.

Along with the mechanisms of legal regulation on mandatory marking of generative content, it is also necessary to develop technological marking methods. First of all, there is a need for uniform international standards in the field of marking generative content. The process and technologies of distributing deepfakes have no geographical boundaries. Generative content is created, used and distributed worldwide. Therefore, it is necessary to develop unified standards for marking generative content in order to enable users to easily detect them, regardless of the language, the technology platform used or the state in which the content consumer resides.

Among the currently used marking methods, graphic marking has become quite widespread. Markers in such systems are a short text to a digital product (for example, "created with the help of AI") by analogy with the marking of foreign agents or extremist organizations (federal laws No. 255-FZ of July 14, 2022 "On Control over Activities of Persons under Foreign Influence", No. 114-FZ of July 25, 2002 "On Countering Extremist Activities"), or graphic objects used by analogy with the marking of information products in accordance with the Federal Law No. 436-FZ of December 29, 2010 "On the Protection of children from information harmful to their health and development|. Such marking is obviously of a notification nature, but due to the simplicity of its presentation form, it cannot be used in specialized areas, for example, in the field of legal proceedings.

Watermarking is a technologically more complex option. “Watermarks” are digital signatures that are embedded in digital files, such as images, video recordings, and phonograms. They can be used to identify the source of origin of a digital product, prevent unauthorized copying or distribution, as well as to track the movement of content” [8].

For example, Telesystem, which has been operating on the Russian market for more than 20 years, develops and manufactures EDIC-mini voice recorders [13]. Files with phonograms recorded on dictaphones of this brand are provided with certain digital markers of the authenticity of the recording (with additional metadata). It protects the recording from unauthorized use and verify the integrity of files. This function is in high demand in forensic investigation and expert practice.

Hour One, a company that creates AI avatars, uses the watermark “AV” (AlteredVisuals) [14] to mark its videos. As the developers point out, this is done out of respect for the end user’s right to know that the video was generated using AI technologies.

To prevent the spread of fake news and misinformation, large corporations already have experience confirming the authenticity of digital products. The Coalition for Content Provenance and Authenticity introduced an Official Content Credentials Icon (C2PA standard) [18].

The C2PA technology standard developed by the coalition allows creators to embed metadata into digital products to verify their origin and related information. The C2PA standard is intended not only for generative images, its use is also planned by camera manufacturers, media companies that create visual content to certify the source and origin of media content. Companies such as OpenAI, Adobe, Microsoft, PublicisGroupe, Leica, and Nikon have integrated this standard into their activities.

Such labeling will contain both metadata stored in the file that is invisible to the user’s eye, and a visible CR watermark that will appear in the upper-left corner of each image.

Images generated using neural networks such as ChatGPT and DALL·E have included C2PA metadata since February 2024. To check the “history” of an image, the user can use their Content Credentials Verify service [16]. However, this marking approach is applied only to graphic forms of generative content. In addition, the user can remove the marking by taking a regular screenshot of the generated image, thereby creating a new file with “clean” metadata. A watermark can also be deleted when framing an image.

Moreover, Microsoft has developed Azure OpenAI Service, which adds invisible watermarks to all images generated using DALL·E [17]. Azure AI Speech service embeds watermarks for users to determine whether speech is synthesized using Azure AI Speech and which voice is used during generation [18].

Google has introduced a beta version of a tool for embedding digital watermarks directly into images, audio, text or video created using AI – SynthID [19]. This tool allows users to embed a digital watermark directly into digital products created by artificial intelligence. And subsequently, based on these watermarks, it is possible to verify content, but only if it was generated by Google’s AI tools.

There are services offering to embed watermarks for audio files to prevent the use of user data by artificial intelligence models, tracking them back to the source. The developers of systems for synthesizing and cloning sounding speech also offer their own technologies for creating watermarks. For instance, Assemble AI has developed a complex deep neural network watermark “PerTh” [20], which helps embed invisible data into the generated speech, creating an invisible watermark. According to the developers, this watermark is difficult to remove, but the verification option is available only for phonograms generated by Resemble AI.

The MyVocalAI neural network [21] offers various possibilities for synthesizing and cloning sounding speech and provides users with the ability to create watermarks for marking generative content. However, users with a paid subscription (starting with a standard subscription for about 1,000 rubles per month) get the technical ability to remove markings by creating phonograms with cloned voices not only by directly recording their voice, but also by downloading any phonograms of any speaker.

Due to the fact that a significant amount of generative content is distributed on social networks, it is reasonable to create services on various social platforms that would automatically recognize generative content and notify

users about it. Despite the technological complexity of this process, some corporations have sufficient resources to create such services.

For example, the Youtube video hosting service announces a new obligation for users to disclose information about whether the uploaded video is altered or synthesized [22]. In addition, the company states that their latest update will also automatically identify similar content and put these labels in the video description [22]. However, users mention that the mechanism for detecting generated content has not yet been debugged and is extremely unstable, there are cases when authentic content has been recognized as generated. The company also plans to adjust the algorithm for deleting generative content in the next few months if users determine that a custom face or voice is used to create it. However, cases of abuse of this right cannot be excluded. For example, public figures can apply for the removal of videos discrediting them, insisting on this content to be generated while it is authentic.

Meta announces plans about marking images, videos, and audio on their social networks (Facebook, Instagram and Threads) with the text "Made with AI", if the developed algorithms themselves reveal signs of content generated using neural networks, or if the user notifies about it him/herself. If a photorealistic image is generated by their Meta AI neural network, it is already marked as "Imagined with AI" [23].

In June 2024, Meta introduced a AudioSeal technology for proactive detection of voice cloning with localized watermarking [24]. Along with the technology for creating watermarks, a tool for detecting watermark data was also introduced, which greatly simplifies the verification of a cloned voice. According to the developers, the applied watermarks are stable even in cases of various sound recording editing options. The watermark itself represents a certain signal in the generated phonogram, which is not distinguishable to the human ear. There is no information about the stability of marking in the situation of qualified installation of a phonogram in open print.

AudioSeal is a method of speech localized watermarking, it jointly trains a generator that embeds a watermark in the audio, and a detector that detects the watermarked fragments in longer audios.

Descript, a service for creating various types of content, also makes it possible to create AI speakers, having their 30-second consent to record. When saving a phonogram with a recording of a cloned voice, a user has a choice whether to save the file with metadata (which will contain information that the Descript service was used when creating the file) or without them.

In our opinion, an easy-to-implement mechanism for marking generative content, primarily for detecting it among authentic ones, is embedding information about the service used to generate it in the metadata of the file. The inclusion of an identification number of this generation can also become an additional authentication mechanism. But it is important to take into account the technological possibility of changing or deleting information in the metadata of generated objects, which significantly reduces the effectiveness of using such marking for the purpose of judicial establishment of circumstances.

Some mechanisms of legal regulation can be taken, for example, from blockchain technologies. For example, specialized resources [25] forecast technological solutions that, in our opinion, can be implemented in domestic legislation.

Despite obvious capacities of blockchain technologies in content marking, technological costs of implementing this process are high.

Technologically, watermarks are divided into two main types, namely "fragile" and "durable". Fragile ones are easily destroyed when users manipulate the media. Durable ones are more resistant to manipulation, but such watermarks are more difficult to detect for an ordinary user who does not have special knowledge.

The video encoding technology can be used in immutable blockchains as a mechanism for verifying video authenticity (in the industry, such technologies are implemented by companies such as FactomAxiom).

Similar technologies can initially be implemented in procedural legislation, for example, for the formal evaluation of digital files by the court. In any case, such solutions seem to be real, applicable and in demand by law enforcement practice.

Thus, the existing methods of content marking represent a variety of technologies, each of

which has its own characteristics and areas of application. At the moment, there are the following main types of marking:

– watermarking;

– embedding service information in the file metadata;

– blockchain technologies;

– visual marking.

Each of these types has both advantages and limitations, and the choice of a specific marking method depends on the tasks and nature of the content itself.

Certain mechanisms for legal regulation of the distribution of generative content related to its mandatory marking have already been introduced in some countries. For example, Paragraph 133 of the Artificial Intelligence Act developed and approved by the European Parliament contains measures for marking generative content [26]:

– watermarks, metadata identification, cryptographic methods to confirm the origin and authenticity of content, record-keeping methods, fingerprints or other methods;

– such methods should be sufficiently reliable, compatible, and effective, as far as technically possible, taking into account available methods or a combination of them, such as watermarks, metadata identification, cryptographic methods for confirming the origin and authenticity of content, registration methods, fingerprints, etc.

Specific measures of legal regulation of the generative content turnover are developed in the People's Republic of China. The following provisions of the Regulation on the Administration of the Deep Information Synthesis Service on the Internet are applied at the level of interdepartmental interaction [27]:

– Article 7 provides for the responsibility of companies providing resources for content synthesis,

– Article 9 prescribes the use of mechanisms for identifying users of generative synthesis systems,

– Article 16 obliges to mark generative content. Such a measure not only reduces the criminogenic potential of generative content, but also creates conditions for establishing a single source of origin of distributed files.

The progressive experience of Chinese colleagues should be necessarily implemented into domestic legislation. However, successful counteraction to crime associated with generative content requires development of framework international acts.

In the USA, the AI Disclosure Act of 2023 [28] requires generative artificial intelligence to disclose that their output has been generated by artificial intelligence.

However, it is important to note that antimarking services already exist. For example, a popular open access service successfully removes watermarks from various images (http://dewatermark.ai.ru).

At the same time, a lack of international technological standards for marking generative content actually negates any legal regulation measures. Without joint coordinated efforts to create unified mechanisms for legal regulation of generative content marking, all of the above methods are not effective.

Experts of the High-Level Advisory Body on Artificial Intelligence of the United Nations proposed their international recommendations on the regulation of artificial intelligence and published "Governing AI for Humanity: Final Report" in September 2024 [29]. The report notes the danger of the spread of deepfakes and emphasizes the importance of developing common standards for authenticating content and its digital origin.

The importance of strengthening international cooperation in the use of artificial intelligence technologies as one of the main tasks of AI development in the Russian Federation is also mentioned in the National Strategy for the Development of Artificial Intelligence [13].

Conclusion

Taking into account the considered structure of subjects, the mechanisms for legal regulation of generative content are reduced to the following:

– establishing the obligation of technology companies and users creating generative content to use watermarks to mark generative content;

– establishing liability for the refusal of marking or the removal of any type of marking;

– establishing liability for the misuse of biometric personal data to create a deepfake;

– taking into account public danger of acts committed using a deepfake as a means of committing an illegal act;

– development and procedural regulation of means to ensure the authenticity of digital evidence in court proceedings.

Summing up a certain intermediate result in considering the issue of developing legal regulation mechanisms and algorithms for marking generative content, it should be concluded that AI technologies for creating generative content have already become widespread both in various countries and legal systems, and have become part of many types of human activity. This is the reason for the actual impossibility of establishing any effective legal prohibition on the creation of generative content.

Therefore, the creation of national legal norms is important, but will not ensure the information security of the state in the long term. Only international cooperation in the field of AI legal regulation and creation of uniform universal norms on mandatory marking of generative content will be an effective means in combating deepfakes and protecting society and the state from their harmful effects.

REFERENCES

1. Bodrov N.F., Lebedeva A.K. The concept of deepfake in Russian law, classification of deepfake and issues of their legal regulation. *Yuridicheskie issledovaniya = Legal Studies,* 2023, no. 11, pp. 26–41. (In Russ.).
2. Finance worker pays out $25 million after video call with deepfake chief financial officer. *CNN.* Available at: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html (accessed September 20, 2024).
3. *Admitad otkazal v vyplate 600k, t.k skazal chto, ya dipfeik* [Admitad refused to pay 600k, because it said that I was a deepfake]. Available at: https://pikabu.ru/story/admitad_otkazal_v_vyiplate_600k_tk_skazal_chto_ya_dipfeyk_11295845?utm_source=linkshare&utm_medium=sharing (In Russ.). (Accessed September 20, 2024).
4. Deepfakes are coming for the financial sector. *The Wall Street Journal*. Available at: https://www.wsj.com/articles/deepfakes-are-coming-for-the-financial-sector-0c72d1e5 (accessed September 20, 2024).
5. The government did not support a draft law on criminal liability for deepfakes. *Informatsionnoe agentstvo TASS* [TASS News Agency]. Available at: https://tass.ru/obschestvo/17922853 (In Russ.). (Accessed September 24, 2024).
6. *O vnesenii izmenenii v chast' pervuyu Grazhdanskogo kodeksa Rossiiskoi Federatsii : proekt federal'nogo zakona No. 718834-8 (vnesen 16.09.2024 senatorami Rossiiskoi Federatsii A.A. Klishasom, A.G. Sheikinym, N.S. Kuvshinovoi, R.V. Smashnevym, deputatom Gosudarstvennoi Dumy D.V. Bessarabovym)* [On Introducing Amendments to Part One of the Civil Code of the Russian Federation: Draft Federal Law No. 718834-8 (introduced on September 16, 2024 by Senators of the Russian Federation A.A. Klishas, A.G. Sheikin, N.S. Kuvshinova, R.V. Smashnev, State Duma Deputy D.V. Bessarabov)]. Available at: https://sozd.duma.gov.ru/bill/718834-8 (accessed September 24, 2024).
7. *O vnesenii izmenenii v Ugolovnyi kodeks Rossiiskoi Federatsii: proekt federal'nogo zakona № 718538-8 (vnesen 16.09.2024 deputatom Gosudarstvennoi Dumy Ya.E. Nilovym, Senatorom Rossiiskoi Federatsii A.K. Pushkovym)* [On Amendments to the Criminal Code of the Russian Federation: Draft Federal Law No. 718538-8 (introduced on September 16, 2024 by State Duma Deputy Ya.E. Nilov, Senator of the Russian Federation A.K. Pushkov)]. Available at: https://sozd.duma.gov.ru/bill/718538-8 (accessed September 24, 2024).
8. Bodrov N.F., Lebedeva A.K. The concept of a deepfake in Russian law, its classification of deepfakes and problems of legal regulation. *Yuridicheskii vestnik Dagestanskogo gosudarstvennogo universiteta = Herald of Dagestan State University*, 2023, vol. 48, no. 4 (68), pp. 173–181. (In Russ.).
9. Bodrov N.F., Lebedeva A.K. Threats and challenges in the era of generative artificial intelligence, taking into account the criminogenic potential of deepfakes. In: *Sankt-Peterburgskii mezhdunarodnyi kriminalisticheskii forum: materialy mezhdunar. nauch.-prakt. konf.* [Saint Petersburg International Forensic Forum: proceedings of the international scientific and practical conference. Saint Petersburg, June 10–11, 2024]. Saint Petersburg, 2024. Pp. 62–65. (In Russ.).

10. *Ofitsial'nyi otzyv ot 22 iyulya 2024 g. No DG-P4-23438 na proekt federal'nogo zakona "O vnesenii izmenenii v Ugolovnyi kodeks Rossiiskoi Federatsii", vnosimyi v Gosudarstvennuyu Dumu deputatom Gosudarstvennoi Dumy Ya.E. Nilovym* [Official review No. DG-P4-23438 of July 22, 2024 on the draft federal law "On Amendments to the Criminal Code of the Russian Federation", submitted to the State Duma by State Duma Deputy Ya.E. Nilov]. Available at: https://sozd.duma.gov.ru/ (accessed September 24, 2024).
11. Warning that "shallowfake" images are the "next big scam" to hit Britain: Fraudsters are mocking-up pictures of car damage to con insurers – with number of cases surging by 300 per cent in a year. *Mail Online*. Available at: https://www.dailymail.co.uk/news/article-13373513/shallowflake-scam-warning-car-insurance.html (accessed September 18, 2024).
12. Bodrov N.F., Lebedeva A.K. Deepfake as an object of forensic examination. In: *Natsional'nye i mezhdunarodnye tendentsii i perspektivy razvitiya sudebnoi ekspertizy: sb. dokladov nauch.-prakt. konf. s mezhd. uchastiem, g. Nizhnii Novgorod, 22–23 maya 2024 g*. [National and international trends and prospects for the development of forensic examination: collection of reports of the scientific and practical conference with international participation, Nizhny Novgorod, May 22–23, 2024]. Nizhny Novgorod, 2024. Pp. 42–50. (In Russ.).
13. *EDIC-mini — diktofony dlya vashei bezopasnosti* [EDIC-mini voice recorders for your safety]. Available at: https://www.telesys.ru/Products/EM (accessed September 16, 2024).
14. *Ethics*. Available at: https://hourone.ai/ethics/ (accessed September 16, 2024).
15. Introducing official content credentials icon. Available at: https://c2pa.org/post/contentcredentials/ (accessed September 16, 2024).
16. *Content credentials*. Available at: https://contentcredentials.org/verify (accessed September 16, 2024).
17 *Watermarks in preview in Azure OpenAI Service*. Available at: https://techcommunity.microsoft.com/t5/ai-azure-ai-services-blog/watermarks-in-preview-in-azure-openai-service/bap/4253344 (accessed September 26, 2024).
18. Azure AI Speech. Available at: https://azure.microsoft.com/en-us/products/ai-services/ai-speech?msockid=04b5abc717656a530541bfde16f56b4b (accessed September 16, 2024).
19. *DeepMind.* Available at: https://deepmind.google/technologies/synthid/ (accessed September 26, 2024).
20. *Resemble.ai.* Available at: https://www,.resemble.ai/watermarker/ (accessed September 20, 2024).
21. *Myvocal.ai.* Available at: https://myvocal.ai/billing (accessed September 20, 2024).
22. YouTube Blog – Official Blog for Latest YouTube News&Insights. Available at: https://blog.youtube. (accessed September 20, 2024).
23. *Cloud Level.* Available at: https://about.fb.com/news/ (accessed September 20, 2024).
24. Roman R.S., Fernandez P., Défossez A., Furon T. et al. *Proactive detection of voice cloning with localized watermarking*. Available at: https://arxiv.org/abs/2401.17264 (accessed September 20, 2024).
25. *Deep fake challenge*. Available at: https://deepfakechallenge.com/ (accessed September 20, 2024).
26. *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206* – C9-0146/2021 – *2021/0106(COD))*. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html (accessed September 6, 2024).
27. [Departmental rule "Regulation on the Administration of the Deep Information Synthesis Service on the Internet"]. Available at: https://www.cac.gov.cn/2022-12/11/c_1672221949318230.htm (accessed September 6, 2024).
28. *H.R.3831 – AI Disclosure Act of 2023*. Available at: https://www.congress.gov/bill/118th-congress/house-bill/3831/text?s=1&r=1 (accessed September 6, 2024).
29. *Governing AI for humanity: final report.* Available at: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_ru.pdf (accessed September 30, 2024).

INFORMATION ABOUT THE AUTHORS

NIKOLAI F. BODROV – Candidate of Sciences (Law), President of the Union of Criminalists and Criminologists, Moscow, Russia, associate professor at the Forensic Expertise Department of the Kutafin Moscow State Law University (MSAL), Moscow, Russia, bodrovnf@gmail.com, https://orcid.org/0000-0002-9005-3821

ANTONINA K. LEBEDEVA – Candidate of Sciences (Law), Member of the Union of Criminalists and Criminologists, Moscow, Russia, associate professor at the Forensic Expertise Department of the Kutafin Moscow State Law University (MSAL), Moscow, Russia, tonya109@yandex.ru, https://orcid.org/0009-0004-9344-2103

Received October 10, 2024