Section Four.
Internet, New Technologies and Privacy

Privacy by Design – Searching for the Balance Between Privacy, Personal Data Protection and Development of Artificial Intelligence Systems[1]

Zanda Davida <zanda.davida@hotmail.com>
RIGA, Latvia

Dominik Lubasz <dominik.lubasz@lubasziwspolnicy.pl>
ŁÓDŹ, Poland

Abstract
The growing use of artificial intelligence systems has put them under the regulatory spotlight all around the world. The EU considers to regulate artificial intelligence systems as a part of initiative of creating ethical and legal framework for trustworthy artificial intelligence. It is no secret that data is the fuel of artificial intelligence systems, but the problem with the insufficient protection of data subjects is remaining. The article aims to analyze the interaction between the GDPR and the draft of Artificial Intelligence Act and search for the balance between privacy, personal data protection and development of artificial intelligence systems. The article analyses the legal framework and compare many guidance documents issued by the international organisations. The authors reveal that the draft of Artificial Intelligence Act does not contain instruments for the insufficient protection of data subjects, especially in aspects concerning control and transparency, and that it lacks the promised horizontality of the draft legislation and the creation of a legal framework for all artificial intelligence systems and not just selected ones. Moreover, the article proposes solutions on how to minimize the privacy risks associated with the development and use of artificial intelligence systems. It suggests to develop further guidance and regulation and to consider a more horizontal approach.

1 The research leading to this publication was supported by the National Science Centre (*Narodowe Centrum Nauki*) in Poland on the basis of decision no. 2018/31/B/HS5/01169.

Keywords:
data protection, privacy, privacy by design, artificial intelligence, AI systems, data protection by design.

1. Introduction

With the expansion of artificial intelligence, trust deficits in both the technology and its developers are becoming an important issue. Concerns arise about the application of technology in an instrumental way, arbitrary or discriminatory effects of automated decisions made using it. As new technologies gain importance and become easier to implement, it becomes necessary to analyse whether these technologies may violate the law or ethical standards. One of the main areas of concern is the area of data protection and privacy implications, especially since the demand for data due to the development and use of AI-based solutions is greater than ever. In this context, the availability, quality, and quantity of data is an issue of concern, as they are the basis for self-learning systems to fulfil their original purpose, i.e. to find relationships between information allowing to draw specific conclusions, make decisions, and through them, influence the environment. The accuracy of the algorithms themselves is also not without significance. Algorithm bias can independently generate cognitive deficiencies that intensify the negative effects on data subjects.[2].

Thus, there is a need to identify solutions to minimize the risks related to the use of techniques based on artificial intelligence, particularly from the perspective of the possibility of a discriminatory effect, harming human dignity and privacy, and leading to restrictions on freedom of expression, access to information, and manipulation of opinions.[3].

---

2 On the subject of potential risks, see further: Tjerk Timana and Zoltan Manna (ed) 'Data Protection in The Era of Artificial Intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies' (GDVA, 2019), 11. Overview of trends in AI guidelines in different countries and organizations: see: Anna Jobin and Marcello Ienca, 'Artificial Intelligence: the global landscape of ethics guidelines' (2019) AL/Digital ethics project <https://www.researchgate.net/publication/334082218_Artificial_Intelligence_the_global_landscape_of_ethics_guidelines> accessed 22 June 2021.

3 ibid; Advisory Board on Artificial Intelligence and Human Society, Report on Artificial Intelligence and Human Society, (2017) <https://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_en.pdf> accessed 22 June 2021.

The need to build, or essentially support the recovery of trust in technology was also seen as an important element. Trust in technology or trustworthiness of technology ultimately became a key element in the search for a target regulatory framework.

In the field of privacy and data protection, for the legal assessment of the social and ethical implications of AI, it was obvious to reach for the mechanisms developed in the creation of 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)[4]. It had to be assessed whether legal instruments such as risk-based approach, privacy by design, data protection by design, data protection impact assessment or privacy impact assessment more broadly could form the basis for a legal framework for artificial intelligence. This assessment had to be done in the context of a European AI strategy supporting "the creation of ethical, secure and state-of-the-art AI solutions in Europe" based on three pillars: (i) increasing public and private investment in AI for its wider deployment, (ii) preparing for socio-economic change, and (iii) providing an appropriate ethical and legal framework to strengthen European values[5]. This vision is presented in the communications issued by the European Commission on 25.4.2018. "Artificial Intelligence for Europe".[6] and in the "Coordinated

4 European Parliament, 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 [2016] OJ L199/1–88 (General Data Protection Regulation).

5 See also European Commission, 'Member States and Commission to work together to boost artificial intelligence „made in Europe"' (Press release, 7 December 2018) <http://europa.eu/rapid/press-release_IP-18-6689_en.htm> accessed 6 March 2019. See more Agnieszka Jabłonowska, Maciej Kuziemski, Anna Maria Nowak, Hans-Wolfgang Micklitz, Przemysław Pałka and Giovanni Sartor, 'Consumer law and artificial intelligence. Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence' (EUI Working Papers, 2018) 4-11; Sandra Wachter and Brent Mittelstandt, 'A Right to Reasonable Interferences: R-thinking Data Protection Law in the Age of Big Data and Ai' (2019) Columbia Business Law Review 1.

6 Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Artificial Intelligence for Europe' (Communication) COM (2018) 237 final.

Plan on Artificial Intelligence" of 7.12.2018.[7], as well as in the "Declaration of Cooperation on AI" signed on 10.4.2018.[8] and in the White Paper On Artificial Intelligence – A European approach to excellence and trust issued on 19.2.2020 r.[9]

The first result of the work initiated in the above documents was the establishment of the High-Level Expert Group on AI (HLEG), which was first tasked with developing ethics guidelines for trustworthy artificial intelligence, followed by policy and investment recommendations. In the guidelines developed in April 2019, HLEG pointed out the need to create conditions for the development of human-centric artificial intelligence in Europe primarily by giving it the characteristic of "trustworthy" artificial intelligence. A trustworthy artificial intelligence should have certain characteristics[10], which revolve around providing guarantees of autonomy and control, as well as protection, to human beings subjected to the influence of AI-enabled processes. These principles are:

1. Human Agency and Oversight;
2. Technical Robustness and Safety;
3. Privacy and Data Governance;
4. Transparency;
5. Diversity, Non-discrimination and Fairness;
6. Societal and Environmental Well-being;
7. Accountability[11].

---

7 Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 'A coordinated plan for artificial intelligence' (Communication) COM (2018) 795 final.

8 European Commission, 'EU member states sign up to cooperate on artificial intelligence' (News, 8 March 2020) <https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence> accessed 6 March 2019.

9 Commission, 'White Paper on Artificial Intelligence – A European approach to excellence and trust' (Communication), COM (2020) 65 final.

10 European Commission, 'Assessment list of trustworthy artificial intelligence' (Study, 1 March 2021) <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> accessed 27 April 2021.

11 To assess compliance with these principles, the High Level Expert Group on Artificial Intelligence has developed and made available the ALTAI tool – High Level Expert Group on Artificial Intelligence, 'The Assessment List on Trustworthy Artificial Intelligence' <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence> accessed 27 April 2021.

Viewed individually, each of these characteristics is necessary but not sufficient to achieve trustworthy artificial intelligence. Under ideal conditions, all seven characteristics interact harmoniously with each other, and their scopes overlap. However, if in practice it turns out that the interactions between these features lead to conflicts, society should make efforts to correct them accordingly.

The analysis carried out in the European Union led to the presentation by the European Commission on 21 April 2021 of a draft regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts[12]. This Regulation is to lay down harmonised rules on artificial intelligence, providing rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain applications of remote biometric identification systems. Non-discrimination legal solutions are also to be implemented by introducing requirements aimed at minimising the risk of algorithmic discrimination, in particular with regard to the design and quality of datasets used for developing AI systems, complemented by obligations for testing, risk management, documentation and human oversight throughout the life cycle of AI systems. Although the Regulation on AI is intended to be horizontal and, despite its broad definition of AI systems, it only regulates certain aspects related to the operation of AI systems and only certain systems. This is also important in the context of the planned relationship of this regulation to the GDPR. According to paragraph 1.2 of Explanatory Memorandum, the AI Regulation is only intended to supplement the General Data Protection Regulation (Regulation (EU) 2016/679) and the Law Enforcement Directive (Directive (EU) 2016/680) with a set of harmonized rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain applications of remote biometric identification systems.

Under EU law, the GDPR is thus to remain the basis for assessing the horizontal compatibility of AI systems in the area of personal data. The objectives of this legal act, which are to ensure a high level of protection of individuals' rights and a technology-neutral approach in which the implementation of the requirements is based on a risk analysis of the processing from the perspective of the rights and freedoms of data subjects, and the

12 Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM 2021 206 final.

elements of proper implementation of the regulation are not only the security of processing (Article 32 RODO) but also the principles of processing (Article 5 RODO), such as transparency, fairness, data minimisation and purpose limitation[13], are in line with the trends in the development of technology and the assumptions of the legal framework of trustworthy and human-centric artificial intelligence defined in the above mentioned documents[14]. Also the legal mechanisms and instruments used in this act, which allow the assessment of technical solutions from the perspective of their impact, both at the design stage as well as during implementation and use, on data subjects, i.e. the obligation to take into account data protection by design, data protection by default or to conduct a data protection impact assessment, which have a technology-neutral mechanism, allows the assessment of fit for purpose.

2. *Concepts of privacy by design and data protection by design*

The original concept of privacy by design was originally created by Ann Cavoukian, during her time as Privacy Commissioner for the State of Ontario[15]. This concept is the result of work to consolidate the practice of incorporating privacy protection into new infrastructure projects currently underway in Canada and as a specific, both philosophical and practical response to the difficulties of guaranteeing adequate privacy protection in

13 See more: Piotr Drobek in Edyta Bielak-Jomaa and Dominik Lubasz (eds) *RODO. Ogólne rozporządzenie o ochronie danych. Komentarz* (Wolters Kluwer 2017) 340.

14 Dominik Lubasz and Katarzyna Witkowska, 'Europejska reforma ochrony danych osobowych z perspektywy pełnomocnika przedsiębiorcy' in Kinga Flaga-Gieruszyńska, Jacek Gołaczyński and Dariusz Szostek (eds) *Media elektroniczne. Współczesne problemy prawne* (C. H. Beck 2016) 176.

15 IPC, 'Privacy by Design. The 7 Foundational Principles' <https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf> accessed 23 March 2021. See also Dominik Lubasz and KatarzynaWitkowska, *RODO. Ogólne rozporządzenie o ochronie danych. Komentarz* (LEX/el., comments on Art. 25 No 5), see also Wojciech R. Wiewiórowski, 'Privacy by design jako paradygmat ochrony prywatności' in Grażyna Szpor and WojciechR. Wiewiórowski (eds) *Internet. Prawno-informatyczne problemy sieci, portali i e-usług* (C. H. Beck 2012) 13–29 and the literature referred to therein, and Michał Bienias, 'Ochrona danych w fazie projektowania oraz domyślna ochrona danych (privacy by design oraz privacy by default) w ogólnym rozporządzeniu o ochronie danych' in Grzegorz Sibiga (ed) *Ogólne rozporządzenie o ochronie danych. Aktualne problemy prawnej ochrony danych osobowych 2016* (2016) 20 Monitor Prawniczy 53–57.

light of rapidly developing technology[16]. It is based on seven principles: proactive rather than reactive ; privacy as the default setting; privacy embedded into design ; full functionality understood as achieving positive sum, not zero sum; protection of privacy from the beginning to the end of the information life cycle; visibility and transparency and respect for user privacy[17].

A strategy for privacy by design in the implementation of new technological solutions was also presented by ENISA (European Union Agency for Network and Information Security) in its report Privacy and Data Protection by Design. It indicated two possible approaches, i.e. data-oriented - aiming at limiting the negative impact by, among others, data minimisation, separation or generalization, and process-oriented concerning mainly organizational aspects and procedures ensuring the realization of the right to autonomy, in particular by informing data subjects, enabling data control, enforcing protection and, finally, demonstrating compliance[18].

These approaches allow to put the data subject in the centre of attention when designing personal data processing processes with the use of modern technologies, including artificial intelligence, especially because of their technological neutrality. The applied procedure of focusing the evaluation perspective on a human being allows, at the same time, to realize the postulate of striving to regain trust by technology and to include mechanisms to ensure this in the design (trust by design)[19].

---

16 It is worth noting that in the context of artificial intelligence, A. Cavoukian modified the original concept indicating the need for ethical construction of tools using artificial intelligence mechanisms (AI Ethics by design). The core elements of this concept have become: transparency and accountability of algorithms; application of ethical principles to the processing of personal data; ensuring oversight and accountability for the performance of algorithms; respect for privacy as a fundamental human right; data protection as a default setting; proactive identification of security risks, thereby minimising risks; robust documentation to facilitate ethical design and data symmetry - Ann Cavoukian, 'Ethics by design' <www.ryerson.ca/pbdce> accessed 23 March 2021.

17 This concept was subsequently adopted in the Resolution on Privacy – Design Data Protection and Privacy Commissioners, 'The Resolution on Privacy by Design' (32nd International Conference of Data Protection and Privacy Commissioners, Jerusalem, 27–29 October 2010. <http://www.giodo.gov.pl/pl/1520084/3830> accessed 13 May 2021.

18 ENISA, 'Privacy and Data Protection by Design – from policy to engineering' <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design> accessed 23 March 2021.

19 Witold Chomiczewski and Dominik Lubasz, 'Privacy by design a sztuczna inteligencja' (2020) 20 MoP 67 ff.

The concept of privacy by design implemented as a regulatory instrument in Article 25 of the General Data Protection Regulation shows some particularities resulting from the subject of protection related to the scope of its provisions, namely the protection of natural persons in relation to the processing of their personal data (data protection by design). According to the adopted structure, Article 25 RODO imposes an obligation on controllers to implement appropriate technical and organisational measures designed to effectively implement data protection principles, in particular the principle of data minimisation and to provide the processing with the necessary safeguards in order to meet the requirements of the General Data Protection Regulation, and in particular to protect the rights of data subjects in a specific context of processing. The context of the processing, on the other hand, is to be determined taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of the processing, which are to be the basis for examining the likelihood of risks of the violation of the rights or freedoms of natural persons by the processing under consideration, in order to assess which technical or organisational measures should be implemented to mitigate those risks, both in determining the modalities of the processing and at the time of the processing itself[20].

Notwithstanding the above requirements formulated in Article 25(1) of the GDPR, Article 25(2) adopts a principle that was originally part of the privacy by design concept of Ann Cavoukian, i.e. privacy by default, in the form of the data protection by default principle. According to this principle, the controller is obliged to implement appropriate technical and organisational measures to ensure that, by default, only those personal data are processed that are necessary for each specific purpose of the processing. This obligation relates to the amount of personal data to be collected, the extent of their processing, their storage period and their availability.

The analytical process underlying this obligation shall have a multistage character, starting from determining the full context of the processing, i.e. defining the assumptions and determining the circumstances, the scope, the tool-layer and the manner of performing the data operations, through assessing the impact of the above factors on the rights and freedoms of data subjects to be processed, in particular whether any undesirable effects

20 European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (20 October 2020) <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf> accessed 13 May 2021.

of the processing occur or are likely to occur in relation to the proposed processing from the perspective of the data subject, e.g. discrimination. This assessment shall take place in the context of the proper implementation of the principles of processing contained in Article 5 of the GDPR, in particular the principle of data minimisation and the rights of data subjects (Article 25(2) and 12-22 of the GDPR), and finally the analysis of the adequacy of organisational and technical measures to ensure the implementation of the principles of processing and the protection of the rights of data subjects, including ensuring their security.

3. Functional use of data protection by design model in the design of AI systems

The assumptions of the privacy by design concept adopted in Article 25 of the GDPR, in the form of data protection by design requirements, referring in particular to the assessment perspective, i.e. humonocentric approach in the performance of obligations, taking into account principles such as reliability, transparency and, in particular, data minimisation, and finally, the necessity to ensure data subjects' control over their data prima facie, seem to correspond to the basic problematic issues formulated when designing solutions using artificial intelligence. They address regulatory concerns about the application of technology in an instrumental way, arbitrary or discriminatory effects of automated decisions made with the use of AI, both in the context of the designed algorithms and the possibility of their corruption (algorithm bias), as well as the data necessary to teach AI to make decisions in the designed area, which is primarily related to the availability, quality and quantity of input data[21]. The importance of the issue of using this regulation in an increasingly digital world is highlighted by the European Data Protection Board in its Guidelines 4/2019 on the application of the Article 25 principle of data protection by design, pointing out that this principle plays a key role in promoting privacy and the protection of personal data in society. For these reasons, it is important that controllers take this responsibility seriously and implement the obligations

21 Slomit Yanisky-Ravid and Sean K. Hallisey, 'Equality and Privacy by Design: Ensuring Artificial Intelligence (AI) is Properly Trained & Fed: A New Model of AI Data Transparency & Certification as Safe Harbor Procedures' <https://ssrn.com/abstract=3278490> accessed 13 May 2021.

under the General Data Protection Regulation when designing processing operations[22].

The regulation of Article 25 of the GDPR applies to all controllers regardless of their size, organisational complexity or the level of complexity of the planned processing operations. The provision of Article 25(1) of the GDPR implies first of all that compliance with the data processing requirements is to be an equivalent goal to the business purposes for a controller designing new processing operations[23]. Data protection must therefore be embedded first in the R&D project and then in the implementation and maintenance and be an inherent part of the whole process in its individual phases. This principle is therefore not only about the compliance aspect, but also, and perhaps above all, about developing a specific organisational culture of working on new solutions to ensure compliance[24].

The primary obligation is to implement appropriate measures and necessary safeguards to ensure the effective implementation of data protection principles and, consequently, the rights and freedoms of data subjects. Article 25 sets out both design and default elements to be taken into account, addressing the context, the nature of the processing purposes and scope, as well as the state of the art and the cost of implementation, and the risks to data subjects, of varying probability and severity. The analysis of these elements should be made at an early stage of planning a new processing operation and repeated during the processing, through regular reviews of the effectiveness of the chosen measures and safeguards.

As emphasised by the European Data Protection Board in the abovementioned Guidelines, effectiveness is at the heart of the concept of data protection by design. The requirement for effective implementation means that any measure should produce the intended results in terms of the processing designed by the controller from the perspective of the data subject. For this reason, the provision of Article 25 does not introduce a catalogue of required measures and leaves the decision on adequacy from an effectiveness perspective to the controller. Whether specific measures are effective will therefore depend on the context of the processing in question and an assessment of the relevant elements to be taken into

22 European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (20 October 2020) <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf> accessed 13 May 2021.

23 Monika Susałko in Dominik Lubasz (ed) *Meritum Ochrona danych osobowych* (Wolters Kluwer, 2020) 227.

24 Chomiczewski and Lubasz (n 19) 67 ff.

account when determining the modalities of the processing. This approach implies that controllers should be able to demonstrate that they have ensured adequacy and that the implemented measures and safeguards achieve the desired effect in terms of personal data protection by minimizing the risks for data subjects related to the envisaged forms of processing. To this end, as underlined by the European Data Protection Board, the controller may define appropriate key performance indicators (KPIs), or provide a justification of its assessment of the effectiveness of the chosen measures and safeguards, in order to demonstrate their effectiveness, in line with the accountability principle (Article 5(2))[25].

In developing a model approach for the use of data protection by design instruments in the design of IS systems, reference can be made to already developed guidelines[26]. Particularly noteworthy is the concept developed by the Norwegian Data Protection Supervisory Authority in its guidelines on Software Development with Data Protection by Design and by Default[27]. This is primarily supported by the AI definition proposed in the draft AI Regulation, which indicates that 'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with[28]. Consequently, the logic of proceeding in the design and development of AI systems will be similar to that presented in the aforementioned Guidelines, taking into account and adapting the model to the particular characteristics of AI.

In the above mentioned Guidelines, the implementation plan is divided into 7 phases:

1. Training,
2. Requirements,
3. Design,

---

25 European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (20 October 2020) <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf> accessed 13 May 2021 7.

26 ibid.

27 Norwegian data protection supervisory authority, 'Software development with Data Protection by Design and by Default' (Guidelines) <https://www.datatilsynet.no/en/about-privacy/virksomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/?print=true> accessed 13 May 2021.

28 ibid, Art. 3(3).

4. Coding,
5. Testing,
6. Release
7. Maintenance.

The key to the correct implementation of the presented approach in the development of IS systems is the prior definition of the nature and purpose of the project. It is important to determine the full context of the activity, and from the point of view of personal data protection, the processing of data, the determination of the scope of data, the sources of their acquisition and the manner, needs, and therefore the purpose of processing, the manner of performing particular operations, the tools used, both internal and provided by external entities. Defining the project framework in the above-mentioned scope shall allow to perform the compliance analysis and the risk analysis of the designed solution in the subsequent steps of planning and implementation.

4. Training

In the presented model, the first phase is a training allowing, in accordance with data protection by design requirements, to determine the level of knowledge, including personal data protection and processing security and cyber security, which needs to be absorbed in the organisation in order to be able to properly carry out the subsequent steps. The end result is to reach, by means of appropriate staff qualification improvement, a state in which, when starting to work on a new solution, everyone in the organisation understands both the need for and the risk of data protection and security, and knows what requirements apply, what they should pay attention to and what tools enable them to transform their knowledge of data protection and information security into instruments, technical and organisational measures that secure them. This applies both to internal requirements, including policies, procedures, including risk assessment procedures, and to external requirements, in particular relevant legislation in the area of personal data and information security or cyber security[29]. It is important to consider that external requirements may include sectoral regulations. For example, the requirement to comply with best practices, standards, code of conduct for the chosen technology, business practices.

29 (n 27).

The data subject is entitled to rely on the due and professional care of the AI system developer. Namely, if there is a generally accepted good practice in the industry, then the data subject can be sure that this will also be taken into account in the training activity. The exception may be where sectoral rules (e.g. business practices) are not mandatory, in which case the AI system developer may ignore them, but must sufficiently inform the data subject that normal business practices have not been followed.

5. Requirements

The next phase is the identification of data protection and information security requirements for the final product. The correctness and completeness of the requirements identification will have a significant impact on the correctness of the end result, and from this perspective it is a key step. In order to define correct requirements, it is necessary to establish the context, i.e. to determine the data requirements starting from the scope, sources, the categories of data subjects, the identification of the user and the owner, i.e. the future controller, as well as the further entities involved in the processing, i.e. the processors and other recipients. This will allow to identify the relevant legislation, both general and sectoral, guidelines, applicable codes of conduct, norms and standards. The analysis of the requirements, as in the first stage, must be both external and internal and from the perspective of personal data protection include regulations, in particular provisions of the General Data Protection Regulation, but also business practices and policies, internal procedures such as control, audit, compliance procedures, etc.

In relation to the implementation of data protection by design model, among the legal requirements at the forefront is ensuring compliance of the solution to be developed with the principles of personal data processing formulated in Article 5, i.e. the principles of lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality. Already in this phase of the analysis the controller will have to demonstrate its assessment as well as its decision on how to ensure compliance in accordance with the principle of accountability, which should be included in the compliance procedure by creating appropriate documentation, check lists, etc.

From the perspective of the construction of AI-based mechanisms, the intensity of compliance difficulties may vary, but in particular the verification of the principle of legality, fairness, transparency, data minimisation, and security embedded in the principle of confidentiality and integrity

will be crucial[30]. This is also indirectly related to the direction of the conceptual work on the legal framework for AI indicated in the European Commission and HLEG studies related to the expectation of constructing trustworthy human-centric artificial intelligence.

The primary issue relates to the requirement for AI developers to ensure compliance with the principle of lawfulness and fairness. It is not limited to the determination of an adequate legal basis, but also requires that AI algorithms or models are constructed in such a way as to ensure the correctness and non-discriminatory character of the processing or the effect of such operations, as well as to exclude the possibility of basing on them biased or detrimental automated decisions affecting the rights and freedoms of data subjects. It is an element of the assumption of the necessity to take into account the interests of the data subject in shaping the processing, including the expectation that the interference with privacy caused by the data processing is not excessive. Consequently, it results from the controller's obligation to introduce measures to prevent arbitrary and discriminatory treatment of data subjects and to implement solutions and self-learning mechanisms to exclude from the processing which is the basis of the decision data which are incorrect, unduly processed or inadequately processed, as well as those which ensure the correction of factors causing the inaccuracy of personal data and which will be oriented at the maximum reduction of the risk of errors and at securing personal data in a way which takes into account the potential risks for the interests and rights of the data subject. This applies both to the algorithms themselves and to the data used , in particular their quantity and quality.

The transparency principle, on the other hand, is aimed at ensuring that the data subject is aware of the purpose, scope and context of the processing, in order to enable him to exercise control over his own data[31]. The challenges in ensuring compliance with the principle of transparency are mainly related to the assessment of the extent to which it is possible to explain how AI systems work, from the perspective of the person whose

30 Dominik Lubasz and Monika Namysłowska in Dominik Lubasz (ed) *Meritum Ochrona danych osobowych* (Wolters Kluwer, 2020) 1013.

31 The WP29 indicates that consent should also specify the consequences of the processing - see WP29, 'Opinion 15/2011 on the definition of consent' (13 July 2011, WP 187), <http://www.giodo.gov.pl/pl/file/5341> accessed 13 May 2021 18. See also Arwid Mednis, 'Cechy zgody na przetwarzanie danych osobowych w opinii Grupy Roboczej Art. 29 dyrektywy 95/46 Nr 15/2011 (WP 187)' in Grzegorz Sibiga (ed) *Aktualne problemy prawnej ochrony danych osobowych* (2012) 7 MoP 26. Łuczak (n 13) 466.

data are to be processed. In specific cases, it may prove difficult to provide sufficient information to the data subject, mainly when the AI system is based on deep learning, unsupervised or only partially supervised, when learning is not based on methods using symbolic reasoning principles. The selection of methods to underpin the operation of AI mechanisms, taking into account the requirement of their explainability, is consequently one of the elements of assessing compliance with the principle of transparency. This is because the choice of a particular technological solution should not exclude a priori the possibility of implementing this principle already at the design stage. Technological difficulties cannot exempt one from the obligation of transparency. Therefore, the core of the issue is to find a solution that will allow the proper exercise of data subjects' rights and to provide them with relevant information about the principles of operation of the algorithm so that the data subject can understand the scope and the consequences of the processing of his/her data as well as contest the decisions taken with regard to him/her.

The assessment of the compliance of the AI systems with the data protection by design requirements will also be influenced by the designated purpose of the data processing, which determines the individual parameters of the processing, such as the scope of the data or the time limits, underlying the principle of purpose limitation and data minimisation. The identification and indication of the purpose must therefore take place before the processing starts, and properly informing the data subject about it is one of the elements that allow the data subject to act within his autonomy and control over the processing. The purpose will be different for the different phases of the development of AI systems, starting from the learning phase, validation, testing and finally implementation. It has a significant impact because the correlation between the purpose and the scope of data resulting from the data minimisation principle determines which data fall under the notion of data necessary to fulfil the purpose and which do not. This scope will vary from one phase to another, and will also be case-specific. In the design phase of the AI the main difficulty lies in the fact that at the stage of data acquisition its developers are not always able to predict, especially in models that are not based on symbolic principles of reasoning, how much data will be necessary to achieve a satisfactory learning outcome of the AI in order for it to reach practical applicability[32]. Thus, it can be extremely difficult to draw the line between adequate and inadequate data. It may be even more difficult to demonstrate this

32 Chomiczewski and Lubasz, (n 19) 67ff.

relationship based on measurable criteria. The identification of solutions can start with analysing the possibility of using prepared data and test environments in the learning phase of the AI, and in the application phase performing regular reviews of the area of its operation and adjusting the scope of the data as it changes[33].

The scope of the data as one of the factors also affects the risks of the processing to the data subjects, the more data and individual personal information is collected about the data subject and the more individuals are assessed in the processing, the greater the risks to those individuals[34]. Consequently, it is necessary to search for adequate safeguards corresponding to the increased risk, the level of which must increase as the risk increases. These safeguards are subject to the confidentiality and integrity principle enshrined in Article 5(1)(f). The construction of this principle is based on a risk-based approach and is concretised in particular in the provisions of Articles 25 and 32 of the GDPR. Its essence is the obligation to ensure that personal data are processed in a way providing adequate security, including protection against accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

Security requirements will be determined by identifying the threats to which AI systems may be exposed, the correlated vulnerabilities of these threats, and the likelihood and severity of the impact on data subjects. These factors influence the parameters for selecting adequate and appropriate security measures. In order to identify risks and make decisions on mitigating them, it is necessary to conduct a risk analysis focused on identifying and addressing the above-mentioned factors affecting the level of risk. This assessment shall be made from the perspective of the impact of the risk on the data subjects. Furthermore, it shall take into account the state of the art, the cost of implementation, and the nature, scope, context and purposes of the processing affecting the risk of violation of the rights or freedoms of natural persons with varying degrees of likelihood and severity arising from the processing by the particular system. The flexibility of the choice of measures is therefore limited by their adequacy to the potential risks of the processing operations and in particular to their security[35].

---

33 ibid.

34 ICO, Guidance on AI and data protection, op. cit., p. 60.

35 Norwegian data protection supervisory authority, 'Software development with Data Protection by Design and by Default' (Guidelines) <https://www.datatilsynet.no/en/about-privacy/virksomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/?print=true> accessed 13 May 2021.

The proactive identification of security risks, a feature still derived from Ann Cavoukian's original privacy by design principle, and thus the minimisation of risks, technical robustness, are important qualities when designing AI-based solutions, in the data protection by design model of Article 25 of the GDPR.[36]

The specific nature of the technologies used in connection with the development of AI systems, taking into account the requirements of Article 35(1), (3), and (4) of the GDPR, makes it obligatory in some cases, and advisable in others, to implement the specific requirement of conducting a data protection impact assessment. It consists in assessing the impact that the envisaged solution or processing operation may have on the rights and freedoms of natural persons whose data are or will be used. Deepening the analysis in this mode is one of the elements that may be key to prove the fulfilment of the principle of fairness, which, as it was mentioned, requires the limitation of the negative impact of the processing on data subjects.

Equally important from the perspective of the legal requirements that AI developers need to take into account is the issue of designing software, algorithms, and processing, including decision-making processes using AI mechanisms, to ensure that the rights of data subjects, as enshrined in Articles 12-22 of the GDPR, are respected[37]. This includes both the appropriate design of the information policy, in particular in the context of the proper implementation of the principle of transparency, and the preparation for the exercise of specific rights under the GDPR, i.e. the rights of access to data, rectification of data, erasure of data, restriction of processing, data portability, objection and not to be subject to decisions based solely on automated processing of data. In the latter case, the possible authorisation of decisions based solely on automated processing of data, including profiling, with the effects of Article 22 requires the fulfilment of additional requirements of legitimacy and an in-depth analysis.

Notwithstanding the above regulations, from a privacy and personal data protection perspective, the requirements formulated in the draft regulation presented by the European Commission on 21 April 2021 establishing harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts will have to be taken into account

---

36 High Level Expert Group on Artificial Intelligence, 'The Assessment List on Trustworthy Artificial Intelligence' <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence> accessed 27 April 2021, *op. cit*., p. 20.

37 Susałko (n 23) 230.

in the future[38]. As I mentioned, this regulation is intended to lay down harmonised rules on artificial intelligence, providing for rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain applications of remote biometric identification systems. If this Regulation is finally adopted, there will therefore be new requirements for certain AI systems, additional to the GDPR, as the AI Regulation is only intended to complement and not replace the regulation of the General Data Protection Regulation.

In the context of the scope of the analysis, if this is the final form of the AI Act, an analysis will have to be made as to whether the proposed solution does not fall under the black list of AI-related practices which are prohibited in Article 5 of the AI Regulation. Subsequently, it will also have to be analysed whether the designed AI system will not qualify as a high-risk AI system, as referred to in Article 6 of the draft AI Regulation. In the case of such qualification, a number of additional obligations to be imposed on both manufacturers and users of IS systems will be actualised.

At the end of the deliberations concerning the verification of the requirements to be met by the designed AI system, it is worth returning to the HLEG guidelines on trustworthy and ethical artificial intelligence, and the follow-up to these guidelines in the form of The Assessment List For Trustworthy Artificial Intelligence (ALTAI) For Self Assessment, which provides a system for verifying the compliance of designed AI systems with the principles of trustworthy and ethical artificial intelligence formulated by HLEG and indicated above. In order to facilitate the analysis, a web application for assessing compliance has also been prepared - ALTAI - The Assessment List on Trustworthy Artificial Intelligence[39].

6. Design

In the design phase the mechanisms that address the diagnosed requirements should be taken into account and appropriately designed, in parti-

38 Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM 2021 206 final.

39 High Level Expert Group on Artificial Intelligence, 'The Assessment List on Trustworthy Artificial Intelligence' <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence> accessed 27 April 2021.

cular by ensuring compliance with the implementation of the principles of processing, security of processing and the rights of data subjects. In the last mentioned point it is not only about ensuring that these rights are respected and fulfilled, but also that these rights are easily realisable, accessible and intuitive, and that the impact of the processing is limited to what is necessary to achieve reliable results.[40]

These requirements must be precisely reflected in the design. It needs to be assessed and designed by which means these requirements are to be achieved. The means must be adequate for each specific requirement, and this adequacy can be measured e.g. using instruments such as data protection impact assessments. Among the measures proposed by the Norwegian supervisory authority in the aforementioned study on the principle of data protection by design are, inter alia, measures such as minimisation and limitation of the processing and scope of data, security, storage separation, data aggregation, default data protection, i.e. configuring privacy settings in such a way that they are most conducive to ensuring privacy by default.

AI system developer should implement appropriate technical and organisational measures to ensure that, by default, only personal data which are necessary for each specific purpose of the processing are processed. AI system developer and user are accountable for implementing default processing settings and options in a way that only processing that is strictly necessary to achieve the set, lawful purpose is carried out by default. The basic requirement is that data protection is built into the processing by default.

Process requirements for design, on the other hand, include informing, i.e. designing an appropriate level of transparency of the processing for the data subject, ensuring control through the right of access, update, erasure of one's own data, enabling the exercise of rights, and finally designing in such a way that the controller can document how the requirements of the General Data Protection Regulation have been implemented. It is important to emphasize that the General Data Protection Regulation aims to protect not only privacy, but also fairness and other fundamental rights including private life. Therefore data protection law is protecting people also against abuse of information asymmetry.[41] The position of the data subject in relation to data controllers is improved by the transparency of

40 Chomiczewski and Lubasz, (n 19) 67 ff.

41 Paul De Hert and Serge Gutwirth, 'Privacy, data protection and law enforcement: Opacity of the individual and transparency of power' in Erik Claes, Antony Duff and Serge Gutwirth (eds) *Privacy and the Criminal Law* (Intersentia, 2006).

processing activities.[42] The principle of transparency is one of the key principles also in consumer law. Therefore, the inadequate compliance with the requirements may lead to both – breach of data protection law and consumer law. The GDPR contains more detailed transparency obligations than the EU consumer law. Nevertheless, requirements for design must be fulfilled in a broad manner because data are important not only for data protection but also for consumer policy. For example, AI system developer must see the distinction between the terms and conditions and the privacy policy. It must be clear to the data subject and communicated in plain and intelligible language that terms of use of AI system are related to the rights and obligations of the data subject as consumer and the trader under the contract, while a privacy policy provides information about what the AI system does with the personal data[43]. It must also be ensured that data subjects are motivated to read by encouraging interaction. The more interaction is provided, the more data subject choice would be meaningful. Similar methods are used in consumer law.[44] The GDPR makes explicit reference to pictograms – „standardized icons in order to give in an easily visible, intelligible and clearly legible way a meaningful overview of the intended processing".

The design phase is the last moment when AI system developer must objectively assess usage methods regarding to protect specific groups of the data subjects, namely vulnerable persons including children. For example, profiling children should be avoided. It is prohibited to do direct exhortation to children.

7. Coding

A project that takes into account the defined requirements, both data- and process-oriented, enables it to be coded correctly. The knowledge base, absorbed in the organisation and then designed in detail taking into account all requirements, should be translated into an appropriate execution level in which all designed requirements are addressed and then tested.

---

42 Natali Helberger, Frederik Zuiderveen Borgesisu, and Agustin Reyna, 'The Perfect Match? A Closer Look at the Relationship between EU Consumer Law and Data Protection Law' (2017) 54, 5 Common Market Law Review 11.

43 General Data Protection Regulation, Recital 42.

44 BEUC, 'EU Consumer Protection 2.0 Structural Asymmetries in Digital Consumer Markets' (March 2021) <https://www.beuc.eu/publications/beuc-x-2021-018_eu_consumer_protection.0_0.pdf> accessed 13 May 2021.

In this context, consideration must also be given to excluding the use of specific functions and modules that may prove to be unsafe. Furthermore, the guidelines point out the need to perform static code analysis and code reviews on a regular basis, which ensures that guidelines for secure coding are being followed and can be measured to ensure controls are working.

8. Testing

The testing phase is aimed at verifying whether the data protection and information security requirements have been implemented as planned and properly met. In the case of software, it is indicated that it needs to be tested for vulnerabilities using dynamic tests, fuzz tests and penetration tests[45]. It is important to clearly define the requirement for testing for data protection and information security, in particular in high risk and specific AI systems, for example facial recognition, product safety components, immigration and border control AI systems[46]. In line with a risk-based approach, high-risk AI systems will need to include the implementation of adequate risk correlated mitigation measures. The data subject has a legitimate right to expect that higher data protection requirements will be met in the testing of higher risk AI.

9. Release

A successfully completed testing phase allows the decision to launch the product on the market. In this phase, it is important to plan not only the sales and marketing strategy, including communication, but also the strategy for managing incidents that may occur after the release and the procedure for updating the designed and implemented security features. At release phase it is also important to have strategy how to react to individual complaints of data subjects. The AI system basically operates

45 Norwegian data protection supervisory authority, 'Software development with Data Protection by Design and by Default' (Guidelines) <https://www.datatilsynet.no/en/about-privacy/virksomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/?print=true> accessed 13 May 2021.

46 Irena Nesterova, 'Mass data gathering and surveillance: the fight against facial recognition technology in the globalized world' (SHS Web of Conferences 74, 03006, 2020) 6.

independently, so human supervision is essential in specific and individual cases that may indicate general data security risks regarding AI system.

Finally, it is the responsibility of the AI system controller to demonstrate compliance with data protection principles. If the principle of accountability is not effectively implemented, trustworthy AI cannot be achieved.

10. Maintenance

The last-mentioned elements ultimately become the goal of the maintenance phase. In this phase, business continuity, incident and security breach management plans are implemented, the purpose of which is to guarantee the ability to continuously ensure confidentiality, integrity, availability and resilience of the processing systems and services, as well as the ability to quickly restore the availability of and access to personal data in the event of a physical or technical incident. On the basis of observed incidents, work is to be undertaken to ensure that identified non-conformities are resolved, as well as to initiate further product development activities. For these activities, the same strategy should be applied as in the development of the project based on the requirements of data protection by design, starting with the development of the concept and the context allowing the identification of knowledge needs, and therefore the training stage, as the first stage of the cycle. This process is not a one-off, but a continuous process in which it is necessary to update the conclusions, taking into account changing factors, above all those related to the context of processing.

11. Summary and conclusions

The risks related to the development and application of solutions using artificial intelligence mechanisms, as well as the search for and definition of protective objectives, especially in the area of fundamental rights, in particular the right to privacy and the protection of personal data, clearly directs the analysis, among other things, to the search for and evaluation of the fit for purpose of existing legal regulations. In the area of personal data protection, which is also a fundamental right according to Article 8 of the CFR, and in the context of the will to create human-oriented artificial intelligence, the potential should be seen in the provisions of Regulation 2016/679, at the regulatory basis of which lies the will to ensure a high

level of protection of the rights of natural persons through, inter alia, the legal instruments data protection by design and data protection by default[47]. Objective of taking into account the perspective of the data subject in the creation process, as an immanent element of the assessment, and at the same time will allow to perform the analysis from the perspective of the principle provided for in the GDPR, to ensure the realization of the rights of data subjects and the security of the processing.

The adoption and process-oriented implementation in the development of new technological solutions of a personal data protection strategy in the design phase based on Article 25(1) and (2) of the GDPR, using both data-centric and process-oriented approaches, addresses the main concerns related to the application of technology in an instrumental way, arbitrary or discriminatory effects of automated decisions made using AI. It allows reflection, definition of requirements and appropriate design of solutions to avoid discriminatory bias of algorithms as well as problems of availability, quality and quantity of input data, through the use of a federated learning model[48]. At the same time, applying this method to AI projects should allow trust in the technology to be rebuilt and this goal to be woven into the design (trust by design), which is ultimately a key element in creating human-centric AI.

In the context of the submission by the European Commission on 21 April 2021 of a draft regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts[49], the regulatory environment for artificial intelligence will evolve. However, if the strategy of building a trustworthy human-centric AI in the European Union based, inter alia, on the principles of human agency and oversight, technical robustness and safety, transparency, diversity, non-discrimination and fairness is upheld, the GDPR will remain the only leading regulation in this area, and data protection by design

47 High Level Expert Group on Artificial Intelligence, 'The Assessment List on Trustworthy Artificial Intelligence' <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence> accessed 27 April 2021, p. 26.

48 Shlomit Yanisky-Ravid and Sean K. Hallisey, 'Equality and Privacy by Design: Ensuring Artificial Intelligence (AI) is Properly Trained & Fed: A New Model of AI Data Transparency & Certification as Safe Harbor Procedures' <https://ssrn.com/abstract=3278490> accessed 13 May 2021.

49 Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM 2021 206 final.

mechanisms will have to be embedded in the design and use of AI. This corresponds with the outlined intentions of the proposed AI Regulation, which is merely to complement the General Data Protection Regulation with a set of harmonised rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain applications of remote biometric identification systems.

The discussion on the regulation of artificial intelligence is now going to gain momentum, and the European Commission's proposal should be seen as an encouragement for this. However, voices are already being raised that the draft does not contain instruments for the insufficient protection of data subjects, especially in aspects concerning control and transparency, and that it lacks the promised horizontality of the draft legislation and the creation of a legal framework for all AI systems and not just selected ones.

There is an urgent need to develop further guidance and regulation and to consider a more horizontal approach. While Europe is still debating this, the effective use of data protection instruments in the design phase of AI systems is becoming increasingly important. The article proposes solutions on how to minimize the privacy risks associated with the development and use of AI systems.