Article

DOI: 10.17803/2713-0533.2024.3.29.491-513

Digital Profiling and the Legal Regime of Derived Personal Data

Artur N. Mochalov
Ural State Law University named after V.F. Yakovlev,
Yekaterinburg, Russian Federation

Abstract: The paper discusses some aspects of the legal regulation of personal data profiling in various jurisdictions. It focuses on derived personal data, also known as inferences, which are the outputs of digital profiling and automated decision-making. Although the extraction of new knowledge about individuals based on the processing of personal data has become common practice in both the commercial and public sectors, there have been only a few attempts to establish specific legal frameworks for derived personal data. These include the European Union, California (USA), and Singapore. Using a comparative legal approach, the author analyzes the characteristics of derived personal data and how the rights of individuals are protected in relation to derived personal information in these jurisdictions and in Russia as well. After examining the relevant laws and regulations, the author concludes that these attempts to regulate derived personal data are an effort to adapt traditional legal frameworks to the challenges posed by Big data. At the same time, the protection of personal data when using Big data technologies and artificial intelligence requires advanced regulatory approaches. Today, data extraction processes are often hidden from data subjects and not under their control. The author believes that the automated processing of personal information, including digital profiling and the extraction of new personal data, should be made more transparent and allow users to opt out.

Keywords: personal data; derived personal data; inferences; profiling; data mining; privacy

Acknowledgements: The reported study was funded by Russian Science Foundation, project number 24-28-01378

Cite as: Mochalov, A.N., (2024). Digital Profiling and the Legal Regime of Derived Personal Data. *Kutafin Law Review*, 11(3), pp. 491–513, doi: 10.17803/2713-0533.2024.3.29.491-513

Contents

I. Introduction

The term "profiling" in relation to personal data means a set of practices of creating, discovering or constructing knowledge about a person from large sets of data from a variety of sources (Niševic, 2020, p. 104).

By analyzing personal information, computer algorithms allow to obtain new knowledge about people, which was not initially known to the controller. For example, processing data about a user's social network behavior, it is possible to get reliable information about their age, education level, interests, hobbies, beliefs, and even political preferences. A few years ago, in the USA there was a public outcry due to the actions of Cambridge Analytica Company. They processed data from users' accounts in a popular social network to identify potential

Republican Party supporters and target them with a campaign for Donald Trump (Day, 2020).

Today, digital profiling of internet users is a common practice in data-driven business models. News services and marketplaces, for example, use profiling to generate personalized recommendations, while advertising operators use it to create targeted ads. The material for profiling comes from digital footprints left by Internet users on various sites, including "likes" on social networks, search queries, or the delivery time of goods purchased on marketplaces. In recent years, information transmitted by smart devices has also been actively used by controllers for profiling the users (Wiedemann, 2022). In all cases, profiling produces new knowledge about individuals by deriving hidden and non-obvious information from a primary data set. This information can be used for evaluating individuals (social scoring systems) and predicting their behavior. In this regard, the results of profiling are widely used, particularly by banks to assess the solvency of customers, employers to select candidates for vacant positions, and law enforcement agencies to identify persons prone to illegal behavior (Westerlund et al., 2021, p. 34). The social rating system has been most developed in the People's Republic of China (Vinogradova et al., 2021, pp. 9–10).

Despite the widespread use of profiling in the processing of personal data by computers, there is relatively little special regulation regarding derived personal information about an individual. The article will focus on three jurisdictions where such regulation exists — the European Union, Singapore, and the State of California (USA). It will be shown that there are still many controversial issues in establishing the legal regime for derived personal data, which the legislator approaches differently in each case. The article will discuss some features of derived personal data that distinguish them from "classical" (primary) personal data. Then, the specifics of the implementation by data subjects of individual rights in relation to derived data will be analyzed, considering their characteristics. In particular, the right to access derivative data about themselves, the right of rectification of derivative data and the right to delete them ("right to be forgotten") will be considered.

II. Conceptual and Legal Framework

Profiling is based, first, on social and psychological patterns of people's behavior and, second, on the statistical correlations, which allow to "calculate" certain characteristics of a person based on information about his or her previous activity, to determine his or her interests and predict likely actions in the future (Day, 2020, pp. 596–599). Classifying people according to their psychological types and understanding their behavior is nothing new in science. However, with the advent of Big data technologies and machine learning, it has become possible to process information about a large number of people at once, identify previously unknown patterns, and quickly obtain accurate results (Adjerid and Kelley, 2018). This has led to the development of a new economy based on data, where personal information has become a valuable digital asset.[1]

In computer science, the term "data profiling" has a narrower meaning and refers to the process of preparing and technically analyzing data for subsequent use. Profiling is aimed at improving the quality of data, eliminating errors, contradictions, and duplications. In the process of profiling, it is possible to identify patterns, rules, and trends in data, and determine dependencies between different data elements. The extraction of new non-obvious knowledge by computer processing of existing information is called data mining (Naumann, 2014).

The use of data mining to process information about a person has raised difficult questions, since statistical correlations, as it turned out, completely ignore the requirements of personal data legislation (Roig, 2017, p. 6). From the data on a user's behavior in a social media platform, it is possible to extract information that infringes their privacy, such as their philosophical convictions or political opinions. These inferences about a person's characteristics may be biased (Chander, 2017). There are cases where computer algorithms have denied employment opportunities to all female candidates or have considered individuals with common Afro-American names to be potential criminals. Therefore,

[1] *See* Personal Data: The Emergence of a New Asset Class. An initiative of the World Economic Forum January 2011. Available at: https://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf [Accessed 06.04.2024].

the rights of individuals in relation to profiling and the use of extracted personal information cannot be effectively safeguarded solely through traditional protections provided by personal data legislation. Instead, specific regulatory measures are necessary.

The European Union has taken the lead in the legislative framework for data profiling. Enacted in 2016, the General Data Protection Regulation, commonly known as GDPR,[2] establishes a framework for automated processing of personal data that involves using personal data to evaluate or predict specific aspects of an individual's personal life, such as their work performance, economic situation, health, personal preferences, interests, reliability, behavior, location or movements. From the definition of profiling in Art. 4(4) of GDPR, two features of profiling are seen: (a) it is always automated processing of personal data; (b) the special purpose of processing is the assessment of "certain personal aspects relating to a natural person," including the analysis or prediction of his or her behavior. The GDPR does not clearly distinguish between the process of creating a person's profile and making a decision based on the created profile. Some contributions suggest that profiling does not include the automated decision-making stage (Wiedemann, 2022).

The GDPR is silent about the legal nature of estimated or inferred knowledge about a person obtained during profiling or decision-making based on a digital profile, and does not use the term "derived data." In particular, it avoids the question of whether such knowledge relates to personal data or is a separate type of information. However, Art. 4(1) of the Regulation does not link the assignment of information to personal data with the method of obtaining it. On this basis, it can be concluded that even if information about a person is not *collected*, but *created* on the basis of primary data, this is not a reason not to consider it personal data (Wachter and Mittelstadt, 2019, p. 518). The Article 29

[2] 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). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [Accessed 06.04.2024].

Data Protection Working Party[3] (hereinafter Art. 29 Working Party), in its Guidelines on automated individual decision-making and profiling within the framework of the General Data Protection Regulation (GDPR), acknowledged the existence of inferred or derived data about individuals. These data were described as "new personal information that has not been directly provided by the data subjects themselves."[4]

Unlike the GDPR, the California Consumer Privacy Act (CCPA),[5] adopted in this American State in 2018, explicitly refers to personal information "inferences drawn from any of the information identified in this subdivision [definition of personal information] to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes." "Inference" means the derivation of information, data, assumptions, or conclusions from facts, evidence, or another source of information or data. The scope of CCPA is narrower than the GDPR. This Act applies only to the processing of data about consumers who are citizens of the State of California by commercial corporations. It does not regulate profiling carried out by law enforcement bodies and other government agencies.

The third example of legal regulation of derived data can be found in the Singapore Personal Data Protection Act (PDPA) following the amendments made to it in 2020.[6] Derived data is defined under the PDPA to refer to new data elements that are created by an organization in the course of business from other personal data about the individual (or another individual), in the possession or under the control of the organization. Like the CCPA, the PPDA treats derived data as personal data.

[3] The Working Party was set up under Art. 29 of Directive 95/46/EC as an independent European advisory body on data protection and privacy.

[4] Art. 29 Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Adopted on 3 October 2017 WP251rev.01. Available at: https://ec.europa.eu/newsroom/article29/items/612053 [Accessed 06.04.2024].

[5] The California Consumer Privacy Act of 2018. Available at: https://theccpa.org [Accessed 06.04.2024].

[6] Available at: https://sso.agc.gov.sg/Act/PDPA2012 [Accessed 06.04.2024].

Despite the widespread use of profiling, special regulation of derived data is currently rare. For example, it is not included in the Chinese Personal Information Protection Law 2021 (hereinafter PIPL)[7] or in the Russian Federal Law "On Personal Data" No. 152-FZ.

III. The Two Key Features of Derived Personal Data

There are at least two important features of derived personal data that distinguish them from "classic" personal data. The first feature is that derived data is not "collected" in the usual sense (i.e., received directly from the data subject or from third parties), but is created (or "calculated") as a result of automated processing of other (primary) personal data. The second feature is the probabilistic or inferred nature of derived personal data. These features will be discussed in more detail later.

III.1. Derived Data as Non-Collected Data

Unlike ordinary personal data that is collected from the data subject or from third parties, derived data does not have a collection stage. The process of obtaining derived personal data is most often hidden from the subject, and the subject may not even know that the controller has become aware of personal information that he or she did not provide. Accordingly, with respect to derived personal data, there is most likely no explicit consent of the data subject to their processing.

As noted above, the assignment of information to personal data does not depend on the method of its receipt. The Article 29 Working Party pointed out that there are three types of personal data based on their origin:

— "actively and knowingly" provided by the data subject;

— "observed" data that characterizes the subject's activity (for example, the history of search queries or information transmitted by trackers of devices such as fitness bracelets);

[7] Personal Information Protection Law of the People's Republic of China (Adopted at the 30th Meeting of the Standing Committee of the Thirteenth National People's Congress on 20 August 2021). Available at: http://en.npc.gov.cn.cdurl.cn/2021-12/29/c_694559.htm [Accessed 06.04.2024].

— inferred or derived data, which are created by the controllers themselves on the basis of data provided by the subjects and are the result of computer algorithms.[8]

From the perspective of the Art. 29 Working Party, derived data, although not explicitly provided by the data subject, should nevertheless be treated as personal data within the scope of the GDPR. However, certain legal safeguards granted to data subjects under the GDPR, such as the right to portability of data, are restricted to collected and observed personal data and do not extend to derived data.

In California, the State Attorney General's Office issued Opinion No. 20-303, dated 10 March 2022, explaining certain aspects of the CCPA in relation to derived personal data.[9] The document effectively equates derived data with collected data, since inferences constitute a part of the consumer's unique identity and become part of the information that the business has "collected about" the consumer.

The logic of these explanations suggests that the restrictions imposed by law for the collection of personal data should also apply to computer generating new derived data. In particular, this applies to the rule that the amount of personal data should be the minimum necessary to achieve the purpose of their processing. The purposes of processing derived personal data, in turn, must be legitimate, pre-defined and clearly formulated. Derived personal data should not be obtained and used for purposes incompatible with the purposes of primary data collection.

The difficulty, however, lies in the fact that the process of profiling and subsequent decision-making can involve personal information obtained from different sources and collected by various controllers for different purposes. The primary data for creating digital profiles is provided by the subject at different times and in different circumstances. Moreover, when forming a digital profile and discovering new knowledge about a subject, data is used that relates both to this subject and to other

[8] Art. 29 Working Party. Guidelines on the right to data portability Adopted on 13 December 2016. Available at: https://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf [Accessed 06.04.2024].

[9] Available at: https://www.dwt.com/-/media/files/2022/03/20-303.pdf [Accessed 06.04.2024].

persons with similar characteristics (for example, to predict the subject's consumer preferences or the probability of non-repayment of a loan). During processing, personal data can be combined with information that is not personal data. Based on some derived personal data, other derived personal data may be created. Therefore, in practice, it is almost impossible to correlate derived personal data and the purposes of their use with the purposes of collecting primary personal data. Traditional guarantees of a subject's control over the use of their data, such as the principle of consent of the subject, limitation of the amount of data processed and limitation of purposes, become illusory when it comes to Big data technologies (Gonçalves, 2017, p. 98; Savelyev, 2015).

III.2. Inferred Nature of Derived Data

The second characteristic of derived data is its inherently probabilistic nature. Despite the advancements in Big data processing techniques that allow for relatively accurate assumptions, these data are still the product of computer calculations based on statistical correlations. For instance, a social media user's age group can be inferred with a high degree of probability based on their membership in certain communities, their likes, emojis, and comments. Likewise, gender, citizenship, nationality, religious affiliation, and political views can also be "calculated" to some extent. However, these predictions remain approximations and may not always hold true.

The outcome of profiling may also result in the generation of sensitive data that falls under specific categories necessitating the individual's explicit consent for processing. In numerous jurisdictions, this encompasses, for instance, data pertaining to an individual's medical condition or beliefs. As mentioned previously, derived data is not obtained directly from the subject, and its creation is typically not accompanied by the acquisition of the subject's consent. However, what if the automated analysis of personal information reveals sensitive data about an individual? Some may argue that the probabilistic nature of these conclusions (for instance, regarding an estimation of a person's health based on their purchases) exempts the entity that obtained this data from seeking the individual's permission.

In literature, there is a suggestion to employ the term "quasi-health data," which refers to inferred data about an individual's condition (just "indirectly related to health"), particularly based on information obtained from smart devices. Such data may be confidential, but in a legal context, it should not be construed as health information. (Malgieri and Comandé, 2017). However, this approach does not answer the question of the legal conditions for processing sensitive derivative data designated as "*quasi-*" (Fischer, 2020, p. 39). The presumed nature or even inaccuracy of inferred information does not mean that such information is not personal data or does not relate to a specific person. However, if new data related to specific categories is calculated during the analysis of data (including those that do not belong to special categories), its processing will require the consent of the subject. The same position is shared by European commentators.[10]

Certain types of derived personal data may constitute an opinion or evaluation, such as inferences regarding an individual's preferences, trustworthiness, financial capacity, or projected future conduct. The distinctive aspect of this information is its unverifiability.

Among the compared legal acts, the CCPA most clearly refers assumptions and conclusions to the personal information that falls into the scope of this Act. The GDPR is silent on whether personal data includes information about a person that is an estimate, prediction, or analytical conclusion. The Article 29 Working Party assumes that non-verifiability of information about a person is not an obstacle to considering it as personal data[11] (Wachter and Mittelstadt, 2019, p. 520). At the same time, in the law enforcement practice of the European Court of Justice, this question has not received a clearness. In its decision of 17 July 2014, it concluded that the legal analysis of immigrants' applications may contain personal data but cannot be classified itself as personal data within the meaning of Directive

[10] Art. 29 Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

[11] Art. 29 Data Protection Working Party. Opinion 4/2007 on the Concept of Personal Data on 20 June 2007. Available at: https://www.europarl.europa.eu/cmsdata/183970/20080130ATT20135EN.pdf [Accessed 06.04.2024].

95/46/EC, which was in force prior to the adoption of the GDPR.[12] In a subsequent decision on 20 December 2017, the court expanded the definition of "personal data" to include written answers submitted by a candidate at a professional examination and any comments made by an examiner with respect to those answers.[13] However, such an extension of the scope of the personal data legislation was limited by the Court to specific circumstances. Following the logic of these decisions made in non-digital contexts, there is no reason to state unequivocally that the conclusions drawn as a result of profiling are personal data within the meaning of the GDPR.

IV. Rights of Data Subjects with Respect to Derived Personal Data

Considering the features of derived personal data, a number of controversial issues arise regarding the implementation of the rights of subjects — in particular, the right to access data, the right of rectification and the right to delete or to demand erasure of data ("right to be forgotten").

IV.1. Right to Access

In the context *of the right of access to personal data*, the question of who owns the derived personal data is important. In theory, the model of "ownership" of personal data, which means that the data subject is regarded as owner of information about himself or herself, has become widespread. This doctrinal model is resulted in general rule of the need for the subject's consent to the processing of personal data, which is enshrined in the legislation of most countries, or the right of

[12] CJEU — C-141/12 and C-372/12 — YS v. Minister voor Immigratie, Integratie en Asiel. Available at: https://curia.europa.eu/juris/document/document.jsf?text=&docid=155114&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=671001 [Accessed 06.04.2024].

[13] CJEU — C-434/16 — Peter Nowak. Available at: https://curia.europa.eu/juris/document/document.jsf;jsessionid=BC736E3C6C1250DFC36D8A676A461F8C?text=&docid=198059&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=798290 [Accessed 06.04.2024].

data portability, which is guaranteed, for example, in Art. 20 of the GDPR (Bouchagiar and Bottis, 2018, pp. 226–227). However, derived personal data, as noted above, is not received from the data subject or from a third party. From this standpoint, is it reasonable to regard them as information that continues to belong to the data subject, or does such information, from the moment of its creation, become the sole "property" of the controller?

Assuming that it is the controller who derived the personal data that owns them, then, then the controller may refuse the subject access to them, referring, for example, to the fact that this information is a trade secret. Indeed, derived personal data may have the characteristics of a trade secret: they are created in the course of the activities of a holder, are not known to third parties, and have commercial value for the holder (Bottis and Bouchagiar, 2018, p. 208).

In the abovementioned Opinion the Attorney General of California emphasized, that according to the CCPA, "if the business holds personal information about a consumer, the business must disclose it to the consumer on request." Without explicitly addressing the question of ownership of information that is generated internally, the CCPA guarantees the right to access personal data in any situation. "The plain language of the statute, as well as the legislative history, persuade us that the CCPA purposefully gives consumers a right to receive inferences, regardless of whether the inferences were generated internally by the responding business or obtained by the responding business from another source," the Opinion says.

Similarly, Article 15 of the GDPR refers to the right of a data subject to request confirmation from the controller, that "personal data *concerning* him or her" (not "*collected*" *from* him or her) are being processed. It is obvious that the scope of Art. 15 extends beyond only "collected" personal data. However, it is not clear how far it extends (Custers and Vrabec, 2024). Recital 63 of the GDPR indicates restrictions on the right to access personal data, in particular if this violates the rights and freedoms of others, including the right to trade secrets and intellectual property results. The definition of trade secret in the EU Trade Secret Directive[14] is so broad as to include nearly any data handled by a commercial entity, in particular information about consumers' behavior (Wachter and Mittelstadt, 2019, p. 607). This means that trade secret protection considerations significantly limit the access of subjects to derived personal data.

Unlike the right of access to data, *the right to data portability* provided for in the GDPR does not apply to derived data. According to Art. 20 of the GDPR, the data subject has the right to receive from the controller the personal data related to him in a machine-readable format, which he *provided* to this controller, and transfer them to another controller — if the processing of such data is carried out in automated systems *based on the consent of the data subject*.

This approach is likely aimed at protecting the economic interests of data controllers who have invested resources in data mining to extract valuable personal information. Unlike primary data, which can be collected multiple times from a subject or third parties, derived data are a unique product of a controller's efforts (the result of computer algorithm processing) and have greater economic value. Therefore, freely transferring such data in a machine-readable format from the controller that created it to other controllers that did not invest resources in obtaining it would disproportionately limit their economic interests.

IV.2. Right to Rectification

The specific features of derived personal data are manifested in the exercise of the *right of the data subject to rectification* of inaccurate or irrelevant information.

In various jurisdictions, accuracy and adequacy are usually proclaimed among the fundamental principles of personal data processing. At the same time, derived personal data, as mentioned above, are inferred. This means that there is a possibility of error

[14] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0943 [Accessed 06.04.2024].

in identifying certain characteristics of a person based on statistical correlations identified in the primary data. Does this mean that if an error is detected, the derived data must be updated at the request of the subject? Article 16 of the GDPR refers to the right of the subject to require the controller immediately rectification of inaccurate personal data concerning him or her. From this wording, it can be concluded that such a right should apply to both collected and derived personal data. However, if the procedure for clarifying the verifiable data is clear, then the probabilistic or estimated characteristics of a person may not always be changed at the request of the subject. For example, a data subject may say that their music preferences and individual recommendations for a playlist on a music listening service are defined incorrectly and do not correspond to their wishes. In turn, the service administration can claim that the selection of music was performed correctly, as a result of computer calculations based on data about tracks previously listened to by the user, as well as information about the music preferences of other users of the service with similar characteristics and interests (Custers and Vrabec, 2024, p. 55).

PDPA (after the changes made in 2020) demonstrates another approach. Article 22 of the PDPA, like the GDPR, establishes the data subjects' right to send requests to controllers for correction of data about themselves. However, Article 22(6) contains a number of exceptions from this rule, including derived personal data and "opinion data kept solely for an evaluative purpose." The term "evaluative purpose" is defined in Art. 2(1) of the PDPA. This includes, in particular, "the purpose of determining the suitability, eligibility or qualifications of the individual to whom the data relates" in such fields as employment or education.[15]

The Singapore legislator likely assumes that derived and opinion data do not belong to the individual who is the subject of the data, but rather to the entity that created the information. From the perspective

[15] *See* Advisory Guidelines on Key Concepts in The Personal Data Protection Act (Revised 1 October 2021). Available at: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Key-Concepts/Advisory-Guidelines-on-Key-Concepts-in-the-PDPA-1-Oct-2021.ashx?la=en. P. 107 [Accessed 06.04.2024].

of the Personal Data Protection Commission of Singapore, accuracy in derived personal data shall be achieved through accurate categorization and selection criteria (i.e., adequate business rules) at the data processing stage.[16]

IV.3. "Right to be Forgotten"

The rights of a personal data subject usually include the right to request the termination of processing of their personal data and their deletion (if there are no other legal grounds for their storage and processing by the controller) — the so-called "*right to be forgotten*." The subject may be interested in prohibiting the processing of derived personal data not only if they are incorrect or irrelevant, but also if such data is sensitive information for the subject that he would not have provided to the controller at his own will, including if the processing of such information requires the individual's mandatory consent in accordance with the law. The acquisition of such knowledge about the subject without their consent may be regarded as a disproportionate invasion of their privacy. The subject may also wish to stop processing and delete conclusions and assessments based on the analysis of the primary data, if the use of this information by the controller or other parties poses a risk of discrimination to them.

At the same time, the CCPA guarantees the consumer's right to request the deletion of only the personal information that was collected from this consumer. It can be concluded that in California, the consumers' right to delete the data does not apply to inferences. A similar conclusion can be drawn from the analysis of the Singapore PDPA, which does not provide for the "right to be forgotten," but only speaks about the possibility of the subject to withdraw consent to the processing of personal data. At the same time, it seems that if the derived data belongs to special categories of data, the processing of which can only be carried out with the consent of the subject, then the subject will

[16] Advisory Guidelines on Key Concepts in The Personal Data Protection Act (Revised 1 October 2021). Para. 16.9.

have the right to demand that the processing of data about him or her be stopped on the grounds that this data is being processed illegally.

The GDPR rules differ significantly from the CCPA and PDPA. Article 17(1) of the GDPR assigns the subject "right to obtain from the controller the erasure of personal data concerning him or her" regardless of the way, in which this data was obtained. However, right to be forgotten is not absolute and can only be implemented if certain conditions are met. In particular, the subject may request the deletion of data about them if such data is processed illegally or are processed for direct marketing purposes, including profiling. In addition, Article 21 of the GDPR establishes the right of the subject to object to the processing of data about him or her, including profiling, if the data is processed in the public interest or for the purpose of ensuring the legitimate interests of the controller or a third party. In the event of an objection, if there are no legally binding legal grounds for processing, the personal data must also be deleted.

The subject's interests during automated processing of personal data (including profiling) are also protected by special guarantees, provided by Art. 22 of the GDPR, among which the right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." This provision is criticized in the literature due to its limited practical application (Davis and Schwemer, 2023). First, it applies only to the cases where automated processing, including profiling, leads to legally significant consequences. Conclusions or assessments derived from personal data through their computer processing may have serious consequences for the subject, including long-term ones, but may not always be described in terms of "legal effects." Secondly, Article 22(1) only deals with cases where a legally relevant decision is based solely on automated processing and, therefore, does not apply to semi-automated procedures, when part of the data processing operations is performed with the participation of a person. Moreover, Article 22(2) of the GDPR sets out a number of significant restrictions in the implementation of the right.

IV.4. Other Rights and Special Guarantees

The new EU Artificial Intelligence Act,[17] adopted by the European Parliament on 13 March 2024, consolidates guarantees of the rights of individuals against unfair derivation and use of specific types of personal data. In light of the fact that data mining often involves machine learning and the outputs generated by AI algorithms are often unpredictable and difficult to explain (Fischer, 2020), Article 5 of the Act prohibits placing such AI systems on the market and using them, in particular, for the evaluation or classification of natural persons or groups of persons based on their social behavior or known, inferred or predicted personal or personality characteristics, with the social score leading to discriminatory or unfavorable treatment of certain natural persons or groups in social contexts that are unrelated to the contexts in which the data was originally generated or collected, or of such treatment is unjustified or disproportionate to their social behavior or its gravity. The prohibition will come into effect on 1 January 2025. It will also be illegal to use AI to assess or predict the risk that a natural person will commit a criminal offense, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; to infer emotions of a natural person in the areas of workplace and education institutions; to categorize individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. Should any data regarding an individual be obtained through the use of AI technologies in contravention of this prohibition, it shall be deemed subject to erasure in accordance with Art. 17(1d) of the General Data Protection Regulation (GDPR). Nonetheless, these restrictions do not impede the employment of AI for purposes related to security and law enforcement, nor do they apply in certain other situations.

Special guarantees of subjects' rights related to the processing of derived data may also be provided for cases where such data is processed for marketing purposes, including in recommendation services. In

[17] The EU Artificial Intelligence Act. Available at: https://artificialintelligenceact.eu [Accessed 06.04.2024].

China, for instance, the PIPL does not explicitly regulate profiling or the use of derived personal data. However, it does provide in Art. 24 that commercial marketing targeting individuals based on automated decisions must be accompanied by options that are not specific to their personal characteristics, and individuals must have convenient means to opt out. In March 2022, the Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services came into force in China.[18] Article 17 of the document obliges the algorithmic recommendation service providers "to provide users with a choice to not target their individual characteristics, or provide users with a convenient option to switch off algorithmic recommendation services." Moreover, algorithmic recommendation service providers shall provide users with functions to choose or delete user tags used for algorithmic recommendation services aimed at their personal characteristics. Such regulation allows subjects to avoid derivation of inferences for marketing purposes on the stage of providing the primary data.

V. Legal Regime of Derived Personal Data in Russia

The Russian Federal Law "On Personal Data" dated 27 July 2007 No. 152-FZ (hereinafter FLPD), like the Chinese PIPL, does not contain special rules concerning profiling or the derived personal data.

In 2019, by Decree of the Government of the Russian Federation No. 710, the concept of "digital profile" was introduced into official circulation.[19] However, the term "digital profile" in this document is used in a different sense than in the GDPR or CCPA. Digital profile in

[18] The Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services: Order of the Cyberspace Administration of China, the Ministry of Industry and Information Technology of the People's Republic of China, the Ministry of Public Security of the People's Republic of China, and the State Administration for Market Regulation No. 9. Adopted 31 December 2021. Available at: https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm [Accessed 06.04.2024].

[19] Decree of the Government of the Russian Federation No. 710 dated 3 June 2019 "On conducting an experiment to improve the quality and connectivity of data contained in State information resources."

the context of the Decree means a set of up-to-date and reliable data and other information about individuals or legal entities generated in the Unified Identification and Authentication System or other information systems of state and local government bodies, as well as organizations subordinate to them, in order to provide such information with the consent of the subjects to the entities who have requested access to it (Vinogradova et al., 2021, p. 8). Creating a digital profile, therefore, is limited to the information that is used in the public sector, does not involve the use of inferred data, and does not aim to discover or evaluate personal qualities or predict person's behavior. Profiling in the sense that it is used in the GDPR or CCPA is regulated in Russia only by the general provisions of personal data legislation.

Article 16 of the FLPD outlines the rights of individuals in relation to automated processing of their personal data. However, the provisions of this article only apply to situations where decisions are made based solely on automated processing, which have legal consequences for the individual or significantly impact their rights and legitimate interests. Such decisions can only be taken with the explicit consent of the individual, and they have the right to object to such decisions.

The right of objection provided for in Art. 16(4) of the FLPD cannot be exercised in situations where automated processing of personal data is undertaken for marketing or other purposes that do not have direct legal effect on the individual. Moreover, this provision does not apply when the processing of data is not exclusively automated. Consequently, a significant portion of digital profiling activities falls outside the scope of Art. 16 of the FLPD.

The Law does not prohibit the extraction of new knowledge from processed personal data about individuals. At the same time, conditions for the processing of personal data in accordance with Art. 6 of the FLPD may include not only the explicit consent of the individual, but also other circumstances, such as the fulfillment of contractual obligations by the data controller (or "operator") to the individual or the exercise of rights and legitimate interests by the operator or third parties. In practical terms, this latter circumstance can be interpreted broadly to include rights and interests of the operator related to economic activities.

Derived data, within the scope of the FLPD, remains personal data as defined by the Law as "any information that relates directly or indirectly to a specific or an identifiable natural person." The Law does not tie the subject's right to access, clarify, block, or delete personal data to the method of obtaining such data, as stipulated in Art. 14(1) of the FLPD. This implies that these rights can be exercised with respect to derived data as well, provided that the data is incomplete, obsolete, inaccurate, obtained illegally, or unnecessary for the specified purpose of processing. The FLPD acknowledges certain exceptions, primarily related to security concerns and the conduct of law enforcement operations. Additionally, the individual's exercise of their right to access personal data might be denied if it results in a violation of the rights or legitimate interests of other parties.

Starting from 1 October 2023, Art. 10.2-2 of the Federal Law "On Information, Information Technologies and Information Protection" No. 149-FZ also applies in Russia, which provides for the specifics of submitting information using recommendation technologies based on the collection, systematization and analysis of information related to the preferences of Internet users. This Article obliges providers of recommendation services to disclose information about user preferences that are used to generate recommendations, as well as not to violate the rights and legitimate interests of citizens and organizations. At the same time, unlike the Chinese PIPL, the Russian Law does not provide for the right of users to refuse using the recommendation technologies in relation to them or to prohibit processing of certain information about their preferences. It is also debatable whether the Russian legislator considers information about user preferences as personal data. The current regulation provides a significant degree of flexibility for operators who process personal data from Russian internet users for marketing purposes. This includes the use of neural network technologies that may lead to potential violations of the rights of individuals whose data is processed (Minbaleev and Storozhakova, 2023, pp. 76–78).

VI. Conclusions

The timid efforts of legislators to regulate the use of derived personal data represent their desire to adapt traditional legal mechanisms to processes of digital profiling that rely on Big data and artificial intelligence technologies. At the same time, existing approaches to personal data regulation do not work in the context of Big data (Bottis and Bouchagiar, 2018; Gonçalves, 2017; Savelyev, 2015).

The legal framework for personal data protection remains highly conservative, continuing to view personal data as information originating from an individual, belonging to them, and typically requiring their consent for use. However, valuable personal information is increasingly extracted through computational processes, often without the consent of the data subjects. The processes of discovering non-obvious personal data during profiling and its subsequent use by controllers for estimation and prediction subject's behavior are typically hidden from the data subjects and beyond their control. In light of these developments, there is a need for alternative regulatory paths in personal data protection, shifting the emphasis from merely how data is collected to how it evolves (Wachter and Mittelstadt, 2019, p. 615). The new approaches require increased transparency in automated decision making and the expansion of mechanisms allowing data subject to opt out. (Gonçalves, 2017). Ultimately, the regulation should proceed from the need for a fair and reasonable balance between the interests of data subjects and the controllers, based on mutual confidence and accountability.

References

Adjerid, I. and Kelley, K., (2018). Big data in Psychology: A Framework for Research Advancement. *American Psychologist*. 73(7), pp. 899–917, doi: 10.1037/amp0000190.

Bottis, M. and Bouchagiar, G., (2018). Personal Data v. Big data: Challenges of Commodification of Personal Data. *Open Journal of Philosophy*, 8, pp. 206–215, doi: 10.4236/ojpp.2018.83015.

Bouchagiar, G. and Bottis, M., (2018). Personal Data Protection Models: Aspects of Ownership. *16th International Conference e-Society 2018*. Available at: https://ssrn.com/abstract=3167011 [Accessed 06.04.2024].

Chander, A., (2017). The Racist Algorithm? *Michigan Law Review*, 115, pp. 1023–1045.

Custers, B. and Vrabec, H., (2024). Tell me something new: data subject rights applied to inferred data and profiles. *Computer Law & Security Review*, 52, 105956, doi: 10.1016/j.clsr.2024.105956.

Davis, P. and Schwemer, S.F., (2023). Rethinking Decisions Under Article 22 of the GDPR: Implications for Semi-Automated Legal Decision-Making. *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023)*. Available at: https://ssrn.com/abstract=4478107, doi: 10.2139/ssrn.4478107 [Accessed 06.04.2024].

Day, P., (2020). Cambridge Analytica and Voter Privacy. *Georgetown Law Technology Review*, 4.2, pp. 583–607.

Fischer, C., (2020). The legal protection against inferences drawn by AI under the GDPR. July 2020. Available at: https://arno.uvt.nl/show.cgi?fid=151926 [Accessed 06.04.2024].

Gonçalves, M.E., (2017). The EU Data Protection Reform and the Challenges of Big data: Remaining Uncertainties and Ways Forward. *Information & Communication Technology Law*. 26(2), pp. 90–115, doi: 10.1080/13600834.2017.1295838.

Malgieri, G. and Comandé, G., (2017). Sensitive-by-distance: quasi-health data in the algorithmic era. *Information & Communications Technology Law*, 26(3), pp. 229–249, doi: 10.1080/13600834.2017.1335468.

Minbaleev, A.V. and Storozhakova, E.E., (2023). Problems of legal protection of personal data in the process of using neural networks. *Courier of Kutafin Moscow State Law University (MSAL)*, 2, pp. 71–79, doi: 10.17803/2311-5998.2023.102.2.071-079. (In Russ.).

Naumann, F., (2014). Data Profiling Revisited. *ACM SIGMOD Record*, February 2014, doi: 10.1145/2590989.2590995.

Niševic, M., (2020). Profiling Consumers Through Big data Analytics: Strengths and Weaknesses of Article 22 GDPR. *Global Privacy Law Review*, 1(2), pp. 104–115.

Roig, A., (2017). Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). *European Journal of Law and Technology*, 8(3).

Savelyev, A.I., (2015). The Issues of Implementing Legislation on Personal Data in the Era of Big data. *Pravo. Zhurnal Vysshey shkoly ekonomiki*, 1, pp. 43–66. (In Russ.).

Vinogradova, E.V., Polyakova, T.A. and Minbaleev, A.V., (2021). Digital profile: the concept, regulatory mechanisms and enforcement problems. *Law Enforcement Review,* 5(4), pp. 5–19, doi: 10.52468/2542-1514.2021.5(4).5-19. (In Russ.).

Wachter, S. and Mittelstadt, B., (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big data and AI. *Columbia Business Law Review*, 2, pp. 494–620.

Westerlund, M., Isabelle, D.A. and Leminen, S., (2021). The Acceptance of Digital Surveillance in an Age of Big data. *Technology Innovation Management Review*, 11(3), pp. 32–44.

Wiedemann, K., (2022). Profiling and (automated) decision-making under the GDPR: A two-step approach. *Computer Law & Security Review*, 45, 105662, doi: 10.1016/j.clsr.2022.105662.

Information about the Author

Artur N. Mochalov, Cand. Sci. (Law), Associate Professor, Department of Constitutional Law, Ural State Law University named after V.F. Yakovlev, Yekaterinburg, Russian Federation
artur.mochalov@usla.ru
ORCID: 0000-0003-2502-559X