Professional paper / Stručni rad

AI AND GDPR: RISKS AND OPPORTUNITIES FOR DIGITAL MARKETING

MARIJA BOŠKOVIĆ BATARELO
Parser compliance d.o.o.
Petračićeva 6, 10 000 Zagreb, Croatia
info@parser.hr

ABSTRACT

This article examines the period of more than three years of the applicability of the General Data Protection Regulation (GDPR) and its relationship with artificial intelligence. The article defines main term and depicts current risks and opportunities when it comes to interaction between AI and GDPR. New marketing solutions, that are based on artificial intelligence, foster new risks that are not only technical, but also ethical, social, and legal. For that reason, GDPR introduced terms, such as, data protection by design and data protection impact assessments. With multidisciplinary approach to compliance, specific risks can be turned into opportunities and technological innovations that foster privacy and data protection in the context of digital marketing.

KEYWORDS*:* GDPR, AI, risks, data protection, privacy, algorithms, privacy by design, differential privacy, compliance, cohorts

1. INTRODUCTION

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 – GDPR) is fully applicable from 25 May 2018, more than three years. GDPR introduced a significant step in regulation of new technologies. Not only because of high administrative fines and supervisory authorities' powers, but because it determined basic principles and steps in ensuring compliance. It established practices of considering data protection by design and by default, as well as risk-based approach, along with organisational and technical measures.

Digital marketing solutions, that are based on artificial intelligence, foster new risks that are not only technical but also ethical, social, and legal. In that way also legal texts, such as GDPR, must consider multidisciplinary aspects of regulation. Some law experts claim that GDPR was dead before it even became applicable, while financial experts claim that GDPR put too much burden on companies.

It is a rather common notion that European Union (EU) somehow regulates, while USA and China support innovation. On the other hand, the EU is not the only one concentrated on regulating data protection. Similar legislations have been already introduced in California, Australia, Japan, Brazil, India, Brazil, South Africa, New Zealand, Chile, Thailand, South

Korea, China, Canada, and UK. This article will examine influence of GDPR on artificial intelligence, correlation of basic legal principles on digital marketing activities that are based on artificial intelligence, certain legal risks with possible opportunities for future developments.

2. DEFINITIONS

The EU High-Level Expert Group on Artificial Intelligence characterises the scope of research in artificial intelligence (AI) as follows: *"As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors, and actuators, as well as the integration of all other techniques into cyber-physical systems)."*

Some IT experts claim that AI does not actually exist yet. What we see today are more complex algorithms, which automate certain processes, predict, and monitor behaviour, decrease costs and time, while reducing number of people who are necessary to perform specific task. However, the concept of an algorithm is more general than the concept of an AI. Not all algorithms involve AI, but every AI includes algorithms.

3. DATA PROTECTION LAWS

Digital economy, as well as digital marketing, are based on vast processing of personal data. Usually, data sets include personal data or mix of personal and unpersonal data. On 14 November 2018, the EU published Regulation on a framework for the free flow of non-personal data in the EU. Recitals of the Regulation describe that the expanding AI represent major sources of non-personal data, for example because of their deployment in automated industrial production processes. Specific examples of non-personal data include aggregated and anonymised datasets used for big data analytics. It is important to emphasise that, if technological developments make it possible to turn anonymised data into personal data, such data are to be treated as personal data, and GDPR is to apply accordingly.

Each company, while processing data, needs to comply with many different fields of law, such as, competition law, consumer protection law, data protection law, labour law, criminal law, antidiscrimination law, etc. Consumer protection law and data protection law share the common goals of correcting imbalances of informational and market power, and, along with competition law, they contribute to ensuring that people are treated fairly.

4. PERSONAL DATA

As a starting point to analysis on correlation of GDPR and AI it is important to define when GDPR applies. As a basic principle, GDPR applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.

Personal data is defined as: *„Information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or*

indirectly, by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. "

GDPR applies to the processing of personal data in the context of the activities of an establishment in the EU, regardless of whether the processing takes place in the EU or not. It also applies to the processing of personal data of data subjects who are in the EU by a controller or processor not established in the EU, where the processing activities are related to the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the EU the monitoring of their behaviour as far as their behaviour takes place within the EU.

Since AI based systems process data by automated means, once this processing is concerned personal data, GDPR shall apply. Definition of personal data is quite broad and, once it is somehow possible to re-identify individuals, the data will be considered as personal. Digital marketing activities are mostly concerned with online identifiers, such as IP address, cookie IDs, which are generally also considered as personal data.

5. DATA PROTECTION IMPACT ASSESSMENT

Article 35 of GDPR requires that a data protection impact assessment is preventively carried out regarding processing that is likely to result in a high risk to the rights and freedoms of natural persons. The assessment is required when the processing involves a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person. Each EU supervisory authority for personal data protection published the list of cases in which data protection impact assessment were obligatory. Such cases, for example, include, smart meters, smart cameras, profiling, etc.

Thus, an impact assessment is usually required when AI-based profiling contributes to automated decision-making affecting individuals and their rights and freedoms. Each controller considers, on case-by-case basis, whether a data protection impact assessment is necessary for some specific case. This kind of assessment is important to examine relevant risks, to mitigate certain risks and to implement certain measures. After the assessment, it is important to evaluate and monitor the risks and cooperate with a relevant supervisory authority.

6. ROLES IN DATA PROCESSING

The part of AI technologies that puts most of the legal risks to developers, users and implementors is determining roles in personal data processing. In theory, it is quite simple. Data controller is the one who determines purpose and means of personal data processing. Data processor is the one who performs processing on behalf of data controller. Typical data processor is often a company that provides data storage services. The situation becomes complicated when few companies jointly determine purposes and means and then they must agree on responsibilities. This is often the case with AI based technologies and each joint controllers' situation should be examined on case-by-case basis.

Data controllers (companies) usually engage various IT and marketing providers and by this they outsource processing of large amounts of personal data. Nevertheless, they remain accountable for any processing activities carried out on their behalf. They must assess the risks and have appropriate contractual and technical safeguards in place to mitigate those risks. Similar situation can happen with all sorts of AI based technologies because it is in their core systems to process data, analyse it, and use it for more automated and smarter processes. This possesses high risks for companies that need to be very diligent and keep confidential information in accordance with special laws, for example, government institutions, banks, attorneys at law, doctors, financial advisors etc.

7. PROFILING

Through AI technologies people can be determined by a certain profile, according to their health, gender, age, interests etc. The system learns about personal aspects of users, analyses the data, and predicts future behaviour. This is how, while we write an e-mail, the system offers to us a suggestion of words and sentences, and while we surf the Internet, we get certain types of advertisements.

GDPR defines profiling as the following: *'profiling'[...] consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location, or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.*

Profiling is not forbidden under GDPR. It can be performed based on defined legal grounds (Article 6 – legal obligation, consent, legitimate or public interest, vital interests of data subjects, contractual relations). Article 9 GDPR defines that special category of personal data (data on health, sex life or sexual orientation) are, in general, forbidden to process and prescribes exemptions. It is up to data controller to assess on case-by-case basis, which legal ground is applicable.

8. TRANSPARENCY

Transparent and clear information are closely connected with consent. Data subject cannot give informative consent if he/she does not understand the processing. Consent, according to Article 4 GDPR, should be freely given, specific, informed, and unambiguous, and be expressed through a clear affirmative action: *'consent' of the data subject means any freely given, specific, informed, and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.*

While assessing the legal ground, data controllers often do not choose consent as a legal ground. The reason for those lays down in the fact that consent is the most difficult legal ground to manage. Data subject has the right to withdraw consent and once the consent is withdrawn data controller must stop processing data and delete the data accordingly. In AI based technologies, this is technically highly demanding process, and that is why we usually see legitimate interest and contractual obligation as a common legal ground for analysing data, subsequent processing

of personal data for automation and analytics, and for learning from the data and improving systems.

9. ETHICAL FRAMEWORK

GDPR includes not only compliance with the relevant laws, but also ethical context. Laws usually do not define in detail what is allowed. It is up to data controllers to assess, analyse, define, and be accountable. Principle of accountability, defined in GDPR, puts burden of proof on data controllers since they must prove compliance. Compliance is usually proven through documents, logs, trainings, assessments, and analysis.

Ethics comes to the focus also in cases when data controllers use legitimate interest as legal ground for processing personal data. It is not enough only to state that data can be used for improving services or marketing activities. This analysis should strike the right balance between interests of data controller and right and freedoms of data subject. It should also consider social responsibility.

According to the EU High-Level Expert Group on AI, to implement and achieve trustworthy AI, seven requirements should be met, building on the principles mentioned above:

1) Human agency and oversight, including fundamental rights.
2) Technical robustness and safety, including resilience to attack and security, fall back plan and general safety, accuracy, reliability, and reproducibility.
3) Privacy and data governance, including respect for privacy, quality and integrity of data, and access to data.
4) Transparency, including traceability, explainability and communication.
5) Diversity, non-discrimination, and fairness, including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation.
6) Societal and environmental wellbeing, including sustainability and environmental friendliness, social impact, society, and democracy.
7) Accountability, including auditability, minimisation and reporting of negative impact, trade-offs, and redress.

These terms are quite broad and open to interpretation. For technical part of development teams, it is not easy to incorporate law and ethics into the code. The code is quite specific while, legal and ethical terms are subject of consensus and interpretation.

10. DATA PROTECTION BY DESIGN

The protection of the rights and freedoms of natural persons regarding the processing of personal data requires that appropriate technical and organisational measures have been taken to ensure that the requirements of GDPR are met. To be able to demonstrate compliance, data controllers adopt internal policies and implement measures which meet the principles of data protection by design and data protection by default. These measures, with due regard to the state of the art, could include minimising the processing of personal data, pseudonymising and anonymisation techniques, trainings of staff, and involvement of data protection officer through all the stages of a project.

If we consider broader framework, it is important to have in mind also privacy by design, security by design, and ethics by design. This would probably mean that, in the future, an AI projects will involve more multidisciplinary teams. For example, if a company is using AI based technologies on a web shop, it must ensure that it does not discriminate specific group of people, that biases are considered, and that the procedure is transparent to the buyers.

11. OPPORTUNITIES

It seems that possible opportunities lay down in technical solutions and training IT and marketing staff about legal matters. New projects consider synergy of different competences in one person, or synergy of people of different backgrounds involved in one project.

In 2019, Google published that it invested in research which advanced innovation in the field of privacy. Using this cryptographic protocol, two parties can encrypt their identifiers and associated data, and then join them. They can then do certain types of calculations on the overlapping set of data to draw useful information from both datasets. All inputs (identifiers and their associated data) remain fully encrypted and unreadable throughout the process. Neither party ever reveals their raw data, but they can still answer the questions at hand using the output of the computation. This result is the only thing that is decrypted and shared in the form of aggregated statistics. For example, this could be a count, sum, or average of the data in both sets. The crucial question whether it is possible to preserve privacy and still use and analyse the data became even more relevant during coronavirus pandemic. Google and Apple provided this type of technology for stop-covid mobile apps that functions on previously described principles. It means that a person can track contacts, provide information on disease, and still not reveal the identity.

During 2020, Google also started testing, through Google Chrome, results of the project Privacy Sandbox, named Federated Learning of Cohorts (FLoC). FLoC proposes a new way for businesses to reach people with relevant content and ads by clustering large groups of people with similar interests. Data about behaviour and performance of groups of users related by common attributes will be put in a cohort. A cohort is a group of users who share a common characteristic that is identified in the report on analytics dimension. For example, all users with the same Acquisition Date belong to the same cohort. The Cohort Analysis report enables analysis of cohort behaviour. This system plans to replace third-party cookies and provide privacy-preserving alternative for digital marketing.
Most of the critics in AI based solutions for more privacy-friendly digital marketing goes to possible discrimination by putting people into separated groups (cohorts). Also, European Commission started investigation on Google FLoC solutions due to the possible breach of EU competition law rules. The technology has not been perfect and there has been room for the improvement. However, at this point, this represents significant opportunity for balancing AI and GDPR in the context of digital marketing.

12. CONCLUSION

AI, as well as digital marketing activities, are not possible without data. Many of the used data are personal data or mixed sets of data that include personal data. Definition of personal data is quite broad and GDPR puts a lot of compliance burden on companies. It is important, from compliance perspective, to perform data protection impact assessments and include experts

from different backgrounds. In some cases, it is even advisable to coordinate the project with supervisory authorities. It is crucial to follow the EU laws and relevant national laws, to have in mind recognised risks of technology and have a broad perspective that includes also ethical principles and guidelines. Significant opportunities for vast usage of data lay down in techniques, such as, differential privacy and multi-party computation. If we do not consider compliance and ethics as a burden, future out-of-the-box solutions will play crucial role in advanced digital marketing activities.

REFERENCES

1. B.J. Koops, Trouble with Data Protection Law, *International Data Privacy Law*, doi: 10.1093/idpl/ipu023, Forthcoming. Tilburg Law School Paper No. 04/2015 (**Article reference**)
2. G. Sartor, F. Lagiola, The Impact of the GDPR on artificial intelligence, *Scientific Foresight Unit*, PE 641.350, ISBN: 978-92-846-6771-0, doi: 10.2861/293, QA-QA-02-399-EN-N. (**Article reference**)
3. Google, https://blog.google/products/ads-commerce/2021-01-privacy-sandbox/ [January, 25th 2021] (**Internet reference**)
4. S. Ciriani, The Economic Impact of the European Reforms of Data Protection, *Communications & Strategies*, no. 97, 1st quarter 2015, pp. 41-58 (**Article reference**)
5. 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) (**Legal source reference**)
6. Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (**Legal source reference**)
7. The Convention for the Protection of Human Rights and Fundamental Freedoms, the European Convention on Human Rights, opened for signature in Rome on 4 November 1950 and came into force in 1953 (**Legal source reference**)