https://doi.org/10.12697/JI.2020.29.07

Ilya Ilin
Doctoral student
University of Tartu

Aleksei Kelli
Professor of Intellectual Property Law
University of Tartu

The Use of Human Voice and Speech for Development of Language Technologies:

The EU and Russian Data-protection Law Perspectives

1. Introduction

Language technologies[*1] (LTs) have become part of our day-to-day life. Their applications range from services for automatic text translation and spelling- and grammar-checkers to speech-to-speech translators[*2] and applications synthesising the human voice.

The development of LTs does not rely merely on text on a page. It encompasses using the human voice and speech also. Here, 'voice' refers to the process of acoustic waves' creation and 'speech' is the process of phoneme creation.[*3] In a narrow sense, it is possible to regard the human voice as a tool that is used to create speech (the speech vocalisation element).

The voice and speech are crucial elements of the communication process. Communication by voice is the most convenient and the fastest means of interaction between people and also between humans and

[1] Language technologies rely on the use of language resources, where language resources are characterised as copyright-protected databases that may contain copyright-protected works, performances protected as objects of related rights, and personal data. For reasons of space, the authors do not address technical issues such as the relationship between language resources and technologies in this article. Neither is this necessary for the analysis of legal issues related to personal-data protection. For further discussion of the nature of language resources, see: Aleksei Kelli, Krister Lindén, Kadri Vider, Penny Labropoulou, Erik Ketzan, Pawel Kamocki, Pawel Straňák, and Maciej Piasecki, 'Implementation of an Open Science Policy in the Context of Management of CLARIN Language Resources: A Need for Changes?' in *Selected Papers from the CLARIN Annual Conference 2017* (Linköping University Electronic Press / Linköping Electronic Conference Proceedings, Linköpings Universitet 2018) 102–111. https://www.ep.liu.se/ecp/147/009/ecp17147009.pdf; Aleksei Kelli, Kadri Vider, and Krister Lindén, 'The Regulatory and Contractual Framework As an Integral Part of the CLARIN Infrastructure' in Koenraad De Smedt (ed), *Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wroclaw, Poland* (Linköping University Electronic Press, Linköpings Universitet 2015) 13–24. https://www.ep.liu.se/ecp/article.asp?issue=123&article=002; Aleksei Kelli, Kadri Vider, Heiki Pisuke, and Triin Siil, 'Constitutional Values As a Basis for the Limitation of Copyright within the Context of Digitalization of the Estonian Language' in Kalvis Torgans (ed), *Constitutional Values in Contemporary Legal Space II: Collection of Research Papers in Conjunction with the 6th International Scientific Conference of the Faculty of Law of the University of Latvia* (University of Latvia Press 2017) 126–139. DOI: https://doi.org/10.22364/cvcls.2.2016.

[2] In October 2017, Google demonstrated the brand-new headphones known as Pixel Buds, which have an integrated speech-to-speech translation function: Adam Champy, 'Google Pixel Buds – Wireless Headphones That Help You Do More' *Google Blog* (4 October 2017). https://www.blog.google/products/pixel/pixel-buds/.

[3] Alison Behrman, *Speech and Voice Science* (Plural Publishing 2017) 4.

computers. It is much easier to input large volumes of data, utilise a control system, and thereby create a dialogue via voice rather than through other methods of communication.*[4]

Today, more and more products and services are based on LTs that use voice and speech. The practical utilisation of the voice and speech in an LT can be divided into four categories: speech synthesis*[5], voice biometrics*[6], speech analysis*[7], and speech recognition.*[8]

LTs are seldom focused on one particular country. They are disseminated through multiple jurisdictions. Several of the speech-recognition systems now in use are actively distributed by global digital companies (e.g., the Google Cloud speech API or Yandex SpeechKit), and they can be integrated easily into any program, app, or service, developed nearly anywhere in the world. For example, such speech-recognition systems form the core elements of the following products: virtual 'voice assistants' (e.g., Siri*[9], Cortana*[10], Alexa*[11], and Alisa*[12]), Intensive Voice Response (IVR) systems, and vehicular voice-control systems (as used by Tesla, BMW, Ford, and Mercedes–Benz).

To consider the global character of research and business related to LTs, the producers of such technologies need to comply with the relevant regulation, which includes data-protection regulation. The aim for this article is to delineate, evaluate, and compare the legal frameworks for the use of voice and speech in development and dissemination of LTs from the perspectives of EU and Russian data-protection law. Some references to the Estonian legal landscape for data protection*[13] are made also, where there is a need to consult the data-protection rules of a specific EU country. Firstly, Estonian law has been chosen since the authors are familiar with it. Secondly, the EU's data-protection rules leave the Member States considerable flexibility to choose from among various harmonisation and implementation models.

The foundation of data-protection law is the same for Europe and Russia: the European Convention on Human Rights (ECHR)*[14] and the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108).*[15] Since the international framework is limited to the essential principles, it does not extensively harmonise data-protection laws. Therefore, the EU and Russian national laws possess distinctive elements and even conflict with each other in some respects. Such differences in legislation create legal challenges for technology companies that wish to provide their services in Europe and Russia.

[4] Wendy Holmes, *Speech Synthesis and Recognition* (2nd edn, CRC Press 2001) 1. DOI: https://doi.org/10.4324/9780203484685.

[5] Speech synthesis is a technology that converts the text to speech. See: Thierry Dutoit, *An Introduction to Text-to-Speech Synthesis* (Springer Science & Business Media 1997, vol 3) 1. DOI: https://doi.org/10.1007/978-94-011-5730-8.

[6] The voice can be considered to be one of the unique characteristics of the personality that may be used to establish an identity, alongside fingerprints, DNA, and the face or facial geometry. See: Anil Kumar Jain, Arun Ross, and Salil Prabhakar, 'An Introduction to Biometric Recognition' (2004) 14.1 *IEEE Transactions on Circuits and Systems for Video Technology* 4, 5. DOI: https://doi.org/10.1109/TCSVT.2003.818349.

[7] The human voice can provide large amounts of useful information about a person's mental state – for instance, mood, emotional condition, stress level, and any lack of sleep. See: Keng-hao Chang, Drew Fisher, and John Canny, 'AMMON: A Speech Analysis Library for Analyzing Affect, Stress, and Mental Health on Mobile Phones' in *Proceedings of PhoneSense* 2011 (2011). http://people.eecs.berkeley.edu/~jfc/papers/11/AMMON_phonesense.pdf (accessed 10 April 2020).

[8] Speech-recognition technology is a process of automatic speech-to-text transcription. See: Alexander Clark, Chris Fox, and Shalom Lappin (eds), *The Handbook of Computational Linguistics and Natural Language Processing* (John Wiley & Sons 2013) 299.

[9] Speech Interpretation and Recognition Interface, developed by Apple, Inc. Information is available at: https://www.apple.com/siri/ (accessed 10 April 2020).

[10] Voice assistance developed by Microsoft, Inc. See: https://www.microsoft.com/en-us/cortana (accessed 10 April 2020).

[11] Voice assistance developed by Alexa Internet, Inc., a company owned by Amazon, Inc. For information, see: https://www.amazon.com/meet-alexa/b?ie=UTF8&node=16067214011 (accessed 10 April 2020).

[12] Voice assistance developed by Yandex, Inc. Information available in Russian at: https://alice.yandex.ru/ (accessed 10 April 2020).

[13] The Estonian Personal Data Protection Act (Isikuandmete kaitse seadus). Entry into force on 15 January 2019. English translation available at: https://www.riigiteataja.ee/en/eli/523012019001/consolide (accessed 18 June 2020).

[14] Article 8 of: Convention for the Protection of Human Rights and Fundamental Freedoms, ETS No.005, 'Treaty open for signature by the member States of the Council of Europe and for accession by the European Union at Rome' on 4 November 1950 with entry into force on 3 September 1953. https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005 (accessed 10 April 2020).

[15] Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, ETS No.108, 'Treaty open for signature by the member States of the Council of Europe and for accession by the European Union at Strasbourg' on 28 January 1981 with entry into force on 1 October 1985. https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108 (accessed 10 April 2020).

The European data-protection framework is established primarily by the General Data Protection Regulation (GDPR)*[16], which is directly applicable*[17] in all EU member states.*[18] Russian data-protection law relies on the following acts: Federal Law 'On Personal Data'*[19], Federal Law 'On Information, Information Technologies and Information Protection'*[20], and the 'Yarovaya package law'*[21]. This list is not exhaustive. There are also legal acts that do not directly refer to the realm of data protection but do contain separate legal rules affecting the data-protection domain (e.g., Federal Law 'On Communications'*[22], from 2003). On account of the scope for the research presented here and the complexity of Russia's data-protection law, these acts are not the main focus of the article.

The choice of jurisdictions for examination here is based on consideration of the fact that the EU and Russia are neighbours and in a globalised world such as ours, it is not possible or even reasonable to avoid co-operation across the jurisdictions in technology development. The authors' ambition in this regard is limited to addressing co-operation within the framework of LTs, with emphasis on data protection. The research holds further relevance in that extensive comparative analysis of the Russian data-protection laws (significantly amended in 2015*[23] and 2017*[24]) and the General Data Protection Regulation*[25] with regard

[16] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), dated 27 April 2016, with entry into force on 25 May 2018. https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed 10 April 2020).

[17] However, the GDPR (*ibid*) allows derogation from the regulation in certain fields, such as research; see its Article 89. It is also relevant with regard to the development of LTs. For reasons of space and the focus of this article, this derogation is not addressed.

[18] The GDPR's territorial scope is not limited to the EU states alone. It applies also to the European Economic Area (EEA) countries and in certain circumstances to non-EU-, non-EEA-based companies. The territorial scope of the GDPR is described further on, in section 3 of the paper.

[19] Федеральный закон «О персональных данных» (Federal Law 'On Personal Data') N 152-FZ, dated 27 July 2006, adopted by the State Duma on 8 July 2006, approved by the Federation Council on 14 July 2006, with entry into force on 26 January 2007. Unofficial English translation available at: https://pd.rkn.gov.ru/authority/p146/p164/. All translations from Russian into English are by the authors of the present paper unless otherwise noted.

[20] Федеральный закон «Об информации, информационных технологиях и о защите информации» (Federal Law 'On Information, Information Technologies and Protection of Information') N 149-FZ, dated 27 July 2006, adopted by the State Duma on 8 July 2006, approved by the Federation Council on 14 July 2006, with entry into force on 26 January 2007. Unofficial English translation available at: http://www.wipo.int/wipolex/ru/details.jsp?id=15688 (accessed 10 April 2020).

[21] Its unofficial name, after one of the authors of the law, Irina Yarovaya. The package consists of the two pieces of Federal Law legislation that introduce amendments to the acts on combating terrorism: (i) Федеральный закон «О внесении изменений в Федеральный закон «О противодействии терроризму» и отдельные законодательные акты Российской Федерации в части установления дополнительных мер противодействия терроризму и обеспечения общественной безопасности» (Federal Law 'On Amendments to the Federal Law "On Counteracting Terrorism" and Certain Legislative Acts of the Russian Federation Regarding the Establishment of Additional Measures To Counter Terrorism and Ensure Public Safety') N 374-FZ, dated 6 July 2016, adopted by the State Duma on 24 June 2016, approved by the Federation Council on 29 June 2016, with entry into force on 20 July 2016. Available in Russian at: http://kremlin.ru/acts/bank/41108; (ii) Федеральный закон «О внесении изменений в Уголовный кодекс Российской Федерации и Уголовно-процессуальный кодекс Российской Федерации в части установления дополнительных мер противодействия терроризму и обеспечения общественной безопасности» (Federal Law 'On Amendments to the Criminal Code of the Russian Federation and the Code of Criminal Procedure of the Russian Federation with Regard to the Establishment of Additional Measures To Counter Terrorism and Ensure Public Safety') N 375-FZ, dated 6 July 2016, adopted by the State Duma on 24 June 2016, approved by the Federation Council on 29 June 2016, with entry into force on 20 July 2016. Available in Russian at: http://kremlin.ru/acts/bank/41113 (accessed 10 April 2020).

[22] Федеральный закон «О связи» (Federal Law 'On Communications') N 126-FZ, dated 7 July 2003, adopted by the State Duma on 18 June 2003, approved by the Federation Council on 25 June 2003, with entry into force on 1 January 2004. Unofficial English translation available at: http://www.wipo.int/wipolex/en/details.jsp?id=17111.

[23] Федеральный закон "О внесении изменений в отдельные законодательные акты Российской Федерации в части уточнения порядка обработки персональных данных в информационно-телекоммуникационных сетях" (Federal Law 'On Amendments to Certain Legislative Acts of the Russian Federation Regarding the Clarification of the Procedure for Processing Personal Data in Information and Telecommunication Networks') N 242-FZ, dated 21 July 2014, adopted by the State Duma on 4 July 2014, approved by the Federation Council on 9 July 2014, with entry into force on 1 September 2015. Available in Russian at: http://www.consultant.ru/document/cons_doc_LAW_165838/ (accessed 10 April 2020).

[24] Федеральный закон "О внесении изменений в Кодекс Российской Федерации об административных правонарушениях" (Federal Law 'On Amendments to the Code of Administrative Offences of the Russian Federation') N 13-FZ, dated 7 February 2017, adopted by the State Duma on 27 January 2017, approved by the Federation Council on 1 February 2017, with entry into force on 1 July 2018. Available in Russian at: http://www.consultant.ru/document/cons_doc_LAW_212391/ (accessed 12 April 2020).

[25] General Data Protection Regulation (n 16).

to the LT field has not been undertaken before.*[26] The article could also be useful to LT researchers and entrepreneurs who want to cover both the EU and Russia in their studies or products/services. The research results serve as a basis for further investigation pertaining to the personal-data aspects of several jurisdictions' law.

The authors draw on prior research*[27] while relying also on personal experience in the field of legal aspects of LTs. The article broadens the focus of LT-related legal research from that previously established, so as to include Russian data-protection law as well.

The second section of the article addresses the legal nature of human voice and speech from the data protection law perspective. In the third part, the applicability of the EU and Russian data-protection legislation form the LTs perspective is analysed. Under the last section, the principles and rules for voice- and speech-processing are studied.

2. Human voice and speech as personal data

The question of whether human voice and speech should be treated as personal data influences the requirements imposed on development of LTs. Therefore, the authors address particular aspects of the human voice and speech accordingly (see Figure 1). The first of these involves the subject matter of the speech and its content (speech can contain personal data), the second involves the voice as personal data, and the third is related to the question of whether voice belongs to a special category of data that entails additional requirements for its processing (use). The voice is examined without a strong connection to the speech content.

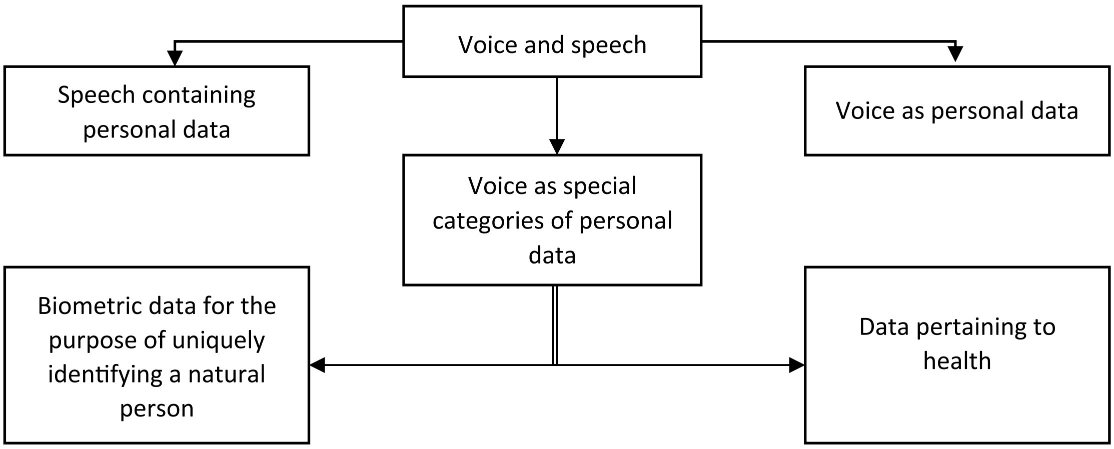
Voice and speech
Speech containing personal data
Voice as personal data
Voice as special categories of personal data
Biometric data for the purpose of uniquely identifying a natural person
Data pertaining to health

Figure 1: Voice and speech from a data-protection perspective

[26] As a matter of fact, even analysis of the impact of the GDPR on the development of language technologies in Europe remains at quite a preliminary level.

[27] See Jane Klavan, Arvi Tavast, and Aleksei Kelli, 'The Legal Aspects of Using Data from Linguistic Experiments for Creating Language Resources' (2018) 307 *Frontiers in Artificial Intelligence and Applications* 71. http://ebooks.iospress.nl/volumearticle/50306; Aleksei Kelli, Kadri Vider, Irene Kull, Triin Siil, Krister Lindén, Arvi Tavast, Age Värv, Carri Ginter, and Einar Meister, 'Keeleressursside loomise ja kasutamisega seonduvaid isikuandmete kaitse küsimusi' (Data Protection Issues Related to the Development and Utilisation of Language Resources) *Eesti Rakenduslingvistika Ühingu aastaraamat* (2018) 14, 77–94. DOI: https://doi.org/10.5128/erya14.05; Liina Jents and Aleksei Kelli, 'Legal Aspects of Processing Personal Data in Development and Use of Digital Language Resources: The Estonian Perspective' (2014) 21.1 *Jurisprudencija* 164. DOI: https://doi.org/10.13165/jur-14-21-1-08.

We begin by considering the facets of speech. Data-protection laws apply if the speech contains personal data. Both European and Russian legal regulations define personal data as information related to an identified or identifiable natural person (the 'data subject').*[28]

The GDPR makes references to various types of personal data (e.g., biometric, genetic, and health data)*[29]; however, the most fundamental line is drawn between the concept of personal data in general and personal data falling in special categories. According to the GDPR, special categories of personal data consist of 'data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation'.*[30] The latter is subject to more stringent requirements.*[31]

The Russian data-protection regulation, in turn, defines three main categories of personal data: general, special, and biometric personal data. Some of the legal acts specify a fourth category of personal data, 'publicly available personal data'*[32]. However, Russia's Federal Law 'On Personal Data' does not classify this as a separate and independent category. The 'special' category of personal data under these laws includes data pertaining to a person's racial or ethnic origin, political opinions, religious or philosophical beliefs, health, or sex life*[33]. The 'biometric data' category covers data related to a person's physiological and biological characteristics that are used for identification purposes*[34] (e.g., fingerprints, DNA, voice, the person's image, the iris portion of the eyes, and/or body structure*[35]).

The three-category division among general, biometric, and special personal data is of fundamental importance in cases of data-processing. For instance, under the general rule, the processing of special-category data is prohibited*[36], while processing of biometric data may be performed, albeit only with the explicit consent of the data subject*[37]. It is important to distinguish data in the special category from the biometric class also because the **level** of protection required is different*[38].

The information space considered to contain personal data is rather extensive. According to the Article 29 Working Party*[39] (WP29), the concept of personal data covers information available in any of various forms (graphical, photographic, acoustic, alphanumeric and so forth) and maintained in storage of numerous types (e.g., on videotape, on paper, or in computer memory).*[40]

According to Russian law, general-category personal data*[41] include such data as the name (surname, patronymic, etc.); the year, month, day, and place of birth; one's address; the identity of one's family; social

[28] Article 4 of the General Data Protection Regulation (n 16). See also Federal Law 'On Personal Data' N 152-FZ (n 19) art 3 (1).

[29] Article 4 of the General Data Protection Regulation (n 16).

[30] Article 9(1) of the General Data Protection Regulation (n 16).

[31] The general rule is that the processing of special categories of personal data is prohibited unless certain circumstances exist, per Article 9 of the General Data Protection Regulation (n 16).

[32] Clause 5 of Decree of the Government of the Russian Federation No. 1119, 'On approval of the requirements for the protection of personal data when processing them in information systems of personal data'.

[33] Federal Law 'On Personal Data' N 152-FZ (n 19) art 10.

[34] *Ibid*, art 11.

[35] Разъяснения Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций «Разъяснения по вопросам отнесения фото-, видеоизображений, дактилоскопических данных и иной информации к биометрическим персональным данным и особенностей их обработки» (Explanations on the Issues of Attributing Photo, Video, Fingerprint Data, and Other Information to Biometric Personal Data and the Features of Their Processing) issued by the *Roskomnadzor* on 30 August 2013. http://www.garant.ru/products/ipo/prime/doc/70342932/ (accessed 12 April 2020).

[36] Federal Law 'On Personal Data' N 152-FZ (n 19) art 10.

[37] *Ibid*, art 11.

[38] Maxim Krivogin, 'Osobennosti pravovogo regulirovaniya biometrichecskih personalnyh dannyh' (Peculiarities of Legal Regulation of Biometric Personal Data) [2017] 2 *Journal of the Higher School of Economics* 80, 82–83. DOI: https://doi.org/10.17323/2072-8166.2017.2.80.89.

[39] The Article 29 Working Party is an advisory committee established via the Data Protection Directive (95/46/EC) (repealed as of 25 May 2018). Its opinions are still relevant since the nature of personal data's protection has not changed.

[40] See page 7 of Article 29 Working Party Opinion 4/2007, adopted on 20 June 2007, on the concept of personal data: http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (accessed 12 April 2020).

[41] Federal Law 'On Personal Data' N 152-FZ (n 19) art 3 (1).

or property status; education, profession, or income[*42]; passport data[*43]; e-mail address[*44]; and information on crossing of state borders[*45].

Another relevant and often misinterpreted issue is the protectability of publicly available personal data. The matter has been settled in EU case law. The European Court of Justice has explained that the use of data collected from documents in the public domain is still processing of personal data.[*46] The public availability of personal data has relevance in the context of processing of special categories of personal data, with the rule being that processing of data in the special categories is prohibited.[*47] However, this prohibition does not apply if the processing involves personal data that have been manifestly made public by the data subject.[*48]

The Russian data-protection laws provide that the personal data in question should be considered publicly available if the data subject gives explicit consent[*49] for inclusion of the data in the relevant publicly accessible sources[*50]. The publicly available data still are subject to the data-protection regulation[*51], but the threshold level of protection is much lower for data in this category than for other categories of personal data. For instance, there is no need to obtain consent for processing[*52] or to ensure a confidentiality regime[*53] for general- and special-category personal data in this case (consent need only be received once for making the data publicly available). At the same time, the rule explicitly does not extend to publicly available biometric data, whose processing still requires the consent of the data subject.

From a language-technology perspective, it is not so relevant when precisely the data subject's rights arise. However, when they end is crucial.[*54] The GDPR does not apply to personal data of deceased persons.[*55] That said, variations may exist in national legislation, creating differences between EU countries in such respects. Therefore, it is important to consult the laws of each specific EU country that is relevant. For instance, under the Estonian Personal Data Protection Act, the protection of rights extends 10 years after the death of the data subject except in cases wherein the data subject died as a minor, for which the term of protection is 20 years. Any heir may give consent for processing.[*56] Other Member States may take different approaches.

Russian data-protection regulation extends to the personal data of deceased persons[*57]. The processing of such data must comply with data-protection rules (including the requirement to gain consent for the processing).[*58] Russia's data-protection law does not specify a duration for the protection of personal data

[42] For relevant case law, see: Presidium of the Russian Supreme Arbitration Court, resolution in case N. A36-5713 / 2014, dated 29 April 2015. https://kad.arbitr.ru/Document/Pdf/21af41bd-86ed-4551-b372-10bb6499cf3d/635e0e09-c758-4f90-85d4-e327555d3daf/A36-5713-2014_20151228_Reshenija_i_postanovlenija.pdf?isAddStamp=True (accessed 12 April 2020).

[43] See the case law: Appeal Definition of the Moscow City Court N 33-14709 / 2014, dated 22 May 2014. https://mos-gorsud.ru/mgs/cases/docs/content/631f39ab-1e88-4428-9ece-d53bdc6b6670 (accessed 12 April 2020).

[44] Consult the case law from: Kalininsky District Court (St Petersburg, Russia) Decision N 12-253 / 2015, dated 26 May 2015. https://kln--spb.sudrf.ru/modules.php?name=sud_delo&name_op=sf&delo_id=1540005 (accessed 12 April 2020).

[45] See case law: Moscow City Court, Appeal Definition N 33-11688 / 2014, dated 10 April 2014. https://mos-gorsud.ru/mgs/cases/docs/content/747dfe19-f877-4121-838c-5f30ab010ebf (accessed 12 April 2020).

[46] Relevant case law is: *Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy and Satamedia Oy*. [2008] C-73/07, from 16 December 2008. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1538028665554&uri=CELEX:62007CJ0073 (accessed 12 April 2020).

[47] General Data Protection Regulation (n 16) art 9(1).

[48] General Data Protection Regulation (n 16) art 9(2) (e).

[49] The data subject has a right to withdraw consent, per Article 8 (2) of Federal Law 'On Personal Data' N 152-FZ (n 19).

[50] *Ibid*, art 8 (1).

[51] *Ibid*, art 6 (1).

[52] *Ibid*, art 6 (1).

[53] *Ibid*, art 7.

[54] For instance, the Russian voice company STC Group demonstrated a novel vocalised by the synthesised voice of a dead Russian actor: «Синтез речи. Беглый обзор» (Synthesis of Speech: A Brief Review) *Stokito on Software Blog* (25 December 2014). goo.gl/Kmcpgh/. An example of the synthesised voice is available at: https://www.youtube.com/watch?v=hvaB1exK9rY (accessed 12 April 2020).

[55] Recital 27 to the General Data Protection Regulation (n 16).

[56] The Estonian Personal Data Protection Act (n 13) s 9.

[57] In the case law, see: Decree of the Federal Arbitration Court of the Eastern Siberian District N A33-14182/2007, dated 1 July 2008. https://kad.arbitr.ru/ (accessed 12 April 2020).

[58] Where a personal-data subject has died, any consent to the processing of his personal data shall be given by his heirs unless the personal-data subject gave such consent while alive. This is addressed in: Federal Law 'On Personal Data' N 152-FZ (n 19) art 9 (7).

of deceased persons. One solution is to rely on an analogy to protection of a person's private life*[59], which is likewise protected after the person's death.*[60] Following this analogy, we could presume that the duration of such protection extends to at least 75 years after the death of the data subject*[61].

The identifiability of a natural person is a critical issue in determination of whether data-protection laws apply. The authors agree with the WP29 reasoning that 'a mere hypothetical possibility to single out the individual is not enough to consider the person as "identifiable"'.*[62]

It is also pointed out in the literature that identifiability depends on the context. Data items not identifying for one person might be identifying for another.*[63] It is also suggested that 'the categorisation of data as identifiable or non-identifiable is a matter of self-assessment by the controller; the controller determines how the data are to be categorised and treated'.*[64] This does not, however, mean that the data are in reality non-personal. The controller cannot avoid liability just by considering all the data processed non-personal.

The use of non-personal data is less subject to legal restrictions.*[65] Data may be non-personal from day 1*[66], or personal data may be anonymised and thereby rendered non-personal. With regard to the latter, one should bear in mind that the definition of personal data's processing is quite broad in the GDPR*[67] and Russian law alike*[68]. Accordingly, the anonymisation process itself is subject to personal-data protection requirements. Secondly, creating entirely anonymised datasets such that the data do not lose their value is a challenging task.*[69] This is especially true for voice and speech.

[59] Mariya Vazhorova, "Соотношение понятий «Информации о частной жизни» и «Персональных данных» " (The relationship between the Concepts of 'Information on Private Life' and 'Personal Data') M Vazharova, tr (2012) 4(87) *Bulletin of the Saratov State Law Academy* 55, 55–56.

[60] See Article 152.2(5) of: Гражданский кодекс Российской Федерации (часть первая) (The Civil Code of the Russian Federation (Part I of IV)) N 51-FZ, dated 30 November 1994, adopted by the State Duma on 21 October 1994, signed by the President of the Russian Federation on 30 November 1994, with entry into force on 1 January 1995. Unofficial English translation available at: http://www.wipo.int/edocs/lexdocs/laws/en/ru/ru083en.pdf (accessed 13 April 2020).

[61] See Article 25(3) of: Федеральный закон «Об архивном деле в Российской Федерации» (Federal Law 'On Archival Affairs in the Russian Federation') N 125-FZ, dated 22 October 2004, adopted by the State Duma on 1 October 2004, approved by the Federation Council on 13 October 2004, with entry into force on 27 November 2004. Available in Russian at: https://rg.ru/2004/10/27/arhiv-dok.html. The law provides for restriction of access to archival documents containing information about the personal and family secrets of a citizen or his private life or including information that creates a threat to his safety, with a set term of 75 years from the date of the creation of these documents.

[62] See page 15 of: Article 29 Working Party Opinion 4/2007, on the concept of personal data, as adopted on 20 June 2007. http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (accessed 13 April 2020).

[63] Manon Oostveen, 'Identifiability and the Applicability of Data Protection to Big Data' (2016) 6(4) *International Data Privacy Law* 299, 306.

[64] *Ibid*, 307.

[65] According to the General Data Protection Regulation (n 16), '[t]he principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable' (Recital 26).

[66] For instance, machine-generated data not containing personal information are not subject to personal-data protection. However, this does not mean that said data are not subject to some currently recognised rights (in the main, database rights and trade-secret protection) or future legal requirements. For further discussion, see, from 10 January 2017: Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Building a European Data Economy' COM (2017) 9 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1537175097689&uri=CELEX:52017DC0009 (accessed 13 April 2020). See also: P Bernt Hugenholtz, 'Data Property: Unwelcome Guest in the House of IP' (2017). https://www.ivir.nl/publicaties/download/Data_property_Muenster.pdf (accessed 13 April 2020).

[67] Article 4 of the General Data Protection Regulation (n 16) defines processing as 'any operation or set of operations which is performed on personal data or on sets of personal data'. The Article 29 Working Party is of the opinion that '[a]nonymisation constitutes a further processing of personal data; as such, it must satisfy the requirement of compatibility by having regard to the legal grounds and circumstances of the further processing'. See page 3 of: Article 29 Working Party Opinion 05/2014, on anonymisation techniques, adopted on 10 April 2014. http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (accessed 13 April 2020).

[68] Federal Law 'On Personal Data' N 152-FZ (n 19) art 3 (3) states that personal data's processing may consist of any action (operation) or combination of actions (operations) performed both automatically and manually with personal data, including collection, recording, arrangement, accumulation, storage, specification (updating or other changing), extraction, use, distribution (including transfer), anonymising, blocking, and destruction of personal data.

[69] See (n 67).

The next two aspects to be considered pertain to the human voice as such. In scientific literature, the voice is considered biometric data.*[70] Both jurisdictions considered here distinguish biometric data from the other categories of personal data. They define said data as data about physical, physiological, or behavioural characteristics of a natural person.*[71] Most commonly, LTs use voice and speech as biometric data for two purposes: 1) to identify and verify a person (voice biometrics) and 2) to acquire and analyse new information about a person (voice and speech analysis).*[72]

From the biometrics perspective, the voice samples (or voiceprints) are used to identify and verify who someone is in a similar manner to DNA, fingerprints, or face recognition.*[73] Depending on the operation mode of the biometric system, the voiceprint may be compared with one particular voiceprint to verify the claimed identity (verification mode) or the system may scan a database of voiceprints to find the matching one and thereby establish the speaker's identity (identification mode).*[74] The voice samples (biometric personal data) within voice-biometrics frameworks are often used in combination with other categories of personal data.

From a speech-analysis perspective, voice and speech patterns can be investigated for purposes of obtaining additional information about the person speaking. For example, voice and speech analysis can be used in medical applications*[75] for its ability to provide information about stress levels, emotional state*[76], or other health details of the person. In the case of detecting mental state, one's level of stress, and other medical information, the data received can be considered to be, in addition, information pertaining to the person's health.

Although the human voice contains biometric information and potentially health-related data, the crucial issue in this regard is whether this means that the voice as such always belongs to a special category of data. The GDPR's definition of special categories of data*[77] refers to two instances of processing wherein the voice can be deemed to belong to special categories: 1) the voice as health data and 2) the voice as biometric data for the identification of a natural person.

If we presume that the voice *per se* (even without any relevant content) always contains health-related information (which is disputable), then it would be regarded as a special category of personal data both in the EU and per Russian data-protection law.*[78] However, a question remains as to what kind of information should be considered health-related and how much of the health-related information can be extracted from the voice.

Russia's data-protection regulation does not provide a definition addressing precisely what information is connected with information pertaining to health. At the same time, Russian regulation of health protection includes the concept of medical secrecy, under which information about requests for medical assistance, information about health and diagnoses or other information received during medical examinations and treatment constitutes a medical secret.*[79] Data with 'medical secret' status receive special legal protection, and the processing and disclosure thereof are prohibited, with certain specified exceptions.*[80] The concept

[70] See discussion by: Joaquín González-Rodríguez, Doroteo Torre Toledano, and Javier Ortega-García, 'Voice Biometrics' in Anil K Jain, Patrick Flynn, and Arun A Ross (eds), *Handbook of Biometrics* (Springer Science & Business Media, 2007). DOI: https://doi.org/10.1007/978-0-387-71041-9_8; Anil Kumar Jain, Arun Ross, and Salil Prabhakar, 'An Introduction to Biometric Recognition' (2004) 14(1) *IEEE Transactions on Circuits and Systems for Video Technology*. https://www.cse.msu.edu/~rossarun/BiometricsTextBook/Papers/Introduction/JainRossPrabhakar_BiometricIntro_CSVT04.pdf (accessed 13 April 2020).

[71] Article 4 of the General Data Protection Regulation (n 16); Federal Law 'On Personal Data' N 152-FZ (n 19) art 11.

[72] Judith Markowitz, 'Voice Biometrics' (2000) 43.9 *Communications of the ACM* 66.

[73] Anil Kumar Jain, Ross Arun, and Salil Prabhakar, 'An Introduction to Biometric Recognition' (2004) 14.1 *IEEE Transactions on Circuits and Systems for Video Technology* 4.

[74] Hariton Costin, Tudor Barbu, Cristi Rotariu, and Iulian B Ciocoiu, 'A Complex Biometric System for Person Verification and Identification through Face, Fingerprint and Voice Recognition' [2006] *Scientific Studies and Research* 361.

[75] (n 7).

[76] Ryan Hafen and Henry Michael, 'Speech Information Retrieval: A Review' (2012) 18.6 *Multimedia Systems* 499.

[77] Article 9(1) of the General Data Protection Regulation (n 16).

[78] (*ibid*); Federal Law 'On Personal Data' N 152-FZ (n 19) art 10.

[79] See Article 13 of: Федеральный закон «Об основах охраны здоровья граждан в Российской Федерации» (Federal Law 'On the Fundamentals of Protecting the Health of Citizens in the Russian Federation') N 323-FZ, dated 21 November 2011, adopted by the State Duma on 1 November 2011, approved by the Federation Council on 9 November 2011, with entry into force on 22 November 2011. An English translation is not available, but the law is available in Russian at: http://kremlin.ru/acts/bank/34333 (accessed 13 April 2020).

[80] Federal Law 'On Personal Data' N 152-FZ (n 19) art 3 (9).

of medical secrecy is associated primarily with medical assistance requests and provision of medical treatment. The information forming a medical secret is a subset of what is deemed to be personal data having to do with health.

In contrast, European data-protection regulation does define the boundaries of information pertaining to health.*[81] According to the GDPR, the information related to health is the personal data that refer to the physical and mental state of the person, along with information about the provision of medical services and related information about health status.*[82] Health-related data are subject to special regulation and protection.

In the authors' opinion, the voice does not always contain health data. Not all television and radio programmes, interview content, etc. should be considered to belong to a special category of personal data. In cases wherein the voice processing is done for collecting data about health , however, it does belong to a special class of personal data, accordingly.

There is no disputing that the voice as such is biometric data. The question is whether this leads to it counting as a special category of data. According to the GDPR, only biometric data used for uniquely identifying a natural person belong to a special category of data.*[83] In other words, it is insufficient to deem the voice biometric data without further consideration. Rather, for it to qualify as a special category of data, the voice processing must be done for identification purposes. In this case, the data processing determines its nature. The situation is similar to that of photos depicting people – after all, one's appearance constitutes biometric data. For the latter case, the GDPR provides the following clarification:

> The processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person.*[84]

The authors of this article presume that the foregoing explanation is valid also for the human voice.

Russian data-protection law treats information about physiological and biological characteristics as biometric data only if the operator*[85] uses it for purposes of identification*[86]. The identification purpose behind the data-processing is the critical criterion for identifying the given personal data as biometric personal data*[87]. In a similarity to the EU approach, the voice is not deemed biometric data in the context of data protection unless it is used for identification purposes.

Whether the voice and speech are considered to be personal data plays a crucial role in the processing and in compliance with the data-protection rules. There is commonality between the European and the Russian approach to personal data and the categories thereof in that technology companies are required to treat information such as voiceprints, health information, and other subject data as personal data and to comply with domestic data-protection regulations on that basis. In the following section, the two regulation systems are analysed and compared. The voice and speech are examined as both non-sensitive, general personal data and personal data belonging to a special category of personal data (biometric data or data pertaining to health).

[81] Even in early jurisprudence of the European Court of Justice, 'health-related data' is accorded extensive scope. For instance, the European Court of Justice has found that 'reference to the fact that an individual has injured her foot and is on half-time on medical grounds constitutes personal data concerning health'. Case C-101/01, criminal proceedings against Bodil Lindqvist (6 November 2003). http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1521039149443&uri=CELEX:62001CJ0101 (accessed 13 April 2020).

[82] See Article 4(15) of the General Data Protection Regulation (n 16).

[83] Article 1 of the General Data Protection Regulation (*ibid*).

[84] Per Recital 51 to the General Data Protection Regulation (*ibid*).

[85] In contrast against the General Data Protection Regulation, the Russian data-protection legislation presumes only one entity processing data, the 'operator', while under the GDPR there are both a 'controller' and a 'processor'. The definition of 'operator' more closely matches the 'controller' definition under the GDPR. This difference is described further on in the paper, in Section 4, which deals with requirements for processing of speech and voice.

[86] Federal Law 'On Personal Data' N 152-FZ (n 19) art 11.

[87] 'Explanations on the Issues of assigning Photo, Video, [and] Fingerprint Data and Other Information to Biometric Personal Data and the Features of Their Processing' issued by the *Roskomnadzor* on 30 August 2013. Available in Russian at: https://pd.rkn.gov.ru/press-service/subject1/news2729/ (accessed 13 April 2020).

3. The applicability of EU and Russian data-protection legislation

The literature emphasises that the right to data protection is a response to technological developments.[*88] The ease of accessing huge volumes of data is rapidly increasing apace with cross-border data flows driven by advances in developments of information and communication technologies and a shift toward a digital economy.[*89] This forces entrepreneurs to comply with the data-protection laws of all countries where their products and services are offered. For example, the social network LinkedIn was banned and now could not be accessed from the territory of Russia because it was in breach of the Russian data-localisation requirement[*90], discussed below – at that time, there was no LinkedIn Corporation legal entity in Russian territory (e.g., branches or representatives' offices).

This section addresses the question of when the EU and the Russian data-protection laws are applicable. The applicability of such laws depends on their territorial and material scope. Let us consider European law first. It defines protection of personal data as a fundamental human right[*91], without any limitation based on nationality or residence.[*92] The GDPR has extraterritorial character and applies both to entities established in the EU and to entities offering goods and services or monitoring data subjects there.[*93] The latter refers to targeting the EU market (i.e., the data subject is within EU territory).[*94] The indicator of targeting the EU market is the use of a language or currency of at least one of the EU member states.[*95]

In practical terms, the extraterritorial effect creates an obligation to comply with the GDPR's requirements. Entities not established in the EU must designate a representative of their operations targeting EU territory.[*96]

In contrast, Russian data-protection law does not have extraterritorial effect. It is not applicable to non-residents processing personal data of Russian citizens abroad. There are two exceptions, however. The first involves a 'data-localisation requirement', and the second is related to the implementation of the Yarovaya package law.

The localisation requirement for Russian citizens' personal data was introduced to Russian data-protection law by a federal law dated 27 April 2017 (242-FZ).[*97] The amendment added a new obligation for data-processing operators: their collection, storage, and use of personal data of Russian citizens must involve only databases on Russian territory.[*98]

This rule mandating local handling of Russian citizens' data must be complied with where the following conditions are met: 1) the information contains **personal data**; 2) personal data are **collected**, meaning the data-processing operator having received the data from third parties; 3) the data are **processed**, or their processing is organised by the operator; and the personal data pertain to **Russian citizens**.[*99]

The restriction of Russia's data protection to Russian citizens creates problems – for instance, how to determine the citizenship of a person who speaks to a voice assistant or how to detect that the voiceprint

[88] Hielke Hijmans, *The European Union As Guardian of Internet Privacy: The Story of Art 16 TFEU* (Law, Governance and Technology Series, Pompeu Casanovas and Giovanni Sartor (eds) vol 31, Springer 2016) 48. DOI: https://doi.org/10.1007/978-3-319-34090-6.

[89] Fabian Hungerland, Jörn Quitzau, Christopher Zuber, Lars Ehrlich, Christian Growitsch, Marie-Christin Rische, Friso Schlitte, and Hans-Joachim Haß, *The Digital Economy* (Strategy 2030 – Wealth and Life in the Next Generation series no 21e 2015).

[90] Case law includes: *LinkedIn Corporation v Roscomnadzor* 02-3491/2016. See the decision of the Tagansky District Court (Moscow, Russia) dated 4 August 2016 and the appeals determination of the Moscow City Court, dated 10 November 2016, on case N 33-38783 / 2016 https://mos-gorsud.ru/mgs/cases/docs/content/c364d1d9-e30c-4ffa-aabb-327c8977adab> (accessed 13 April 2020).

[91] Per Article 8 of the Convention for the Protection of Human Rights and Fundamental Freedoms (n 14).

[92] Recital 2 to the General Data Protection Regulation (n 16).

[93] See Article 3 of the General Data Protection Regulation (*ibid*).

[94] See Article 3 (2) of the General Data Protection Regulation (*ibid*).

[95] Recital 23 to the General Data Protection Regulation (*ibid*).

[96] Recital 80 to the General Data Protection Regulation (*ibid*).

[97] Federal Law 'On Amendments to Certain Legislative Acts of the Russian Federation Regarding the Clarification of the Procedure for Processing Personal Data in Information and Telecommunication Networks') N 242-FZ (n 23).

[98] Federal Law 'On Personal Data' N 152-FZ (n 19) art 18 (5).

[99] Alexander Savelyev, 'Russia's New Personal Data Localization Regulations: A Step Forward or a Self-Imposed Sanction?' (2016) 32.1 *Computer Law & Security Review* 128. DOI: https://doi.org/10.1016/j.clsr.2015.12.003.

being processed belongs to a Russian citizen. The Russian data-protection authority (the *Roskomnadzor*) attempted to solve the problem by issuing an official opinion.[*100] In that opinion, the authority replaced the term 'citizenship' with a reference to the territory. According to the opinion, in the event of doubts about the data subject's citizenship, all information collected and proceeded within the limits of Russian territory must be 'localised' to databases located in Russia.[*101] Applying this principle solves the problem of identification of citizenship. However, it leaves out Russian citizens' personal data collected outside Russian territory.

The application of Russia's data-localisation rule poses a significant hurdle for companies. The above-mentioned LinkedIn case is an excellent example, and it is far from the only one. Recently, the *Roskomnadzor* initiated review proceedings to determine the level of compliance with the data-localisation rule shown by the Facebook group of companies.[*102]

The second exception with regard to the nationally bounded character of Russian data-protection law is found under the Yarovaya package law. This law is not directly connected to data protection, and its material scope differs from that of the Russian federal law 'On Personal Data' and of the GDPR. The Yarovaya package law mostly concerns the public sector (public safety and national security). To some extent, it resembles the EU's Data Protection Police Directive.[*103] Since the Yarovaya package law creates new obligations related to the storage and processing of data, its applicability is analysed below.

The Yarovaya package law is a legislation package consisting of two federal laws that introduce amendments to the acts on combating terrorism. The law obliges the providers of telecommunication services and those organising information's dissemination to store the relevant Internet traffic data (text and voice messages, sounds, photos, videos, and files' metadata) for six months to three years.[*104]

The first issue that arises is that of the 'organiser of information dissemination' concept. The legal definition provided[*105] is too broad and could be taken to refer to virtually every Web page that interacts with a user (e.g., using cookies). Neither does the definition have a national restriction, and it could be considered to cover the Internet giants' companies, messaging services, blog-hosting platforms and owners of blogs that are hosted on such platforms, the owners and 'tenants' of domain names, etc. This legal uncertainty of the definition creates a legal risk for any companies that have a connection with the Russian market that might be covered by the description 'organiser of information dissemination'. That risk leads to the necessity of complying with the legal provisions cited above.

Compliance of communication service providers and organisers of information dissemination with the requirements of the Yarovaya package law could force companies into breaching other obligations – for instance, under their contracts (confidentiality obligations etc.), national legislation (e.g., the various national acts implemented in transposition of Directive (EU) 2016/680), and the GDPR's rules. One of the most significant examples of the far-reaching effects of the Yarovaya package law is the *Telegram* case[*106], involving blocking of services within Russian territory.[*107]

A summary of the framework provided above is presented in Table 1.

[100] Letter issued by *Roskomnadzor* N 08АП-3572, dated 19 January 2015.

[101] See page 5 of the letter of the *Roskomnadzor* (*ibid*).

[102] 'Роскомнадзор направил в Facebook запрос об исполнении российского законодательства' (Roskomnadzor Sent a Request to Facebook on the Implementation of Russian Legislation) (12 April 2020). http://www.interfax.ru/russia/608271 (accessed 13 April 2020).

[103] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

[104] A similar issue was addressed by the European Court of Justice in the context of the directive on privacy and electronic communications (2002/58/EC; 2009/136/EC). The Court found that the directive 'must be interpreted as precluding national legislation governing the protection and security of traffic and location data and, in particular, access of the competent national authorities to the retained data, where the objective pursued by that access, in the context of fighting crime, is not restricted solely to fighting serious crime, where access is not subject to prior review by a court or an independent administrative authority, and where there is no requirement that the data concerned should be retained within the European Union'. Joined Cases C-203/15 and C-698/15, 21 December 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1537508719221&uri=CELEX:62015CJ0203 (accessed 13 April 2020).

[105] Article 10.1 of Federal Law 'On Information, Information Technologies and Protection of Information' N 149-FZ (n 20).

[106] For case law, see: Tagansky District Court (Moscow, Russia) 02-1779/2018. https://mos-gorsud.ru/rs/taganskij/cases/docs/content/03a478c6-798c-4769-80eb-83d4d0a33b34 (accessed 13 April 2020).

[107] Neil MacFarquhar, 'Russian Court Bans Telegram App after 18-Minute Hearing' *The New York Times* (13 April 2018). https://www.nytimes.com/2018/04/13/world/europe/russia-telegram-encryption.html (accessed 18 June 2020).

Table 1: Summary framework

Application of EU and Russian data-protection legislation			
	European data-protection regulation	**Russian data-protection regulation**	
Sources	• The General Data Protection Regulation (GDPR)	• Federal Law 'On Personal Data' • Federal Law 'On Information, Information Technologies and Information Protection' • The Yarovaya package law	
Extraterritorial effect?	Yes	Only per the data-localisation rule and the Yarovaya package law	
Applicability	• EU companies • Non-EU companies with business activities within EU territory (targeting/ monitoring activity)	Data-localisation rule	Yarovaya package law
		• The data subject as a Russian citizen • Processing performed within Russian territory	• The provider of communication services • The organiser of the information dissemination
Specified connection with citizenship?	No	Yes	Neutral / not addressed – citizenship-agnostic definitions

One of the primary data-protection problems encountered in the development of language technologies is related to cloud computing and cross-border data flows. For instance, most voice assistants provide their services by means of cloud computing. Speech-recognition systems too are often built in a manner using cloud services, with Yandex SpeechKit being one example. The main problem currently plaguing the organisation of cross-border data flows between European countries and Russia is legal complication, involving friction among the GDPR, the Russian localisation requirement, and the requirements of the Yarovaya package law. To address the data-localisation rule, the *Roskomnadzor* published a letter aimed at tackling the problems wrought by that rule with regard to cross-border data flows. According to that letter, data of Russian citizens (or, in cases of any doubts about the citizenship of the data subject, data collected within Russian territory) should be initially collected and stored in databases that are physically on Russian territory, after which the material may be copied and transferred to databases situated in other countries.[*108] This leaves several questions, and, at the same time, the legal risk related to rules set forth in the Yarovaya package law are not solved. These various issues could negatively affect further co-operation between European countries and Russia.

4. The principles and rules for voice- and speech-processing

Since voice and speech are protected as personal data, their use (processing) is subject to several requirements. European and Russian jurisdiction both define data-processing in a broad manner, such that it covers virtually all activities performed with the given personal data. For instance, European and Russian data-protection regulations alike provide that the processing involves such operations with data as are

[108] Letter of *Roskomnadzor* (n 100).

carried out by either automatic or non-automatic means and involve such activities as collecting, recording, structuring, storing, using, and transmitting.*[109]

There are usually several parties involved in the processing of data in practice. The Russian and European data-protection scheme differ in how they articulate the identity of the parties performing data-processing activities. Russia's data-protection regulation defines only one body (the 'operator') in this regard that may perform data-processing activities. With its notion of the operator, Russian data-protection legislation refers to the body – defined as a legal person, natural person, or national/local government authority – performing the data's processing and determining the scope, means, and purposes for data-processing.*[110] According to the GDPR, meanwhile, there are two parties involved in data-processing activities (these parties may be represented by a single body): the 'processor' and the 'controller'. The processor is responsible for the technical part of the data-processing and performs the processing on behalf of the data controller.*[111] The data controller determines the means and purposes for processing the data. In comparison with the Russian data-protection regulation scheme, the operator is most similar in definition to 'controller'.

Russian law does not define the processor – the person who technically processes the data. However, under Russia's data-protection regulatory structure, the operator has a right to delegate the data-processing to a third party.*[112] Thereby, the Russian legal approach includes functions of the processor in the legal concept of the third party.

Internationally, the fundamental principles for data-processing are set forth in Article 5 of Convention 108*[113] and reflected in both Article 5 of the GDPR and Article 5 of Russia's federal law 'On Personal Data'. According to Article 5 of the convention, the personal data shall be lawfully obtained and processed,*[114] fairness is required,*[115] processing must be limited in line with the purposes for which the data were stored, the data must be relevant and accurate, and the data shall be kept in a form that permits identifying the data subject for no longer than the purposes for the data's storage necessitate.*[116] These are the fundamental principles that guarantee a certain minimum level of protection in the data-processing. The GDPR complements the list with the accountability principle. This principle for data-processing was developed by the Organisation for Economic Co-operation and Development (OECD). Under it, the data controller too is obliged to comply with the principles mentioned above.*[117] Neither Russian data-protection regulation nor Convention 108 highlights the latter principle.

These fundamental principles for data protection lay the groundwork for the rules on data-processing. The rules developed on their basis can be divided into three groups: those regarding lawful, secure, and transparent processing. Both jurisdictions' rules are discussed in terms of this classification below. Also, voice and speech can be either sensitive data (by virtue of falling into special categories of personal data) or non-sensitive, so the regulatory framework for processing should be investigated with regard to both of these categories as well.

Firstly, the principle of lawfulness of the processing means that the processing should be done in strict compliance with the law and that appropriate legal grounds for such processing must exist.

Under the GDPR, non-sensitive data are lawfully processed if one of the following grounds exists: 1) the data subject's consent, 2) performance of a contract, 3) compliance with a legal obligation, 4) protection of

[109] The full list of the operations that are considered to be data-processing is set forth (for the European approach) in Article 4(2) of the General Data Protection Regulation (n 16) and (for the Russian approach) in Article 3(3) Federal Law 'On Personal Data' N 152-FZ (n 19).

[110] Article 3 (2) of Federal Law 'On Personal Data' N 152-FZ (n 19).

[111] See Article 4(8) of the General Data Protection Regulation (n 16).

[112] Federal Law 'On Personal Data' N 152-FZ (n 19) art 6 (3).

[113] Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, ETS No.108 (n 15) art 5.

[114] See both Article 5(a) and Article 5(b) of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (*ibid*). For case law, see the ECtHR's: *Rotaru v Romania* [GC], 28341/95 04 May 2000; *Taylor-Sabori v The United Kingdom* 47114/99 22October 2002; *Peck v The United Kingdom* 44647/98 28 January 2003; *Khelili v Sweden* 16188/07. See also the CJEU case: *Huber v Germany* C-524-06 16 December 2008.

[115] See Article 5(a) of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, ETS No.108 (n 15). For relevant case law, consult: the ECtHR's: *Haralambie v Romania* 21737/03 29 October 2009; *K.H. and Others v Slovakia* 32881/04 28 April 2009.

[116] See Article 5 of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (*ibid*).

[117] Governance per Article 14 of the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (2013).

vital interests, 5) performance of a task carried out in the public interest, and 6) processing for purposes of pursuing legitimate interests.*[118]

Russian law provides for additional grounds for processing of non-sensitive data. For instance, non-sensitive data may be lawfully processed for purposes of statistics (the Russian data-protection regulations consider this to constitute separate and independent grounds for data-processing)*[119] or if the processing is performed to fulfil non-mandatory terms of the law with regard to information disclosure and so forth.*[120]

In general terms, the GDPR prohibits the processing of special categories of personal data (e.g., biometric and health data).*[121] However, there are the following exceptional cases in which processing is allowed: those of 1) explicit consent; 2) fulfilling one's obligations and exercising specific rights; 3) protection of vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent; 4) performing legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or to former members of the body or to persons who have regular contact with it in connection with its purposes and that the personal data are not disclosed outside that body without the consent of the data subjects; 5) processing related to personal data that are manifestly made public by the data subject; 6) processing necessary for the establishment, exercise, or defence of legal claims; 7) processing necessary for reasons of substantial public interest; 8) processing necessary for purposes of preventive or occupational medicine; 9) what is necessary for the public interest in the sphere of public health; 10) and processing necessary for purposes of archiving in the public interest, for scientific or historical research purposes, or for statistical purposes.*[122]

Russian data-protection law takes a different approach to sensitive and biometric data, so the rules for processing of voice and speech depend on how relevant the terms related to health or biometric data are. In cases wherein the voice and speech involve health data, the regulation of the data-processing is similar to that under GDPR rules. The general rule is to prohibit processing of this type of data.*[123] In contrast, Russia's data-protection law does not restrict the processing of biometric data as a special category of personal data. Instead, there is a requirement that processing be done only after receipt of the data subject's consent.*[124]

For the development of language technologies, the most relevant grounds are the data subject's consent and legitimate interest.

As for the second group of data-processing rules, referring to security, under the European approach, the implementation of the relevant measures is an obligation of the data processor and controller. The Russian approach presumes that the operator implements these measures. Security measures can be divided into two main groups: technical and organisational measures. Implicit to the European approach is that appropriate security measures should be implemented by design*[125] and should be applied by default.*[126] The GDPR provides a list of the technical measures that should be applied in the data-processing.*[127] For instance, among these measures are pseudonymisation and encryption of the personal data and measures to ensure the confidentiality, integrity, and availability of the data. The security requirements set forth under the GDPR follow the ISO 27001 standard.*[128] Organisational measures, in turn, are measures that can be implemented within the company with regard to the employees, other workers, etc. These include provision of information about data-security rules, clarifying these individuals' responsibilities and duties

[118] See Article 6 of the General Data Protection Regulation (n 16).

[119] See Article 6(1-9) of Federal Law 'On Personal Data' N 152-FZ (n 19).

[120] *Ibid*, art 6(1-11).

[121] General Data Protection Regulation (n 16) art 9(1).

[122] *Ibid*, art 9(2).

[123] See Article 1 of Federal Law 'On Personal Data' N 152-FZ (n 19). This provides a list of the exceptions to the general rule set forth in Article 10(2) of 'On Personal Data'.

[124] This is addressed by Article 11 of Federal Law 'On Personal Data' N 152-FZ (*ibid*).

[125] The relevant technical and organisational measures should be integrated into the data-processing process.

[126] See both Article 25(1) and Article 25(2) of the General Data Protection Regulation (n 16).

[127] Per Article 32 of the General Data Protection Regulation (*ibid*).

[128] See the ISO/IEC 27000 family – Information Security Management Systems: https://www.iso.org/isoiec-27001-information-security.html (accessed 13 April 2020).

with regard to data protection.*[129] The Russian approach too presumes that data-processing should employ both technical and organisational safeguards for security*[130]; however, the law 'On Personal Data' makes only general provisions for required security measures.

The last group of rules, that related to transparency of processing, deals with the data subject's right to understand the essence of any automated processing of personal data, the main purposes of that processing, and the identity and habitual residence or place of business of the controller of the data-processing.*[131]

The principles and rules for the data-processing are the basis that should be taken into consideration by those companies conducting business activities in Russian or European territory. Compliance with these data-processing rules demands awareness of the scope of the data subject's legal rights with regard to data protection. These rights are not absolute, and they need to be balanced with the other fundamental rights, such as freedom of expression, freedom of thought, freedom of expression and of information, religious freedom, and linguistic diversity.*[132] The right to linguistic diversity may play an especially significant role in the further development of language technologies and use of voice and speech in their development.

5. Conclusions

With regard to the field of development of LTs, the European and the Russian stance to data protection are quite close in approach but at the same time very far apart. The above analysis of European and Russian legislation shows that these jurisdictions apply similar international legal grounds and follow the same internationally recognised data-protection principles; however, the data-protection regulations are not fully harmonised between the two. For instance, the EU and Russia identify different subjects of data-processing and different scope of obligation for such subjects. Moreover, the EU and the Russian data-protection regulation scheme diverge with regard to the importance of the citizenship of the data subject and differ in the nature of their international application (most importantly, as a general rule, Russia's data-protection legislation does not have extraterritorial effect).

Examination of the relevant laws showed that voice and speech are considered personal data in both jurisdictions. Therefore, there is a need to follow data-protection laws in this connection.

The human voice can be personal data, or it can belong to special categories of personal data. Which rules are applicable depends on such factors as the type of personal data involved (does voice fall under special categories of personal data?), the form of data storage (is the material anonymised or not?), the place where the data-processing takes place, particular circumstances, and the purpose of the processing. Moreover, in some cases, the applicability of the law depends on the citizenship of the data subject and the territorial focus of the processing activities.

The differences and conflicting legal norms between these jurisdictions create legal obstacles to co-operation extending between the two. The reality is that entities involved with language technologies targeted at both EU and Russian territory must simultaneously comply with the regulation systems of both jurisdictions – which are not compatible with each other. This creates a situation wherein a company needs to choose which regulation has to be breached for the sake of other compliance. Therefore, clear grounds exist for further research and investigation aimed at identifying a possible solution that might solve the problem of conflicting norms.

[129] ECtHR: *I. v Finland* No. 20511/03 17.07.2008.

[130] Federal Law 'On Personal Data' N 152-FZ (n 19) art 19.

[131] Article 8(a) of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, ETS No.108 (n 15).

[132] Per Recital 23 to the General Data Protection Regulation (n 16).