Autonomous targeting algorithms and public information under the General Data Protection Regulation

Noé G.

Introduction

Place of birth, education, martial status, family members, etc... All these elements can be displayed in the "intro" section of a Facebook profile, right below the profile picture and the name. This set of personal information is accessible to anyone on the social network, not only to Facebook "friends", and is part of Facebook's "public information", or "*information that can be seen by anyone"*[1]. The notion of "public information", however, is not a Facebook invention but a continuation of the Swedish *"Tryckfrihetsförordningen"* of 1766, the first freedom of information law in the world[2]. This dynamic was massively reinstated by western democracies 200 years later[3] and is today the keystone of many democratic systems[4]. Until the beginning of the 21st century, the publicity of information principle seemed to be unquestionable. However, this general consensus is today called into question with the soaring technical capacities over the processing of data and the development of autonomous algorithms.

With the emergence of data science technologies such as classification[5], regression[6] or recommendation[7], it is today possible to accurately track, identify and target an individual. Furthermore, this can be done without any human intervention by autonomous algorithms, such as Amazon's recommendation algorithm that collects data from its website to spread offers to customers without any human interference[8]. These recommendations are made on the basis of data collected on Amazon's website with the client's agreement and for a limited purpose[9], but in the case of the processing of public data, these consent or purpose filters are non-existent. Any algorithm can download data from any website containing data accessible to anyone and process it for any purpose without supervision or awareness from the data subject.

---

[1] Facebook.com, « What is Public Information on Facebook »

[2] Ackermann & Sandoval-Ballestros, , *Essentials of the right to access public information,* 2006, p. 88.

[3] See the U.S. Freedom of Information ACT of 1966 or the Japanese Shiru Kenri (« right to know ») Supreme Court principle of 1969

[4] H.-J. Blanke and R. Perlingeiro, *Essentials of the Right of Access to Public Information*, chap. 1

[5] S. Marsland, *MACHINE LEARNING - An Algorithmic Perspective*, chap. 7, 2015

[6] S. Marsland, op.cit., chap. 4

[7] K. Pearson, *On the Theory of Contingency and its Relation to Association and Normal Correlation*, 1904

[8] Smith, Linden, *Two decades of recommender system at Amazon.com*, IEEE internet computing journal, 2017

[9] Amazon.com, privacy notice - last updated dec. 2020

Such a situation raises numerous ethical, legal and security issues. Since these algorithms are autonomous, "black-box effects"[10] can appear and produce uncontrollable effects. Despite the negative impact these technologies can have, no specific rules seem to exist for the autonomous processing of public information. If the processing is about personal information, the European General Data Protection Regulation (GDPR)[11] appears to be one of the only possible effective regulatory tool in Europe. However, the question of its scope of application is raised. To what extent can the processing of personal public information by autonomous algorithms be regulated by the GDPR?

Eventhough the GDPR appears to be a suitable regulation to address this issue by granting more security to public data and pinpointing autonomous targeting as a possible threat to fundamental rights (I), its effectiveness limited in practice (II).

1 The GDPR, an efficient regulation for the automated processing of personal public data in appearance

1.1 An enhanced safety of personal public information

The right to access information is today considered as a fundamental right by the constitutions of an non-negligible number of countries[12]. This global tendency is today pursued further by western governments with the willingness to make public the massive amount of data produced by public entities[13]. The notion of public information is then brought into the digital era and nothing seems to stand in the way of this fundamental right's expansion. However, since the 1970's[14], data protection laws emerged in parallel of the publicity of data restraining their so far uncontrolled processing.

The GDPR, the current data protection regulation standard in Europe, sets a frame to the the processing of "*any information relating to an identified or identifiable natural person*", or

[10] L. Edwards & M. Veale, *Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for*, 2017

[11] Regulation (EU) 2016/679 of the European Parliament and of the Council, 27/04/2016

[12] H.-J. Blanke and R. Perlingeiro, op. cit.

[13] French Law "l'Activité, la Croissance et l'Égalité des chances économiques", article 4, 2015

[14] For example Sweden's Data Act (1973), USA's Privacy Act (1974), France's Law on Data Processing (1977)

"personal data"[15]. Henceforth, the processing of such data will have to meet strict criteria like lawfulness, fairness or necessity[16]. Considering the broadness of the definition of personal data in article 4, it may be inferred that personal data in the public domain is also subject to these regulations. This restriction might appear at first sight in contradiction with information freedom as it sets a solid frame to the exercise of this right.

Though, through its processing and organizational requirements, the GDPR tend to ensure a safe environment for public personal data. Any misuse of this data which could directly affect individuals that was not really supervised before appears to have now solid safeguards. The GDPR adds then a precious layer of protection over personal public data so they cannot be handled anyhow by anyone because of their publicity. Furthermore, the question of automated targeting is also handled in the GDPR which adds another layer of protection.

1.2 The automated targeting of data subjects, a practice supervised by the GDPR

Automation and targeting are two difficult subjects to apprehend both technically and legally. First, automation has different degrees. An algorithm might function on the basis of a entire autonomous implementation[17] or, in contrary, only use a limited number of self-governing functions. Thus, it appears to be difficult to assess the extent of an algorithm's autonomy without its source code. Likewise, the targeting of people is sometimes very difficult to assess. When one is receiving an information, there is no way to evaluate its degree of targeting or the algorithmic processing behind it. For example, an advertisement on the social network Instagram can be the result of a meticulous study of the target's profile or just the consequence of stochastic algorithmic choices. The inexactness of the process used behind algorithmic autonomy and targeting makes these technologies difficult to regulate.

However, the GDPR seems to address these problems in a relevant manner. GDPR article 2 states clearly in paragraph one that this regulation "*applies to the processing of personal data wholly or partly by automated means*". Such a sentence demonstrates that the degree of autonomy of an algorithm is not taken into account here. Only a full autonomous treatment of personal data would be subject to more restrictive regulation[18]. Apart from this, just the result of the processing is regarded. To make an analogy with genetics, it is not the genotype, the assembling of genes, that is observed but the phenotype, or the resulting phenomenon of the blending.

The phenotype approach is also chosen in the case of targeting. The targeting of customers

[15] GDPR art4

[16] GDPR art4

[17] For example the A* algorithm is designed to autonomously find a path, see K. Berman and L. Paul, *Algorithms – Sequential, Parallel and Distributed*, chap.23, 2005

[18] GDPR art22

consisting of sets of algorithmic operations performed on personal data, it can be considered as a form of "processing" in accordance with article 4.2. And as it is demonstrated in article 3, the processing of personal data is defined as the main recipient of this regulation. Thus whether the targeting is made by an autonomous algorithm or not, this regulation will apply as long as a data subject has been targeted.

The autonomous targeting seems then to be properly handled by the GDPR. In practice, however, this must be nuanced.

2 The GDPR, a deficient regulation in practice

2.1 The problem of accountability

At first sight, the GDPR appears also to be pretty robust on accountability questions. Two main actors can be found responsible under the GDPR : the controller and the processor[19]. At first sight, such a scission seems efficient as every possible actor involved in the handling of personal data seems to be held liable for any unlawful processing. But firstly, it has been demonstrated that this safety net was porous as entities not bound by the GDPR can be involved in the handling of data[20]. The problem is taken further when no apparent organization is found behind a processing of personal data. If only the result of the targeting is discovered without possibilities of identifying the responsible, as it often the case on the messaging app Whatsapp[21], the GDPR seems useless as no one can be held liable.

Through the notion of controller, the GDPR seems to look for the beneficiaries of the processing to find a responsible. An entity reaping the benefits of a targeting it ordered, like a fashion company advertising on Instagram, could be considered as a controller because it determined the purpose and means of the processing[22]. It could therefore be held liable for any GDPR breach[23], but this system is unfortunately not perfect. Hiding, on the internet or through shell companies, is easily doable. Furthermore, the processing can have no link with the beneficiary. With the help of public data, any citizen with the required knowledge can process alone some data in favour of a cause or support a political party. This problem is indubitaly exacerbated

[19] See art 5 or 82

[20] *The security of personal data under the GDPR:a harmonized duty or a shared responsibility?,* P.T.J.Wolters, International Data Privacy Law,2017,Vol. 7, No. 3

[21] JUS5690 Group 5, *Regulating Political Microtargeting* (movie)

[22] GDPR art4.7

[23] art 82

with high availability of public data. Through this “lone activist” model, the beneficiaries and the processing have no link and no responsible seems to subsist.

Ultimately, the question of accountability is not fully settled by the GDPR and the publicity of data make such a task harder. The GDPR appears therefore as an incomplete solution to regulate the automated processing of personal public data, but a final major flaw remains.

2.2 “Pas vu, pas pris”

The French expression “pas vu, pas pris”, which could be translated into “unseen, not caught”, reveals in itself one of GDPR’s major loophole. As it is technically impossible and in contradiction with fundamental rights of many countries to monitor the use of every computer on the planet, nothing prevents a clandestine processing of personal public data from a basement, a room, or anywhere in the world. Once the necessary data collected, any person with the required savoir-faire can process it and target European data subjects without anyone aware. There is even no need to buy or hack databases to get information thanks to the development of social media as developed in the introduction. It is then very simple to operate discrete targetings and the chances of being caught or even noticed are very slim.

One could argue that a large-scale targeting of people is so complex to set up that it will without doubts leave traces that makes hiding complicated. First, the amount of energy required to power puissant computers or datacenters is not negligible and such a consumption could be noticed very quickly. Moreover, the knowledge needed to manage and process such a quantity of data is rare and wanted, which make the realization of such a process rather hard.

Nevertheless, if this processing is done outside of Europe or in a hostile country, the chances of finding the perpetrators are pretty slim. Also, new technologies like distributed algorithms also participate to the difficulty of observing clandestine processing. These could be done on a large number of regular computers instead of a big one and without the consent of their owners. The GDPR has consequently a high progress margin to handle the autonomous targeting of public personal data.

Table of reference

Facebook.com, « What is Public Information on Facebook »

Ackermann & Sandoval-Ballestros, *Essentials of the right to access public information,* 2006, p. 88.

See the U.S. Freedom of Information ACT of 1966 or the Japanese Shiru Kenri (« right to know ») Supreme Court principle of 1969

H.-J. Blanke and R. Perlingeiro, *Essentials of the Right of Access to Public Information*, chap. 1

S. Marsland, *MACHINE LEARNING - An Algorithmic Perspective*, chap. 7, 2015

K. Pearson, *On the Theory of Contingency and its Relation to Association and Normal Correlation*, 1904

Smith, Linden, *Two decades of recommender system at Amazon.com*, IEEE internet computing journal, 2017

Amazon.com, privacy notice - last updated dec. 2020

L. Edwards & M. Veale, *Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for*, 2017

Regulation (EU) 2016/679 of the European Parliament and of the Council, 27/04/2016
H.-J. Blanke and R. Perlingeiro, op. cit.

Loi "l'Activité, la Croissance et l'Égalité des chances économiques", France, article 4, 2015

For example Sweden's Data Act of 1973, USA's Privacy Act of 1974, France's Law on Data Processing, Files and Individual Liberties of 1977

K. Berman and L. Paul, *Algorithms – Sequential, Parallel and Distributed*, chap.23, 2005

The security of personal data under the GDPR:a harmonized duty or a shared responsibility?, P.T.J.Wolters, International Data Privacy Law,2017,Vol. 7, No. 3