Dániel Eszteri*

Analysis of machine learning techniques in the context of GDPR's provisions on automated decision making and data subjects' rights

I. Introduction

In this article I attempt to present certain issues of artificial intelligence, especially machine learning and the protection of personal data. The General Data Protection Regulation of the European Union (Regulation (EU) 679/2016, GDPR), devotes separate provisions to decision-making affecting personal data exclusively through an automated path. Decisions made by AI have by now become part of our everyday life and the European legislature recognised its social significance and the need for regulation. From the data protection perspective one of the most important areas of the operation of the AI is the phenomenon of 'machine learning', when the software 'learns' based on the input data and makes various decisions. These decisions made by a machine may have an impact on a given person's private life.

The objective of this paper is to shed light on the fact that decisions are going to be made by various software applications in the near future that may even affect us without any human intervention. In most cases these software applications will use nothing but our personal data to make these decisions. Therefore, I wish to call the attention to the social impact of AI and machine learning and the related data protection regulatory problems. As knowing the directions of technological development, the study, analysis and assessment of this field from a legal point of view is very likely to gain significance in the future.

The article starts with a general introduction to the social impact of artificial intelligence (AI) and machine learning. Then it outlines certain issues of machine learning and related data processing relevant for data protection law. In the second half of the paper, I present the relevant provisions of the GDPR and some questions and answers related to their applicability.

II. The concept of artificial intelligence and its impact on social development

There is no uniform concept of artificial intelligence, although the relevant literature includes several attempts at defining it. According to the Oxford Dictionary of Computer Science, AI is the field of computer science which addresses the production of computer programmes to resolve tasks requiring human intelligence.[1] According to the glossary of the AI page in the Council of European website, artificial intelligence is a set of scientific results, theories and techniques whose ultimate purpose is to render machines capable of reproducing the cognitive capabilities of humans. In several areas, the objective of current development projects is to get machines carry out tasks requiring complex thinking which had earlier been done by man.[2] In the opinion of John McCarthy - who approaches the problem primarily from an engineering viewpoint - AI is the science and engineering practice of producing intelligent machines.[3] Therefore, the common point is primarily endowing machinery and instruments, and the programmes and software operating them with intelligence.

The following table illustrates the differences and similarities between a simple programme and AI.[4]

* Ph.D., head of Data Breach Notification Unit, Hungarian National Authority for Data Protection and Freedom of Information (Nemzeti Adatvédelmi és Információszabadság Hatóság), lecturer in data protection law at Eötvös Lóránd University Faculty of Law and National University of Public Service (Hungary).

'Simple' program	Artificial intelligence
- Written by a programmer. - Deterministic: it gives the same answer to the same question at all times. - Gives a Yes-No, 0-1 type result. - The programmer tells (in advance) what the correct result is. - It runs rules, there is no possibility to overwrite the rules.	- Written by a programmer. - It works with probability: there is a certain probability that the answer to the given question will be the same. - Less-more: e.g. 85%-15% type result is given. - The programmer only sets the objective, AI experiments to find the correct results. - It studies samples. - Surprises and shifts in the emphasis may occur.

As far as the history of philosophy is concerned, Stuart Russel and Peter Norvig distinguish four schools in the philosophical development of the AI concept, which are the following:

(1) *System thinking as humans do:* This school regards systems that model the operation and cognition of the human mind as AI.

(2) *System acting as humans do:* This approach is linked to the name of the mathematician Alan Matheson Turing, who held human behaviour as the main criteria of intelligence and the objective to be achieved based on the Turing test named after him.

(3) *Rationally thinking system:* This trend regards the goal of AI development as the creation of machines and software that are more perfect and more rational than human thinking in some sense.

(4) *Rationally acting system:* This is the approach of the modern IT sciences, whose objective is not to have the developed systems think in the classical meaning of the word, or imitate human behaviour, but to behave as rationally as possible (e.g. they should be able to clearly diagnose diseases or forecast natural disasters).[5]

According to John R. Searle, it is necessary to distinguish between weak and strong AI. Searle regards those systems as *weak AI*, which act as if they were intelligent, but despite this there is no information whether or not the system truly has a mind of its own. In contrast, a strong AI refers to systems, which truly think and have an independent mind.[6]

Mankind has forever been preoccupied with the problem of the impact of "created reason" on society. Generally, what lies behind the individual stories is a deeply rooted ancient fear that an artificial being wakening to an independent consciousness could even destroy its own creator. The roots of these pessimistic schools can be found also in pre-20th century literature and in folklore (e.g. the Frankenstein story).

The conflict between humans and artificial beings appeared not only at the level of literary fiction. For instance, the dread of machines gave birth to the Luddite movement during the industrial revolution in the 1810s.[7] The pessimistic theories according to which the created being revolts and destroys its creator are based primarily on the problem of the so-called "technological singularity". According to Ray Kurzweil, the singularity is a future point in time in which the rate of technological change will be so rapid and its effect so deep that human life will be irreversibly transformed. Because of the superhuman intelligence appearing as a result of the singularity, technological development and, in relation to this, social change accelerates so much that those living before the singularity are unable to comprehend the change in the environment. In other words, according to this theory, we are unable to forecast the events after the technological singularity or intelligence explosion with our current images of the future.

And the superintelligent artificial beings appearing in the wake of the singularity can easily crowd out man from existence.[8]

According to other, more optimistic theories the vision of AI overpowering humans stems from the fear of the unknown, just as earlier the fear of ghosts or witches. According to the optimistic approach, if AI is designed appropriately, that is, as agents that achieve the objectives of their masters, then AI arising from the step-by-step progress of current design will serve rather than overpower. People use their intelligence aggressively because they are born with aggressivity due to natural selection. But the machines, which we ourselves build, are not born aggressive, unless we decide to build them as such. On the other hand, it is possible that computers conquer us in the sense that their service becomes indispensable, just as cars conquered the industrialised world.[9]

The points above clearly reveal that the ideal social and the related legal regulation of AI is a question of extraordinary importance and its significance will only grow as time advances.

III. The relationship of artificial intelligence with the protection of personal data

These days, AI-based systems, software applications and devices used on a daily basis provide new types of solutions, which in many cases involve the processing of the personal data of the users. The domestic robots intended for household use or the smart phone applications, which analyse human behaviour, continuously monitor the behaviour and reactions of their users in order to serve their needs as perfectly as possible. It is not fortuitous that personalisation is the keyword for devices and services employing such modern technological solutions in almost every case.

In addition to being personalised, there is increasing demand for technologies, which are capable of forecasting the needs of the user. This presupposes much more complex decision-making mechanisms, which can be best achieved by AI-based self-learning systems. The relevant report of the Norwegian Data Protection Authority (Datatilsynet) describes AI as a system capable of learning based on its own experience and to apply the knowledge obtained in different situations to resolve complex problems. The heart of the concept is that AI learns from the personal data it "sees" (in practice the input data) and makes decisions or "forecasts".[10]

In the sections below, I attempt to review the main issues in the relationship between AI and personal data.

1. Artificial intelligence and machine learning

The above three terms are frequently used as synonyms, although they do not fully overlap. AI serves as an umbrella term, which covers all the procedures when a software makes a decision automatically. Relative to this, machine learning is a narrower concept, which means one branch of AI development. The heart of this is that the system generates independent knowledge out of its own experience. Based on data examples and patterns, the system is able to recognise and determine regularities and rules independently or with human assistance and then makes decisions based on the regularities discovered in the acquired knowledge base.[11]

In the remaining part of the study, I primarily analyse problems not from the viewpoint of AI taken in a broad sense, but from that of the phenomenon of machine learning included in it as this gives rise to the perhaps most important questions to be answered in relation to data protection. Hereinafter wherever I use the term AI, in fact I mean machine learning.

2. The basis of data processing in the course of machine learning

According to the Norwegian Data Protection Authority's report, data processing carried out by an AI system in the course of machine learning can be divided into three steps as follows:

(1) First, a large quantity of so called 'test data' is input in the system and the algorithm tries to find patterns and similarities in this data set. If the algorithm finds identifiable patterns, it will record and save them for subsequent use. After this, the system generates a model on the basis of the recorded and saved patterns. Based on the already identified patterns, with the help of the model, the system is capable of processing the subsequent input data.

(2) After this, the AI system operates as follows: first, new data are uploaded in the system, which are similar to the data used for learning. Then, based on the model, the AI decides which new data are similar to which learned pattern.

(3) Finally, the system provides information on the decision it made based on the acquired patterns in relation to the new input data. [12]

A good example for the systems operating based on the above logic is the development of self-driving cars as the appropriate operation of such technologies is based on processing of a large quantity of personal data. The software operating a self-driving car must, for instance, be able to distinguish between living human beings crossing the roads and walking on the sidewalks and inanimate objects (e.g. a newspaper or bag left on the road) in order to avoid accidents. This is achieved in the course of developing the software operating the self-driving car by having the system analyse a great many photos and videos made of (several tens or even hundreds of thousands) of people, which after a while learns what human beings look like and how they move based on their similar physical and physiological characteristics. Later, when the car eventually perceives a pedestrian crossing the street, it will recognise it as a human being, even if it never met the photo of that person while teaching the software.[13]

It is also important to note that the model generated in the course of machine learning does not necessarily contain the source data, which served as the basis of its learning. In most cases, the AI system generated in the course of machine learning is able to operate independently of the data that served as the basis of learning.[14]

3. The quantity and quality of the data used for machine learning

Characteristically, machine learning requires a much larger quantity of raw data than the human brain does in order to be able to efficiently identify patterns and to set up decision-making models on their basis. So, at first, we might think that the more data we have, the better AI systems we can produce. Yet, the quality of the data used for machine learning, their appropriate prior selection and labelling are much more important aspects. Even before inputting the data in the system, it is necessary to clarify the exact purpose of using the data to carry out specific tasks and because of this, the range of the data used must be restricted to those relevant for the given purpose. The correct selection and preliminary choice of the data used is also a very important criterion.[15]

See the following example: a hospital in the USA wishes to categorise risks linked to patients suffering from pneumonia with an algorithm written for this purpose. In the course of the test, a surprising result was found, namely, that from the viewpoint of recovery the system regularly put patients to the lowest risk category who also suffered from asthma in addition to pneumonia. This was due to the fact that such patients characteristically received much faster and much more thorough care than the others, hence they recovered more quickly. This example shows well that the use of inappropriately selected data or data containing only some parts of the information may produce misleading results.[16]

If we wish to determine exactly how many data are needed for efficient machine learning appropriate to the purpose using the principle of data minimisation[17] referred to in GDPR from the viewpoint of protecting personal data, the following principle can be formulated: at first,

only a restricted quantity of test data may be used for machine learning. After uploading the test data, the system must be continuously monitored to see how accurately and how efficiently it operates from the viewpoint of the purpose to be achieved. If less data are sufficient to achieve the appropriate efficiency, it is unnecessary to use additional ones.

4. The "black box" and the problem of transparency

One of the most frequently expressed concerns in relation to machine learning (and not only from the viewpoint of data protection) is that often it is impossible to predict what sort of result the system will produce. The model applied may produce a result, for which seemingly no explanation exists.[18] This phenomenon is referred to in machine learning as "black box". For an ordinary observer, the system works in practice by absorbing data on the input side, on the basis of which it learns something, then it produces some result. It is, however, extremely difficult to see why exactly it generated that result.

In scientific and technical fields, the black box is a device, system or object, which can only be examined on the basis of its input, output and transmission characteristics, its concrete internal operation is unknown, that is, its implementation is 'opaque' (black). Virtually anything can be referred to as a black box: a transistor, an algorithm or the human mind. The description of the phenomenon covering the concept stems from Wilhelm Cauer, who developed his theory concerning this in 1941, although he did not use the concept at the time. Later, his disciples described the phenomenon as the "black box" analysis.[19] The opposite of the black box is a system where the internal components or logic are accessible to examination. Such a system is sometimes referred to as "white box".

The requirement that data processing must be transparent for the natural person data subject (whose data are being processed) has been included among the principles of data protection for a long time. This principle is expressly named by GDPR in Article 5(1)(a). Accordingly, personal data shall be processed lawfully, fairly and *in a transparent manner in relation to the data subject*. That is to say GDPR names the principles of lawfulness, fairness and transparency at the same time, hence they must be enforced in relation to every data processing operation with respect to one another and simultaneously.

A question therefore is how systems using machine learning can be set up so that they operate with sufficient transparency for the data subject from the viewpoint of the results they produce and they comply with the requirement of transparency with regard to the personal data processed.

According to the report of the Norwegian Data Protection Authority, the principle of operation used by the machine learning system plays an important role from the viewpoint of transparency. By way of example, the report refers to two principles of operation, one of which is termed the decision tree model and the other is the neural network model.[20] The decision tree shows the various decision-making options taking the eventual consequences, chances, usefulness and resources into account, depending on what it is used for. Decision trees are graphs mathematically.[21]

Relative to this, a neural network consists of at least three parts that can be well separated both in terms of function and structure: the input layer, the hidden layers and the output layer. The input layer forwards the data transferred as input to the other parts of the network without modification. A neural network may have several input layers. The hidden layers are positioned between the input and the output, their task is the transformation and coding of information and the creation of abstractions, interim representations. Their number and type, the order of their connection to one another and the number of neurons within them are variable parameters of the network. Finally, the output layer indicates the result. The output function and the number of the output neurons are determined by the nature of the given problem.[22]

The number of branches or layers may change by system in both the decision tree and the neural network models. A good example of this is that Microsoft's engineers won an image recognition challenge in 2015 using a neural network consisting of 152 layers.[23] The size of the network and the connections between the individual layers may render data processing tasks so complex that cannot be understood by humans, even data scientists. They just do not know what happens in the black box. Because of this, undertakings developing data processing based on AI solutions may be seriously challenged by the legal requirement of transparency in that regard.

The primary issue in relation to these data processing methods is whether it complies with the principle of transparency, if the controller informs the data subject only of the general principle of operation used by the AI software or whether it is expected to provide much more detailed information going into the technical details of the operation of the software, attempting also to outline the possible results, in other words, whether it is necessary to open the black box? I will attempt to analyse the information that is appropriate according to GDPR in the subsequent parts of this study.

5. The problem of the need for human intervention

The French Data Protection Authority (Commission nationale de l'informatique et des libertés, in brief: CNIL) published a report in December 2017, which studied a few ethical issues raised by the use of AI and the algorithms.[24] In relation to responsibility, the report poses the philosophical question of who can be accountable for the behaviour of AI. The report brings up a medical example for a case when the possibility may arise of making the algorithm itself accountable for the acts it takes.

According to the report, French law requires that a diagnosis of a disease may only be established by a medical doctor. If any other person carries out such activity, it qualifies as quackery and being unlawful it may raise the suspicion of a crime. Nowadays, however, doctors are frequently assisted by complicated software applications in setting up the accurate diagnosis. Because of this, the software setting up a more accurate diagnosis could even take over the actual role of the doctor as one is inclined to formally accept that the decision is made by the algorithm after a while. And in relation to this, the eventual "accountability" of the software itself may arise in connection with the decision.[25]

Although the report does not put forward a specific recommendation in relation to the issue of accountability, yet it notes in general that in the case of making decisions with a critical (legal) effect, the ideal case would be if the AI systems ensured the possibility of human intervention and/or review. In the case of "mass decisions" requiring particularly deep consideration, the possibility could be left in the hands of the algorithm. In other words, CNIL regards the scale of deployment of AI in decision-making as the greatest challenge.[26]

The need for human intervention appears also among the GDPR provisions. I will present these later.

IV. GDPR notions relevant for this subject matter

Systems based on machine learning are used to make decisions related to personal data with increasing frequency. The personalised advertisements on the Internet and other contents are good examples of how the algorithms operate that analyse human behaviour and learn from it, and how they use our personal data to display more and more personalised targeted advertisements and contents. The notion of automated decision-making is closely related to profiling as the increasingly unique profile of the given person evolves along decisions made by the algorithm. In addition to personalised advertisements, automated decision-making

frequently plays a part in preparing loan assessments and forecasting purchasing power and related profiling. For these decisions, the systems need a large quantity of sufficiently accurate personal data for optimal operation and decision-making.[27]

The notions of automated decision-making and, in close relationship to this, of profiling specifically appear in the GDPR; Article 22 sets forth specific provisions in relation to this.

The Article 29 Data Protection Working Party (abbreviated as WP29), which can be regarded as the legal predecessor of the European Data Protection Board, comprising the data protection supervisory authorities of the Member States of the European Union, published its guidelines on automated decision-making and profiling on 3 October 2017.[28] Below, I present the relevant parts in the GDPR, then I address how the legal requirements can be applied to data protection problems raised by AI and machine learning.

1. The notion of profiling in GDPR

The machine learning systems outlined in the sections above are frequently used to analyse user behaviour and to create user profiles in this way. The profile of a given user may provide information on personal characteristics and personality traits on the basis of which a person may even be clearly identified.[29]

Among its interpreting provisions, GDPR provides a definition of profiling. According to this, profiling means any form of automated processing of personal data consisting of the use of the personal data to evaluate certain personal aspects relating to a natural person. In particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.[30]

Although it may be envisaged that profiling taken in the classical sense could be carried out by a controller even fully manually, such activities are not included in this notion. That is to say, in GDPR terminology profiling means activities based on automated processing of personal data or activities that relies on that. Profiling must include some form of automated data processing; this, however, does not fully exclude the possibility of human intervention according to the WP29 Guidelines.[31] It is important that not exclusively automated decisions may also include profiling.

A decisive element for GDPR terminology is that the purpose of profiling is the assessment of the personal characteristics of a natural person.

According to the guidance, it can generally be said that profiling means the collection of information on a natural person (or a group of natural persons) and the assessment of their characteristics or behavioural patterns with a view to classify them into certain categories or groups. The purpose of the classification is to analyse the range of interests, the expected behaviour or certain capabilities of the data subjects.[32]

This term can be applied also to systems using machine learning because the algorithmic analysis of the personal data input into such systems and drawing conclusions from them may be aimed at profiling purposes.

2. The appearance of automated decision-making in GDPR

The other key notion for the data protection analysis of machine learning is automated decision-making. The study of the relevant EU regulation however reveals that neither Article 22 of GDPR, nor the definitions provided define the notion of automated decision-making, although the term is used by the regulation in several places.

GDPR's wording refers to profiling and automated decision-making together in several places, and sets forth common rules concerning them. It is important to note that irrespective

of this, the two notions are not fully identical. There may be an automated decision-making procedure which does not qualify as profiling and profiling may be carried out without incorporating automated decision-making mechanisms. In most cases, however, the two notions go hand-in-hand and supplement one another, thus discussing them together may be warranted from a data protection point of view. The WP29 guidance captures this duality so that the scope of automated decision-making is different and may partially cover profiling or it may stem from it. According to the guidance, automated decision-making is a capability of making decisions with the help of technological instruments without human intervention.[33]

That is to say, there is no human involvement in decision-making in the case of exclusively automated decision-making, which phenomenon may correspond to the step in machine learning when the decision is made concerning the input data as a result of the processes taking place in the "black box".

V. The general prohibition of automated individual decision-making in GDPR

Below, I will review the provisions concerning automated individual decision-making *only* because the social impact of machine learning on the data subjects is the most spectacular in the case of these data processing activities. Article 22 of GDPR addresses these data processing operations in detail.

Pursuant to Article 22(1) of GDPR, the data subject shall have 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 has similarly significantly impact on him or her*. This provision constitutes a general prohibition on decision-making based exclusively on automated data processing. The regulation includes profiling based on such a decision-making process. This prohibition stands irrespective of whether or not the data subject takes any measure concerning the processing of his/her personal data. Therefore, as a main rule GDPR sets forth a general prohibition on exclusively automated individual decision-making, which has legal effect or similarly significant impact on the data subject.[34]

In order to qualify an activity as human intervention with regard to the decision and therefore the general prohibition of Article 22 should not apply to it; the controller has to ensure that the human review of the decision be of merit and not only a symbolic gesture. It has to be done by a person, who has the authority and the competence to alter the decision. All the relevant data have to be taken into account as part of the analysis.[35] In other words, to be exempt from the prohibition, the final decision must be made by a human being, or the decision proposed by the algorithm has to be reviewed and approved by them.

Furthermore, the rules applicable to exclusively automated decision-making have to be applied only in the cases when the decision has legal effects or similar significant impact on the natural person data subject. GDPR does not define the notions of "legal effect" or "similarly significant", but this wording of the regulation makes it clear that Article 22 extends only to effects involving severe consequences.[36]

The legal effect requires that the machine's decision should influence the legal rights of a person. A legal effect may be something that will influence the legal standing of a person or their rights based on contract. According to WP29, examples of such effects include automated decisions concerning natural persons as a result of which: contracts are terminated, welfare benefits (such as child-related benefits or housing support) guaranteed by law are granted or rejected, entry to a country is refused or citizenship is denied.[37]

The effect of automated decision-making on the rights of people set forth in law or contract concerns cases that can be relatively clearly delineated. In addition, however, there is the more vaguely worded "similarly significant" impact in Article 22 of GDPR, which is also a circumstance subject to the prohibition.

Recital (71) of GDPR may contain some guidance concerning this notion as it lists the following examples: "refusal of an online credit application" or "e-recruiting practices without any human intervention".

It is difficult to accurately determine what should be regarded as of "*significance*" in order to reach the threshold; according to WP29, however, the following decisions may be included in this category: decisions influencing an individual's financial circumstances, such as those concerning his entitlement to credit; decisions which influence an individual's access to health care services; decisions which deny a person the opportunity to be employed, or expose the person to severe disadvantage; decisions which influence access to education, such as university admission.[38]

According to WP29, automated decisions concerning targeted advertisements based on online consumer profiling most of the time do not have similarly significant impact on natural persons (e.g. advertisements of clothing). Yet, there are certain data processing operations even in this category, which may have a significant effect on certain groups of society, such as adults in an exposed situation. For instance, if based on the profile generated a person presumably struggling with financial difficulties is still regularly targeted with advertisements on high interest loans, will potentially aggregate additional debts (provided that he accepts loan offered).[39] The general prohibition of Article 22 shall apply to such cases. According to the main rule, a profile generated about a consumer struggling with financial difficulties (through machine learning) cannot be used for the purpose to target him in an attempt to induce him to take on additional financial risk. The profilers conducting the data processing cannot refer to the fact that taking out the loan is independent of them, as it is the decision of the data subject, since the profiling on which the consumer decision is based is not lawful.

VI. Exemptions from the prohibition of automated individual decision-making in GDPR

As described above, Article 22(1) stipulates a general prohibition on exclusively automated individual decision-making that has a legal effect or similarly significant impact. There are, however, exemptions from this general prohibition set forth in Article 22(2). Accordingly, the prohibition cannot be applied, if the decision is:

a. necessary for entering into or performing a contract between the data subject and the data controller;
b. *authorised by European Union or Member State law,* to which the data controller is subject and which also lays down suitable measures to safeguard the data subject' rights and freedoms and legitimate interests; or
c. based on the data subject's explicit *consent.*

The first exemption is the performance of a contract, on the basis of which controllers may apply automated decision-making processes for purposes related to the contract in a legal relationship that comes into being through the contract. According to WP29, in such a case the controller has to be able to demonstrate that the use of automated decision-making is the most appropriate method of data processing to achieve the purposes specified in the contract. If the purpose specified in the contract can be achieved using another method that will not qualify as necessary.

The second exemption is when automated decision-making in relation to the given data processing operation is made possible by European Union or Member State law. The relevant legislation must also lay down suitable measures to safeguard the data subject' rights and freedoms and legitimate interests. According to Recital (71) of GDPR, such a case may be for instance when the law authorises the state to use automated decision-making mechanisms in order to prevent fraud and tax evasion.

Finally, the third exemption is when the use of automated decision-making is based on the expressed consent of the data subject.[40]

GDPR itself does not define the notion of "explicit consent", but another guidance of WP29 related to consent provides guidelines for its interpretation as follows.

A clearest method of gaining assurance that the consent was expressed is the reinforcement of the consent in a written statement. A signed statement, however, is not the only way to obtain express consent. According to WP29, in a digital or online context it may happen for instance that the data subject can issue the required statement by completing an electronic form, sending an e-mail or uploading a scanned document containing his signature or using an electronic signature. These may comply with the requirement of explicit consent. In theory, even an oral statement may be sufficiently clear to obtain valid express consent. Characteristically, in such cases the controller has difficulties demonstrating that all the conditions of valid express consent were met at the time of recording the statement, hence the recommended form of obtaining an express consent is in writing or recorded in some other form. Finally, one can gain assurance of the validity of the express consent through the two-step control of the consent (a good example of this is the use of two-factor authentication).[41]

VII. The data subject's rights in relation to automated individual decision-making

1. Right to be informed and to access

GDPR requires the controller to provide information in relation to decision-making based exclusively on automated data processing having legal effects or similarly significant impact. The regulation includes profiling based on such data processing in this range.[42] Under this, the following three items of information must be communicated with the data subject:

a. He must be informed of the fact of such data processing;
b. He must be given information of merit on the logic applied; and
c. Finally, he must be informed of the significance of the data processing and its expected consequences for the data subject.[43]

The communication of the fact of automated individual decision-making is a relatively simple requirement; it suffices if the controller provides information that such data processing is taking place. It is important that the data subject must also be aware if automated individual decision-making also implies profiling.

The mode of providing information on the logic applied raises a number of issues. This may be a substantial challenge in the case of the machine learning methods presented in the sections above for the controller as that is frequently based on exceedingly complex data processing tasks that are very difficult to review. An excellent example of this is the phenomenon of the black box.

According to GDPR, controllers must provide "information of merit" on the logic applied. If a controller communicates only in general that it is operating a system based on a neural network may not be sufficient in itself as the data subject will have little idea of what is happening with his personal data in the course of processing.

The information of merit, however, does not necessarily mean that the controller should provide complicated explanations about the algorithm applied or present the algorithm in full. A detailed presentation of the technology would, in most cases, decrease the comprehensibility of the information and impede its reception.[44] Naturally, the complexity of the technology cannot be an excuse for fully waiving the information. The true purpose and importance of the legal institution of informing the data subject can be best realised precisely in these cases. It follows from the spirit of GDPR that data subjects must be aware of the essence and background to the more complicated data processing operations affecting their personal data.

Despite the above, at first sight providing information may seem to be an insoluble problem for controllers because of the nature of machine learning and the black box presented in the sections above. How could the user or operator of AI provide information of merit to the data subject when he does not know what precisely is happening with the data in the black box?

According to the joint research by the Oxford Internet Institute and the Alan Turing Institute of London, it may be a good practice from the viewpoint of the transparency of the decisions made by the algorithms and the related data processing, when the controller provides an opportunity for the data subject to learn the operation of data processing in relation to the so-called "alternative interpretations". The reason for this is that most of the time data subjects are not really interested in how the logic operates, they are much more interested in how they themselves can improve the result established by the algorithm.[45]

The study presenting the research of the British institutions presents the following example. The risk assessment software applied in the course of the evaluation of a loan application arrives at the conclusion that the data subject cannot get the loan. In such a case, if the data subject asks for information on what the reason is for the negative evaluation based on the logic applied, then information concerning the general logic, on the basis of which the software operates and arrives at such a conclusion would not carry all that much additional information for him. What would mean information much more to the point would be, if a public testing system was available, which data subjects can use by entering even fictitious data. In this way, they can have sufficient information of merit about the conclusions that the system arrived at based on the input data. According to the study, such a solution is advantageous also because it respects the rights of the AI developer concerning business secrets and intellectual property in line with recital (63) of the GDPR.[46] Similar solution was used for instance by Google in the case of its machine learning system called TensorFlow.[47]

In addition to the above, it is necessary to note that the controller has to inform the data subject also about the "significance" and "expected consequences" of data processing. According to the WP29 guidance, in order that this information be of merit and comprehensible, real and tangible examples of possible effects should be given. In a digital context, controllers may also use additional instruments to present such effects and may apply visual techniques to explain a former decision. In such a case, the guidance, similarly to the British study, gives the example of providing a comparable application.[48] In other words, when providing information, the controller need not open the "black box" to the data subject, it is sufficient, if it makes the data subject understand how the decision was made and what he can do in order to have a different (more favourable) decision in his case.[49]

In my view, these interpretations practically underline that making a test system accessible is a good solution not only for the applied logic, but also for demonstrating the effects on and significance for the data subject.

2. Requesting human intervention

Pursuant to the GDPR provisions, the controller is under an obligation to enable the data subject to request human intervention in relation to the automated individual decision-making affecting him. This opportunity must be ensured by the controller.[50]

That is to say, in this case human review of the course of the decision-making must be enabled as well as the modification of the decision in the given case in order to eliminate eventual defective conclusions and errors. The data subject is entitled to express his position concerning the decision and to object to it in order to find remedy to the erroneous decision affecting him.[51]

The WP29 guidance underlines that human intervention must be carried out by a person who has the appropriate authority and capability to change the decision. The reviewer has to

thoroughly examine all the relevant data, including any additional information made available by the data subject. In addition, both the guidance and the commentary on GDPR highlight that one of the purposes of the human review of the decision-making may be that based on the result of the decision, the data subjects should not be exposed to detrimental discrimination.[52] The guidance recommends the incorporation of periodic auditing algorithms as a general practical solution in relation to the prohibition of detrimental discrimination, although unfortunately it does not provide a specific example.[53]

Granting the request for human intervention gives rise to a number of questions in practice in relation to the use of AI-based decision-making systems. According to certain criticism, this may be very difficult to implement, particularly in the case of services, which use a dynamic pricing system.[54] In my view - to stick to this example - when the user makes use of such a service and accepts the purchase price calculated through dynamic pricing, after the contract is made and the purchase price is transferred, requesting human intervention can be excluded most of the time because of the provisions of the contract. Realistically, in such cases, human intervention should be possible prior to the contract entering into force, because the decision made on the basis of personal data (the calculation of the price based on the profile of the data subject) is made and it is available well before it is accepted.

In my view, when exercising the request for human intervention, the review of the decision-making mechanisms should be approached from the viewpoint of the effect on the data subject. The person designated to carry out human intervention must examine exactly what data were included in the decision-making and what decision was made and then this would have to be compared to the objections submitted by the data subject. In such a case there is no need to open the 'black box', or the thorough exploration and understanding of the algorithmic process behind the decision. A person authorised to review would not really have a genuine opportunity for this in most cases, because as discussed above, mathematical operations carried out on the data cannot be followed by a human observer beyond a certain level. But the person authorised to review should consider first and foremost whether the decision would have the same result, if it was not made by an algorithm.

VIII. Summary: When does a system based on machine learning operate lawfully?

I attempted to find an answer to the question of what problems arise from AI including machine learning processes from the viewpoint of the protection of personal data. I studied the problem from the viewpoint of the data subject with respect to transparency, purpose limitation, data minimisation and the possibility of human intervention. Below, I attempt to underline the main items, which must be taken into account in the course of the development of AI systems based on machine learning with a view to compliance with the principle of data protection by design and default as referred to in Article 25 of GDPR to guarantee lawfulness.

a. When operating a system based on machine learning, capable of making automated decisions, the controller operating the system has to verify an appropriate legal basis for processing personal data. In the absence of this, the operation of such a system is prohibited.

b. In the course of the development of a system based on machine learning, the least possible quantity of personal data should be used for testing. The efficiency of the system should be continuously monitored and more personal data may be input into the system only with a view to achieving optimal efficiency.

c. In an ideal case, the transparency of an already active system using genuine personal data should be ensured by providing a test system for elaborating 'alternative decisions'

for the data subjects. Using this system, the data subjects can themselves discover the logic applied in the course of the operation of AI.

d. Data subjects should have an opportunity to request human intervention in relation to decisions made exclusively on an algorithmic basis. The person authorised to review has first and foremost to consider whether the result of the decision would be the same if it was not made by an algorithm. In the course of reviewing the decision, what needs to be considered first and foremost is the personal data provided by the data subject and the purpose of the operation of the machine learning system.

To assess the lawfulness of AI and machine learning from the viewpoint of data protection, I came to the above distinction, which sets forth requirements mainly from the viewpoint of transparency and human intervention in addition to legal basis and purpose limitation. In my view, a system capable of making automated decision, which adopts decisions concerning the data subjects based on the analysis of personal data, must in all cases comply with this set of criteria.

I am aware that other distinctions can also be developed, for instance based on the effect on (other) fundamental rights. They would include, for instance, the study of the prohibition of detrimental discrimination which, however, I did not wish to elaborate in detail in this study. Nevertheless, I hope that this writing contributes to the overview of the subject and helps the discourse in this field.

---

[1] László Siba (ed.), *Oxford Dictionary of Computer Science* (Novotrade, Budapest 1989), 158.
[2] https://www.coe.int/en/web/artificial-intelligence/glossary (06. 07. 2021)
[3] John McCarthy – Marvin Minsky – Nathaniel Rochester – Claude Shannon, 'Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' [1955] In: Tech. rep., Dartmouth College, 2-3.; http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf (08. 07. 2021)
[4] Csaba Kollár, 'Relation of AI with human security' [2018] Nemzetbiztonsági Szemle (National Security Review) 2018/1., 10
[5] Stuart J. Russell – Peter Norvig, *Artificial Intelligence – A Modern Approach* (Panem, Budapest 2000) Chapter 26
[6] Balázs Csanád Csáji, 'A mesterséges intelligencia filozófiai problémái (Philosophical Problems of AI)' Exam paper, Eötvös Lóránd University Faculty of Philosophy. Budapest (2002), 4.; http://old.sztaki.hu/~csaji/CsBCs_MI.pdf (29. 06 2021)
[7] Ulrike Barthelmess – Ulrich Furbach, 'Do We Need Asimov's Laws?*'* [2014] Lecture Notes in Informatics, Gesellschaft für Informatik, Bonn, 5.
[8] Ray Kurzweil, 'The Singularity Is Near: When Humans Transcend Biology' (Ad Astra 2014)
[9] Stuart J. Russell – Peter Norvig, Chapter 26
[10] Datatilsynet, *Artificial intelligence and privacy* (Report, January 2018); https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf (29. 06. 2021)
[11] Csaba Szepesvári, '*Machine learning – a short introduction'* (Presentation, Hungarian Academy of Sciences Institute for Computer Science and Control, 22 March 2005), http://old.sztaki.hu/~szcsaba/talks/lecture1.pdf (28. 06. 2021)
[12] Datatilsynet, 7.
[13] See e.g. the self-driving car system developed by AImotive with its seat in Budapest, Hungary which is currently under testing, https://aimotive.com/ (01. 07. 2021)
[14] Datatilsynet, 10.
[15] Datatilsynet, 11.
[16] The Royal Society, *Machine learning: the power and promise computers that learn by example* (2017), 93.; https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf (02. 07. 2021)
[17] GDPR Article 5(1)(c): Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ("data minimisation")

[18] Datatilsynet, 12.
[19] E. Cauer – W. Mathis – R. Pauli, 'Life and Work of Wilhelm Cauer (1900 – 1945)' [2000] Proceedings of the Fourteenth International Symposium of Mathematical Theory of Networks and Systems (MTNS2000) Perpignan, 4.
[20] Datatilsynet, 13-14.
[21] *Artificial Intelligence Electronic Almanac* (*Mesterséges intelligencia elektronikus almanach)* Budapest University of Technology and Economics, 2009. TAMOP - 4.1.2-08/2/A/KMR-2009-0026., http://project.mit.bme.hu/mi_almanach/books/aima/ch18s03 (28. 06. 2021.)
[22] Ibid.
[23] https://blogs.microsoft.com/ai/microsoft-researchers-win-imagenet-computer-vision-challenge/ (01. 07. 2021)
[24] CNIL, '*How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence'* (Report, 2017), https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (29. 06. 2021)
[25] CNIL, 30.
[26] Ibid.
[27] Article 29 Data Protection Working Party (WP29), '*Guidelines on automated decision-making and profiling for the application of Regulation 2016/679 (WP251rev.01)*' (Guidelines 2017), 5.
[28] Ibid.
[29] Attila Péterfalvi – Balázs Révész – Péter Buzás (eds), 'Magyarázat a GDPR-ról (Interpretation on the GDPR)' (Wolters Kluwer Hungary 2018), 198.
[30] GDPR Article 4(4)
[31] WP29 (2017), 7.
[32] Ibid. 8.
[33] Ibid. 8.
[34] Michael Veale – Lilian Edwards, 'Clarity, surprises and further questions in the Article 29 Working Party draft guidance on automated decision making and profiling' (2018) Computer, Law and Security Review 34., 400.
[35] WP29 (2017), 22.
[36] Michael Veale et. al., 401.
[37] WP29 (2017), 22.
[38] Ibid. 23.
[39] Ibid. 24.
[40] Ibid. 25.
[41] Article 29 Data Protection Working Party (WP29), '*Guidance on the consent according to Regulation (EU) 2016/679 (WP259rev.01.)'* (Guidelines 10 April 2018), 20-22.
[42] GDPR Article 15(1)(h)
[43] GDPR Article 13(2)(f)
[44] Attila Péterfalvi et. al.,158.
[45] Sandra Wachter – Brent Mittelstadt – Chris Russell 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*'* (October 6, 2017) Harvard Journal of Law & Technology, 31 (2), 2018.. 863-871.
[46] Ibid. 844.
[47] https://www.tensorflow.org/guide/summaries_and_tensorboard (03. 07. 2021)
[48] WP29 (2017), 28.
[49] Datatilsynet, 21-22.
[50] GDPR Article 22(3)
[51] Attila Péterfalvi et. al., 201.
[52] WP29 (2017), 29. and Attila Péterfalvi et. al., 201.
[53] Michael Veale et. al., 403.
[54] Maja Brkan, 'Do algorithms rule the world? Algorithmic decision-making in the framework of the GDPR and beyond' (2019) International Journal of Law and Information Technology, 11. January 2019.. DOI; 10.1093/ijlit/eay017, 13.