University of Liège

DROIT1357-1 - European law, (big) data and artificial intelligence applications seminar

AI as profiling machines

Ljupcho Grozdanovski
Cyril Fischer

GROUP 5.3
AMAR Salah
BASTIN Clara
SCHMIDT Helena
SIBOYABASORE Cédric

2021-2022

Table of content

Introduction **2**

The profiling and the GDPR **3**
The profiling in a general way 3
Automated decision-making: 4
General provisions on profiling and automated decision-making 5

How AI profiling systems work **7**
Data collection 7
Machine learning 10
Case studies 13

Distinction between AI Act and GDPR : What is the relevance of the future regulation on artificial intelligence with regard to the GDPR ? **17**
Objectives 18
Prohibition of specific AI systems 18
High-risk AI systems 18
Measures in support of innovation 20
Specific requirements for providers of high-risk AI systems 20
Specific requirements for users of high-risk AI systems 21
Specific requirements for certain AI systems 21
European Artificial Intelligence Board 22

Are the GDPR and AI Act technically feasible? **23**
Transparency and explainability of profiling algorithms 23
Correlation and causation 26
Profiling system’s limitation 29

Conclusion **31**

Introduction

In 2014, Amazon started developing an automated profiling system that aimed at performing resume review. But in 2015 they noticed that the algorithm discriminated against women.

The Amazon case raises some questions about profiling, including: How does such an algorithm become sexist? What are the risks? What are the regulations in that regard? Are they sufficient? How can it be fixed?

To answer such questions we will study diverse aspects of profiling systems in general.

First of all, the notion of profiling and automated decision-making will be introduced in a general way. We will also give an overview of the different principles contained in the 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.

After that, a more technical overview of profiling will be given. To do this, we will present how profiling systems technically work. Following this, several case studies of actual usage of profiling will be described. Transparency and the black-box issue will be discussed, as well as an overview of the limitations of current profiling systems.

Finally, to answer the question of whether the AI Act and the GDPR are complimentary or redundant, a distinction will be made between the two to highlight the points of comparison and develop them on the basis of the articles of the proposed regulation.

The profiling and the GDPR

The profiling in a general way

The 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[1] provides in recital n°71 that : "*profiling that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyze or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her*".

Article 4 of the GDPR (General Data Protection Regulation), entitled "Definition", also gives us a definition of profiling in paragraph 4: "*profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements;*"

Profiling is therefore composed of three elements:

A. It must be an automated processing operation
B. It must be carried out on personal data
C. The purpose is to evaluate the personal aspects of a natural person[2].

Profiling must apply some form of automated processing without necessarily excluding human involvement[3]. This definition draws heavily on the definition of profiling: "*Profiling is an automated data processing technique which consists of applying a 'profile' to a natural person, in particular for the purpose of making decisions about that person or analyzing or predicting his or her personal preferences, behavior and attitudes."*[4] The GDPR says that profiling is automated processing of personal data for evaluating personal aspects, in particular to analyze or make predictions about individuals.

The use of the word "evaluate" suggests that profiling involves some form of assessment or Judgment about a person. A simple classification of individuals based on known characteristics such as their age, sex, and height does not necessarily lead to profiling. This will depend on the purpose of the classification. For instance, a business may wish to

[1] We will refer to it under its better-known name "GDPR".
[2] Guidelines on automated individual decision-making and profiling for the purpose of Regulation 2016/679 (adopted on 3 October 2017).
[3] *Ibid*.
[4] Council of Europe: The protection of individuals with regard to automatic processing of personal data in the context of profiling. Recommendation CM/Rec(2010)13 and explanatory memorandum. Council of Europe 23 November 2010.

classify its customers according to their age or gender for statistical purposes and to acquire an aggregated overview of its clients without making any predictions or drawing any conclusion about an individual. In this case, the purpose is not assessing individual characteristics and is therefore not profiling[5].

The GDPR is inspired by the definition of profiling in the Council of Europe, the Recommendation CM/Rec (2010). This Recommendation usefully explains that profiling may involve three distinct stages:

A. Data collection;
B. Automated analysis to identify correlations;
C. Applying the correlation to an individual to identify characteristics of present or future behavior[6].

Broadly speaking, profiling means gathering information about an individual (or group of individuals) and evaluating their characteristics or behavior patterns in order to place them into a certain category or group.

Automated decision-making:

Before we move on to how the GDPR addresses the concepts, a few words about automated decision making are in order. Automated decision-making has a different scope and may partially overlap with or result from profiling.[7]

There are two types of automated decision-making:

a) Solely automated decision-making is the ability to make decisions by technological means without human involvement.

This type of automated decision can be made with or without profiling.

For example: Imposing speeding fines purely on the basis of evidence from speed cameras is an automated decision-making process that does not necessarily involve profiling. It would, however, become a decision based on profiling if the driving habits of the individual were monitored over time, and, for example, the amount of fine imposed is the outcome of an assessment involving other factors, such as whether the speeding is a repeat offense or whether the driver has had other recent traffic violations.[8]

b) Decisions that are not solely automated might also include profiling.

For example, before granting a mortgage, a bank may consider the credit score of the borrower, with additional meaningful intervention carried out by humans before any decision is applied to an individual.[9]

---

[5] Guidelines on automated individual decision-making and Profiling for the purpose of Regulation 2016/679 (adopted on 3 October 2017).
[6] Guidelines on automated individual decision-making and Profiling for the purpose of Regulation 2016/679 (adopted on 3 October 2017).
[7] *Ibid*.
[8] *Ibid.*
[9] *Ibid*.

Three arguments explain the growing use of artificial intelligence by companies and administrations[10] :

a) Optimizations: These systems monitor the correct implementation of regulations in order to detect potential fraudsters[11] or future offenders;[12]
b) Security: for example, in the fight against the danger of terrorism;
c) Objectification : the operation of a neutral machine working on data that do not lie is a very good argument against subjectivity[13].

The use of these artificial intelligence systems by the actors around us generates risks for our individual liberties. In particular, there are concerns about the use of facial recognition, and France believes that such a practice can be used if justified by the public interest[14]. Recently, the Belgian insurance law has been amended to the effect that an applicant for insurance who refuses to acquire or use an object that collects personal data cannot be refused insurance[15].

General provisions on profiling and automated decision-making

a) Data protection principles

These principles are relevant for all profiling and automated decision-making involving persan data.[16]

1. Principles of Lawful, fair and transparent

This principle will be developed below in the section: 'Transparency and explainability of profiling algorithms'. We refer you to it.

---

[10] Y.POULET., « Five years on: the GDPR and the challenges of profiling in the age of artificial intelligence", R.A.E, L.EA., 2021/1, pp. 87-101.
[11] Arrêté Royal du 30 juillet 2018 loi du 18 septembre 2017 relative à la prévention du blanchiment de capitaux et du financement du terrorisme et à la limitation de l'utilisation des espèces, transposant la Directive 2015/849 du Parlement européen et du Conseil du 20 mai 2015 relative à la prévention de l'utilisation du système financier aux fins du blanchiment de capitaux ou du financement du terrorisme, modifiant le règlement (UE) n° 648/2012 du Parlement européen et du Conseil et abrogeant la directive 2005/60/CE du Parlement européen et du Conseil et la directive 2006/70/CE de la Commission telle que modifiée par la Directive (UE) 2018/843 du Parlement européen et du Conseil du 30 mai 2018 modifiant la directive (UE) 2015/849 relative à la prévention de l'utilisation du système financier aux fins du blanchiment de capitaux ou du financement du terrorisme, ainsi que les Directives 2009/138/CE et 2013/36/UE.
[12] Y. POULET., *op cit*.
[13] *Ibid*.
[14] *Ibid*.
[15] Law of 10 December 2020 amending the Law of 4 April 2014 on insurance with a view to establishing in the field of health insurance and individual life insurance a restriction on the processing of personal data concerning lifestyle or health from connected objects, *M.B*., 25 January 2021.
[16] Recital 72 of the GDPR "Profiling is subject to the rules of this Regulation governing the processing of personal data, such as the legal grounds for processing or data protection principles".

2. Further processing and purpose limitation

This principle is found in article 5 paragraph 1 point B. Profiling can involve the use of personal data that was originally collected for something else. Whether this additional processing is compatible with the original purposes for which the data were collected will depend upon a range of factors. These factors are reflected in the article 6(4) :

a) The relation between the purposes for which the data have been collected and the purposes of the further processing;
b) the context in which the data were collected and the reasonable exceptions of the data subjects as to their further use;
c) the nature of the data ;
d) the impact of the further processing on the data subjects ;
e) the safeguards applied by controller to ensure fair processing and to prevent any undue impact on the data subjects.[17]

3. Data minimisation

This principle is found in article 5(1) point C. Controllers must make sure they are complying with the data minimisation principle, as well as the requirements of the purpose limitation and storage limitation principles. Controllers should be able to clearly explain and justify the need to collect and hold personal data, or consider using aggregated, anonymised or (when this provides sufficient protection) pseudonymised data for profiling[18].

4. Accuracy

This principle is found in article 5(1) point D. Controllers should consider accuracy at all stages of the profiling process, specifically when:

- collecting data
- analysing data
- building a profile for an individual
- applying a profile to make a decision affecting the individual.

If the data used in an automated decision-making or profiling process is inaccurate, any resultant decision or profile will be flawed.Controllers need to introduce robust measures to verify and ensure on an ongoing basis that data reused or obtained indirectly is accurate and up to date. This reinforces the importance of providing clear information about the personal data being processed, so that the data subject can correct any inaccuracies and improve the quality of the data[19].

5. Storage limitation

This principle is found in article 5(1) point E. Machine-learning algorithms are designed to process large volumes of information and build correlations that allow organizations to build up very comprehensive, intimate profiles of individuals[20].

---

[17] Guidelines on automated individual decision-making and Profiling for the purpose of Regulation 2016/679 (adopted on 3 October 2017).
[18] *Ibid.*
[19] *Ibid*.
[20] *Ibid*.

How AI profiling systems work

Profiling systems highly depend on the quantity of data. These data can be historical data or data collected by the same organization performing the profiling, and even data by other organizations. Before explaining how these systems work, it is important to first explain how data is collected.

Data collection

When it comes to profiling, we can typically identify three types of data[21]:

The first type of data is called **observed data**. This is data resulting from the data subject's interactions with an organization, a website. Examples of such type of data include the number of times a user has visited a specific web page and information about the device they used to access it. This data collection process is also called *passive* user data collection because the user does not necessarily realize that their data is being collected due to the process not requiring any action from their side[22].
Websites are associated with servers. The main tools to capture observed data are HTTP connections between the client and the website server which keeps track of the users' connections in log files. When browsing and clicking links on a website, the user is actually sending a HTTP request to the server associated with the website which, in turns, sends a HTTP response to the user. HTTP is a stateless Web protocol, meaning that the server has no way of remembering a particular client and their previous requests. This means that the log files the server keeps track of are not mapped with an individual user. However, combined with other web technologies such as cookies, this can be made possible.

Cookies are text files which help the server remember previously provided information and user preferences to avoid remembering them at each visit of the website. They are stored in the web browser that was used to visit the website.[23]

[21] https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-for-the-use-of-personal-data-in-political-campaigning-1/political-campaigning-in-the-online-world/

[22] http://web.mit.edu/ecom/www/Project98/G2/data.htm

[23] Benoit Donnet, *Introduction to computer security*, University of Liège (2020-2021)

Cookies can be divided into two categories: the first-party cookies, which are directly set up and stored by the web page the data subject is visiting and the third-party cookies which are cookies shared between multiple domains and allow building a profile across different websites.[24]

When cookies can identify users, they are considered personal data and are therefore subject to the GDPR[25]. In line with Article 4(11) of the GDPR, users visiting a website have to give explicit content to the processing of their personal data before the website can maintain cookies about them. This content cannot be implied but should rather be given by an affirmative act[26], and therefore users have to accept cookies or not by ticking a box when visiting a web page for the first time. They can choose to activate some cookies while refusing others.

The GDPR has a large scope, as it is concerned with the protection of any kind of personal data. As far as cookies are concerned, there is a proposal for a regulation of the European parliament and of the council called “E-privacy”. This proposal is complementary to the GDPR and would be particularized in the protection of electronic communications personal data. This proposal was justified by the Parliament "*The ePrivacy Directive ensures the protection of fundamental rights and freedoms, in particular the respect of privacy, confidentiality of communications and protection of personal data in the electronic communications sector. It also ensures the free movement of electronic communications data, equipment and services within the Union*[27]."

Recital 23 of the proposition provides that “*The principles of data protection by design and by default were codified under Article 25 of Regulation (EU) 2016/679. Currently, the default settings for cookies are set in most current browsers to ‘accept all cookies’. Therefore, providers of software enabling the retrieval and presentation of information on the internet should have an obligation to configure the software so that it offers the option to prevent third parties from storing information on the terminal equipment; this is often presented as ‘reject third party cookies’. End-users should be offered a set of privacy setting options, ranging from higher (for example, ‘never accept cookies’) to lower (for example, ‘always accept cookies’) and intermediate (for example, ‘reject third party cookies’ or ‘only accept first party cookies’). Such privacy settings should be presented in a an easily visible and intelligible manner.*”[28].

---

[24] https://www.cookiepro.com/knowledge/whats-the-difference-between-first-and-third-party-cookies/

[25] https://gdpr.eu/cookies/

[26] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A32016R0679&from=EN&fbclid=IwAR1WYCQsx67AzwHqGrL-zk4YP7iAZYT22qJvhYVMCaKw00pZuDXDx6XgRJI

[27] Proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), 2017

[28] *Ibid*.

Recital 24 that: "*For web browsers to be able to obtain end-users' consent as defined under Regulation (EU) 2016/679, for example, to the storage of third party tracking cookies, they should, among others, require a clear affirmative action from the end-user of terminal equipment to signify his or her freely given, specific informed, and unambiguous agreement to the storage and access of such cookies in and from the terminal equipment. Such action may be considered to be affirmative, for example, if end-users are required to actively select 'accept third party cookies' to confirm their agreement and are given the necessary information to make the choice.*"[29]

The second type of data is called **provided data**. This represents personal information deliberately provided by the data subject. The data collection process is called *active* data collection because it requires the subject to perform action in order to provide the data.
These data can be shared through direct ways such as form fields, checkboxes or quizzes to the website that will store them, possibly in combination with other user data from cookies. Examples of such type of data include name, location, email address which could be provided when registering on a platform.

The third type of data is called **inferred data**. This term represents data derived from the observed and provided data by the organization. For example, a social media platform may want to infer its users interests based on their observed data and on the data these users provided to the platform.
According to the GDPR[30], inferred data are considered personal data if they are attached to personal observed and/or provided data that allow the identification of an individual. If they do not allow the identification of individuals, then they are not considered as personal data. It is important to mention that we have cited as a source the UK's Information Commissioner's office (ICO), which provides guidance to help compliance with the UK GDPR but the UK GDPR and the EU GDPR have minimal differences which do not impact the points we mention.
To obtain inferred data, more processing of the provided and observed data, which usually may come in an unstructured way, has to be done. In general, this inference is quite arduous and tedious for human beings as it might involve a large number of data subjects. As an example, Facebook reported an average of 1.93 billion daily active users during the third quadrimester of 2021[31]. In that case, deriving meaningful personal data for each user is next to impossible.

The task of inferring data is generally automated through the use of machine learning algorithms.

---

[29] *Ibid*.
[30] https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-for-the-use-of-personal-data-in-political-campaigning-1/personal-data/

[31] Facebook Investor Relations, *Facebook Reports Third Quarter 2021 Results* (2021)

Machine learning

With these users' information accumulated in databases, organizations can apply analysis on a large scale using machine learning algorithms. Currently the state of the art for the profiling system is based on machine learning. Several other methods were used, such as statistical techniques. But with the current size and complexity of data, machine learning techniques are much more appropriate. Note that in the following sections, we will use the words system and algorithm interchangeably.

Machine learning is "*concerned with the design, the analysis, and the application of algorithms to extract a model of a system from the sole observation (or the simulation) of this system in some situations (i.e., by collecting data)*"[32]. It is useful when the solution needs to be adapted to particular cases, which is the case for profiling for which we wish to analyze or predict a natural person's aspects. Machine Learning is a subset of Artificial intelligence (AI), Artificial intelligence being a broader field. AI, according to the HLEG, "*refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to predefined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behavior by analyzing how the environment is affected by their previous actions"* [33]. Therefore, it includes not only machine learning, but also other techniques such as machine reasoning (planning, scheduling,..) and robotics (control, perception,..) which are irrelevant to profiling.

The extraction of useful information from data is called data mining. There are two types of data mining: descriptive data mining and predictive data mining[34].

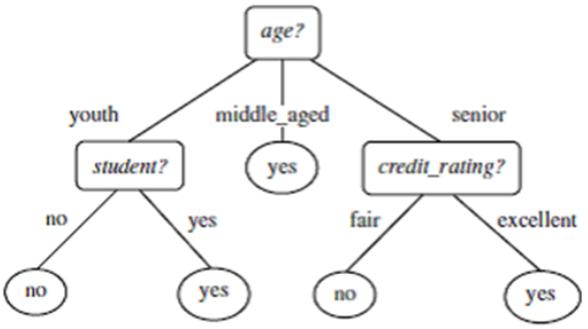
Descriptive data mining aims at finding regularities, interesting patterns and relationships between variables or samples in a database and at gaining more insight into the data. For this purpose, descriptive data mining uses machine learning techniques which we call unsupervised, because they do not need pre-defined labels on the data.
Examples of data mining techniques include clustering, which divides the dataset samples into clusters such that samples in the same cluster are more similar to each other than objects from other clusters. A purpose of clustering could be to identify groups of people within a population based on their financial situation, geographic situation[35].

Predictive data mining, on the other hand, aims at making predictions based on a pattern a supervised machine learning algorithm found in the database. The data samples are each annotated with a label (or output). If the label is a class among a finite number of possibilities, the supervised learning task is called classification. For example, a database of

[32] P. Geurts, L. Wehenkel, *Introduction to machine learning* (2020-2021)
[33] HLEG, *A Definition of AI: Main Capabilities and Scientific Disciplines* (18 Dec. 2018)
[34] B. Schemer, *The limits of privacy in automated profiling and data mining* (2011)
[35] P. Geurts, L. Wehenkel, *Introduction to machine learning* (2020-2021)

previously incarcerated individuals may contain characteristics such as their age, gender, marital status, ethnicity, number of committed crimes and date of last committed crime and the label of each sample would be a “Yes” or “No” respectively indicating whether the individual has committed recidivism or not. An AI profiling model would then be used to predict whether a new individual who had not been observed in the database and so, who does not yet have an output associated to their characteristics (a new sample), will relapse into criminal behavior or not[36].
On the other hand, if the output of the profiling system is a real number, the learning task is called a regression. An example could be with the same database as in the previous example, but in this case the model would predict a positive number representing the recidivism score instead of a “Yes” or a “No”[37].

Classical supervised machine learning techniques include decision trees, which are trees where each node is an attribute, each branch corresponds to an attribute value and each leaf is labeled with a class.



[38]

On the figure above, we can see a decision tree predicting whether a customer will buy a given kind of computer[39].

Decision trees offer certain advantages such as simplicity of design. They are said to be explainable because non-human experts can easily understand why certain predictions were made by the model. However, these models may not be able to capture the complexity of the dataset, which often results in inaccurate predictions.

One problem encountered most of the time with machine learning models and not just with decision trees is the one of “False positives” and “False negatives”, in which an individual that does not fit into a class is wrongly classified by the AI profiling system as being part of that class. For example, a profiling system predicting whether someone is likely to grow up to be a criminal might output the wrong class and this error could have disastrous consequences on the targeted individual[40].

---

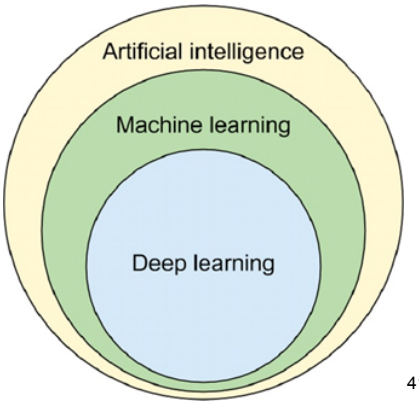
[36] P. Kanani, *Classification of Criminal Recidivism Using Machine Learning Techniques* (2020)
[37] https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

[38] https://zhengtianyu.wordpress.com/2013/12/13/

[39] https://zhengtianyu.wordpress.com/2013/12/13/

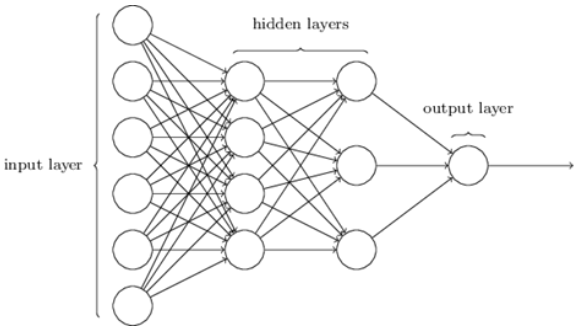
[40] B. Schemer, *The limits of privacy in automated profiling and data mining* (2011)

In recent years, more accurate predictions have been made with the emergence of deep learning models. The deep learning field, which is a subset of machine learning, is inspired by the information processing and neuron connections of the biological brain.



41

Deep learning models are called neural networks and are made of successive layers. Each of these layers play a part in learning the global input-output relationship[42]. The architecture of such models is much more complex than that of a decision tree, but they allow for more accurate predictions.



43

The predictions of neural networks are hard to explain due to the complexity of their architecture. One example of such architecture is given above. To avoid erroneous predictions, some deep learning models output their predictions along with the associated uncertainty rate they have about them to emphasize the fact that the system should maybe not rely on these predictions if the uncertainty is too high and the risks of using these predictions outweigh the benefits[44].

Another problem with predictive AI models, besides the one of misclassification, is that sometimes, the factors on which the output depends is not always present in the dataset[45]. Usually, the more data the more accurate the AI system predictions. However, data subjects have the ability to shield their personal information from data capturing web technologies (for example, disagreeing with the website advertising policy and refusing first-party cookies) and by doing so, they bar websites from collecting their information and this causes the profiling

[41] https://en.wikipedia.org/wiki/Machine_learning

[42] HLEG, *A Definition of AI: Main Capabilities and Scientific Disciplines* (18 Dec. 2018)
[43] https://github.com/d-r-e/multilayer-perceptron
[44] G. Louppe, *Deep Learning*, University of Liège (2020-2021)
[45] B. Schemer, *The limits of privacy in automated profiling and data mining* (2011)

system to output less accurate predictions causing higher False positives rates and False negatives rates. Still, personal targeting can be performed based on the IP address or third-party cookies.

According to the HLEG, "*AI systems must guarantee privacy and data protection throughout a system's entire lifecycle. This includes the information initially provided by the user, as well as the information generated about the user over the course of their interaction with the system (e.g. outputs that the AI system generated for specific users or how users responded to particular recommendations). Digital records of human behavior may allow AI systems to infer not only individuals' preferences, but also their sexual orientation, age, gender, religious or political views. To allow individuals to trust the data gathering process, it must be ensured that data collected about them will not be used to unlawfully or unfairly discriminate against them.*". They must still respect privacy, as humans are subject to the GDPR.

Case studies

Profiling can be applied in many different fields. It is natural that people want to take leverage of the automation and the collection of personal data to better target individuals. We can for example mention political profiling, predictive policing and profiling in advertising. This is of course not an exhaustive list.

Political profiling

Political profiling is the profiling conducted in the context of political campaigns. This makes sense, as candidates may want to target potential voters.

An example of such type of profiling is the one that led to the 2016 Facebook-Cambridge Analytica data scandal:
During the 2016 US presidential elections which opposed Republican candidate Donald Trump to Democratic candidate Hilary Clinton, the London-based data analytics company Cambridge Analytica targeted impressionable voters in swing states in favor of Donald Trump. To do that, they used a Facebook application called "This Is Your Digital" which was developed by psychologist Aleksander Kogan and proposed personality quizzes to their users. With a series of answers, the application determined the user's personality traits. Cambridge Analytica used the users' responses to the quizzes as provided data and their Facebook data, with which they could get additional personal information, as observed data to examine the underlying traits that make up their personalities and to infer whether each user was a persuadable, impressionable voter living in a swing state or not. They built psychographic profiles of the data subjects to assess how they would react to certain advertisements. Then, they would target persuadable voters with advertisements in favor of Donald Trump and the advertisements were personalized with respect to how much the data subject would be open to it.

This political profiling approach was based on the postulate that it is one's personality that informs decision-making and it is one's decisions that drive the way they vote. Two people with the same observed data might not react the same way to the same advertisement [46].

This profiling provoked privacy issues, as Cambridge Analytica also harvested data from the "This Is Your Digital Life" users' friends, who did not consent for the processing of their data. This resulted in up to 87 million Facebook users having their data processed. Following the scandal, the company closed operations[47].

On the 25th of November, the European Commission presented a proposal on the transparency of political advertising. This proposal requires that "*any political advert to be clearly labeled as such and include information such as who paid for it and how much.Political targeting and amplification techniques would need to be explained publicly in unprecedented detail and, would be banned when using sensitive personal data without explicit consent of the individual.* "[48]. This would allow data subjects to have more knowledge about why they see a specific advertisement. According to the GDPR, sensitive personal data include political opinions[49]. With this proposed Regulation, targeting someone based on their political opinion without their consent would be unlawful.

Predictive policing

Profiling can be used for law enforcement to prevent crimes from occurring.

Two famous softwares called PredPol and COMPAS have been used in the United States.

PredPol uses a location-based algorithm. Instead of using or collecting personal data, they are based on historical datasets. The machine learning algorithm cuts a city into multiple 150 by 150 square meters areas and predicts where crime might likely happen during the day. To create their predictions, they use three data variables: the crime type, crime location and crime time. The software was created in 2010 and has been used by several police departments in the United States[50].

Correctional Offender Management Profiling for Alternative Sanctions, known as COMPAS, uses personal data to predict whether a previously incarcerated criminal is likely to commit recidivism once free. It is based on a regression model that outputs a score between 1 and

[46] Karim Amer, Jehane Noujaim, *The Great Hack*. (2019; United States: Netflix)
[47] https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

[48] https://ec.europa.eu/commission/presscorner/detail/en/IP_21_6118
[49] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en

[50] https://www.predpol.com/how-predictive-policing-works/

10 to quantify the individual's likelihood of committing recidivism[51]. It is a decision supporting tool.

Since 2016, both softwares have been accused of inappropriately perpetuating racial bias against neighbors that are more populated with African American and Hispanic individuals in the case of COMPAS[52] and neighborhoods in the case of PredPol[53]. Since then, several police departments across the United States have stopped using PredPol. In December 2021, The Markup and Gizmodo published a report, with graphical support, accusing PredPol of still perpetuating racial biases even after the company promised to be free of them [54]. For example, in the city of Plainfield in New Jersey, the predictive policing software predicted 1940 times more places likely to be the sets of crimes in a neighborhood with a 0% White population than another neighborhood populated with a 63% percentage of White residents.

The problem lies in the datasets the algorithm is trained on. Since they are historical datasets collected by humans, they might contain biases and these biases are reflected in the algorithm predictions.

As for now, none of the two companies have disclosed their machine learning models architectures.

Profiling in advertising: Google's case study

Google has access to a large amount of information about all its users. As they are not supposed to sell directly those data to advertiser under penalty of violating its users' privacy, Google makes money profiling it's user behavior then selling that information to advertisers[55]. From the load of data that Google has on its user, they can infer various useful information for advertisers. It allows companies to perform more targeted advertising and such talk to the group that is most likely to buy their product.
To perform this profiling, Google needs information on users. But it is not a problem for them as they provide a variety of services and it is through them that they collect data on their users. To give some examples[56]:

- Google's search engine can track all their user's searches but also which result they click on.
- Google ADs can also track which ads the user has clicked on in the past.

[51] https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

[52] https://en.wikipedia.org/wiki/COMPAS_(software)

[53] https://en.wikipedia.org/wiki/PredPol

[54] https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them

[55] https://blog.ipleaders.in/how-google-and-facebook-exploit-the-users-personal-data-for-advertisement

[56] Douglas C. Schmidt, Google Data Collection (2018)

- Youtube gives Google a large amount of information about the users viewing habit and interest. It collects the user's search history, subscriptions, comment, watch history and playlist.
- Email sent and received through Gmail as well as the sender and receiver email addresses and the time of sending or receiving are analyzed by Google.
- Google Chrome can track the web behavior. Visited webpages, cookies, browsing history, website-specific permissions, passwords, and add-on data.
- With Google Maps the user's displacement, speed and places visited can be collected. From this, it is possible to infer easily the user's favorite places and the means of transportation he uses.

It is important to note that in 2017 Google announced that they will stop the practice of scanning mail to perform advertising profiling[57].
But the cherry on the top may be Google Analytics, it's the most widely used web analytics service on the web. It's a software that website owners can install on their platform that aims at tracking users' activities. Google analytics only gives access to global statistics on users such as most visited pages, repartition of location of users, repartition of devices used, how users navigate the website. This means that even if, for some reason, you don't use any of Google's services they would still be able to collect information on you if you visit a website using Google Analytics.
Google keeps track of users and their data by storing that information on the user's google account. If the user doesn't have an account or isn't using it when the data is recorded, Google uses cookies to link those data and the user. If you have a google account, you can access the data that Google has on you in your settings, you can also delete them and manage their usage.

Concerning profiling, Google's goal is to predict the maximum amount of information useful for advertisers. Some of these are for example: civil status, interest, revenue, age group, professional field. To do that, Google uses machine learning as it is currently the best way to perform complex tasks with lots of data. More precisely, they are using deep learning, which is a machine learning method that aims at trying to mimic the behavior of neurons. Deep learning yields state of the art performances when dealing with unstructured data and data in text format, which fits the data that Google collects.

[57]https://www.theverge.com/2017/6/23/15862492/google-gmail-advertising-targeting-privacy-cloud-business

Distinction between AI Act and GDPR : What is the relevance of the future regulation on artificial intelligence with regard to the GDPR ?

The proposal of the Regulation (AI Act) "*establishes harmonized rules for the development, placing on the market and use of AI systems in the Union using a risk-based approach"* [58]. This is a new approach in the world of artificial intelligence. These obligations will apply to providers and users of high-risk AI systems[59].

These measures will exist to support innovation, including regulatory sandboxes and in order to " *help small users and suppliers of high-risk AI systems comply with the new rules*" [60].

The enforcement of these harmonized rules will be certified by a governance system at MS level based on the current structures and the creation of a European Artificial Intelligence Committee driven by a cooperation mechanism at EU level[61].

In the AI Act, we find similarities with the GDPR such as the applicability criteria. Indeed, the AI Act would like to incorporate the same structure and philosophy. But, as pointed out above, the proposed Regulation emphasizes innovation, which is not the case with the GDPR in particular. The latter really focuses on everything surrounding the notion of personal data.

Here are the novelties that the AI Act brings:

To start with, the scope of the Regulation is in Article 2 : " *This Regulation applies to :*

(a) providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country ;

(b) users of AI systems located within the Union ;

(c) *providers and users of AI systems that are located in third country, where the output produced by the system is used in the Union* (…) "

Secondly, the proposal of the Regulation defines an artificial intelligence system in Article 3 (1) as " *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* "[62]. This is a unique and future-proof definition of AI[63].

---

[58] 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, p. 4.
[59] *Ibid.*, p. 12.
[60] *Ibid.*
[61] *Ibid.*, p. 4.
[62] 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, p. 39.
[63] *Ibid.*, p. 3.

The AI Act will not override the GDPR, but rather complement it. It is interesting to make a comparison between the two.

Objectives

The objective of the GDPR is to harmonize national privacy laws within the European Union and to give citizens more rights over the use that is made of their personal data.

While the objective of the proposal is :

- ” *Ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values ;*
- *Ensure legal certainty to facilitate investment and innovation in AI ;*
- *Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems ;*
- *Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.* ”[64]

Therefore, the AI Act wants to develop ” trustworthy AI ”, preserving the competitiveness and smooth functioning of the internal market, while relying on ethical technology. So, guaranteeing the protection of individuals and public interests with regard to AI marketed within the European Union.

Prohibition of specific AI systems

The GDPR does not prohibit any activity despite the fact that it imposes strict requirements for certain processing activities. While the AI Act prohibits AI systems that are too risky in all circumstances. These systems that are deemed too risky and ” *relevant to the private sector are those that cause physical or psychological harm to an individual* ”[65].

High-risk AI systems

There is a ” risk-based ” approach, that is to say classification of AI systems according to the risk they can cause to human beings[66]:

- *Unacceptable risk* : a clear threat to EU citizens and introduces a ban of their use;
- *High-risk* : AI systems which may impact humans’ safety or basic rights will be strictly regulated, as further described below;
- *Limited risk* : because of their interaction with humans, they may create certain impact. For those systems, transparency requirements will apply. Users will need to be informed that they are talking to or being serviced by a machine.

[64] *Ibid.*
[65] Privacy and the EU’s Regulation on AI : What’s New and What’s not ?, *in The Journal of Robotics, Artificial Intelligence & Law*. (Consulted on 5 october 2021).
[66] M. KOGUCZAEKOWSKA., What do you need to know about the AI Act ?, *in Timelex*. (Consulted on 5 October 2021).

- *Minimal risk* : other AI, for which no additional obligations are provided in the AI Act[67].

This approach is considered totally innovative and the very first of its kind in the world in terms of AI.

The majority of the requirements of the Regulation apply to high-risk AI systems only. This Regulation lists a number of AI systems that qualify as high-risk[68] like :

1. *Biometric identification and categorization of natural persons :*

    a. *AI systems intended to be used or the 'real-time' and 'post' remote biometric identification of natural persons ;*

2. *Management and operation of critical infrastructure :*

    a. *AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.*

3. *Education and vocational training :*

    a. *AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions ;*

    b. *AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in thests commonly required for admission to educational institutions.*

4. Etc, … [69]

These are the main high-risk AI systems for the private sector, but the Commission can extend this list by applying certain criteria, which can be found in Article 7 of the future Regulation, and a risk assessment method. The possibility of the Commission extending the list of high-risk AI systems is explained by the fact that the future Regulation must be able to adapt to the new uses and applications of AI to ensure legal certainty for everyone.

The Article 7 " Amendments to Annex III " states that : to extend the list in Annex III, the following two conditions must be met :

(a) The AI systems are intended to be used in any of the areas listed in points 1 to 8 of Annex III ;

---

[67] *Ibid.*

[68] Privacy and the EU's Regulation on AI : What's New and What's not ?, *in The Journal of Robotics, Artificial Intelligence & Law*. (Consulted on 5 October 2021).

[69] Annexes to the 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, p. 5 (Annex III).

(b) *The AI systems pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III*[70].

Innovation attracts new risks, so these criteria to apply to high-risk AI systems, not found in the predefined list, helps us to see further, while being legally secure.

Measures in support of innovation

With regard to innovation, Article 53 of the proposed Regulation provides that " 1. *AI regulatory sandboxes (…) shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. (…)* "[71].

This provision would provide a " *legal framework that is innovation-friendly, future-proof and resilient to disruption* "[72]. And this provision is aimed at help SMEs and startups continue to innovate within the new AI rules.

Specific requirements for providers of high-risk AI systems

The GDPR applies to the processing of personal data and not directly to the provider of the systems that enabled this processing[73]. While the AI Act imposes specific requirements on the provider of the high-risk AI system in article 16 of the Regulation, such as :

(a) Ensure that their high-risk AI systems are compliant with the requirements set out in Chapter 2 of this Title ;

(b) Have a quality management system in place which complies with Article 17 ;

(c) Draw-up the technical documentation of the high-risk AI system ;

(d) Etc, …[74]

These obligations include both "*ex-ante* "[75] obligations and obligations that continue after the AI system is put on the market such as regular testing[76].

---

[70] 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, p. 45.
[71] *Ibid.*, p. 77.
[72] *Ibid.*, p. 15.
[73] Privacy and the EU's Regulation on AI : What's New and What's not ?, *in The Journal of Robotics*, Articificial Intelligence & Law. (Consulted on 5 October 2021).
[74] 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, p. 52.
[75] Pre-requisites for the AI system to be put on the market.
[76] M. DESMET., The obligations of providers and users of artificial intelligence systems, *in Md-Lex*. (Consulted on 5 october 2021).

Specific requirements for users of high-risk AI systems

In the proposal of Regulation, there are fewer obligations on users of high-risk AI systems than on providers of such systems. Moreover, these are different from the requirements found in the GDPR[77].

In the article 29 of the proposal of the Regulation, we find the requirements of users of high-risk AI systems are as follows :

1. *Users of high-risk AI systems shall use such systems in accordance with the instructions of use accompanying the systems, pursuant to paragraphs 2 and 5;*

2. *The obligations in paragraph 1 are without prejudice to other user obligations under Union or national law and to the user's discretion in organising its own resources and activities for the purpose of implementing the human oversight measures indicated by the provider ;*

3. *Without prejudice to paragraph 1, to the extent the user exercices control over the input data, that user shall ensure that input data is relevant in view of the intended purpose of the high-risk AI system ;*

4. *Etc,…*[78]

If the user engages in any of these behaviors in Article 29 of the proposal of the Regulation, he risks being considered a provider for the purposes of the Regulation. Therefore, be subject to the obligations of a provider[79].

Specific requirements for certain AI systems

In the proposal of the Regulation, we find in Article 52 transparency obligations for certain AI systems :

1. *Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. « … »*

2. *Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. « … »*

3. *Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other*

[77] Privacy and the EU's Regulation on AI : What's New and What's not ?, in The Journal of Robotics, Articificial Intelligence & Law. (Consulted on 5 October 2021).
[78] 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, p. 58.
[79] M. DESMET., The obligations of providers and users of artificial intelligence systems, in Md-Lex. (Consulted on 5 October 2021).

entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated.

4. *Etc,…*

European Artificial Intelligence Board

At Union level, the proposal establishes a European Artificial Intelligence Board composed of representatives from the Member States and the Commission[80].[81] Each EU member state will designate a competent authority that is responsible for the implementation of the Regulation[82].

The aim of this system is to facilitate the implementation of the Regulation by contributing to effective cooperation between national enforcement authorities and the Commission. But also to provide guidance and advice, and expertise to the Commission to ensure a consistent application of the Regulation[83]. The European Artificial Intelligence Committee will " *collect and share best practices among the Member States* "[84].

The proposal of the Regulation still needs to be adopted by the European Parliament and the Member States. Once the Regulation is adopted, it will " *enter into force on the twentieth day following that of its publication in the Official Journal of the European Union* "[85].

As for the application of the AI Act into the world of artificial intelligence, Article 85 provides for an implementation period of 24 months after its entry into force and a transition period of 12 months[86] for AI systems placed on the EU market prior to the application of the Regulation[87].

[80] 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, p. 15.
[81] *Ibid.*, p. 17.
[82] Privacy and the EU's Regulation on AI : What's New and What's not ?, *in The Journal of Robotics*, Articificial Intelligence & Law. (Consulted on 5 October 2021).
[83] *Ibid.*
[84] 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, p. 15.
[85] *Ibid*., p. 88.
[86] 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, p. 96-97.
[87] Privacy and the EU's Regulation on AI : What's New and What's not ?, *in The Journal of Robotics, Articificial Intelligence & Law*. (Consulted on 5 October 2021).

Are the GDPR and AI Act technically feasible?

From those new regulation we have to ask ourself if they are simply technically doable. In order to do that we will first study the overall transparency of the profiling algorithm. Then in order to gain some insight on the problem of bias the correlation vs causation problem will be tackled. Finally the limitation of profiling systems will be studied.

Transparency and explainability of profiling algorithms

According Article 5§1, a of the GDPR- Lawful, fair and transparent, a person has the right to obtain an explanation of a decision taken by an automated service concerning him. This is an important point because profiling can lead to important decisions such as the refusal of a loan, which are difficult to explain because of the different algorithms.

This is why the GDPR has introduced a fundamental requirement that is found in Article 5 paragraph point a of the Regulation, which is the transparency of processing. Article 12 of the same Regulation states that the controller must provide all information concerning the processing to the data subject in a concise, transparent, comprehensible and easily accessible manner. And to add to that, as said previously, in the Article 52 of the AI Act the transparency is an obligation for certain AI systems.

How can this be developed technically? Nowadays, profiling is performed by machine learning algorithms and the explainability/ transparency of those algorithms may vary in function of their complexity. Take for example a Decision tree algorithm; it is completely transparent, any result output by this model can be easily explained by a series of small questions, such as “is the subject older than 30 ?”, “Does the subject earn more than €1500 a month ?”. The decisions of most simple machine learning algorithms can be understood by non-experts. On the other hand, for complex machine learning algorithms, even experts cannot understand, at least directly, the motivation behind their outputs. Due to this, these algorithms are often referred to as “black box”, the complete opposite of transparency. Unfortunately, complex algorithms are the most popular as they are more effective when dealing with elaborate tasks and a large amount of data.

Transparency and explainability aren’t only interesting to fit the GDPR requirements. They are various benefits to a better understanding of the behavior of the AI model used.

- Explanation of the model to a field's expert. In some applications, a validation of the model from experts of the said application is necessary. This is crucial in the medical field, and to do so, the motivation behind the decision of the model must be explainable.
- Improve end point’s user comprehension
- Ease the debugging of the model.
- Analyze and understand biases that the model can suffer. Those biases can lead to serious problems such as discrimination.

Those benefits lead to improved performances, an increased trust from the users, an identification of biases and of course a compliance to the GDPR rules. Thus, if possible, incorporating an explainability analysis has no downside when designing a model.

Explainability techniques are still an open question in the machine learning world. Some techniques exist, but there is still room for improvement. The most popular ones are LIME[88], SHAP[89] and PDP[90] which can be applied to any kind of machine learning algorithm[91]. They are based on using variation of the inputs and analyzing the relative output of the algorithm. On the other hand, in the field of deep learning there is a lot of research on explainability but these new techniques are only applicable on deep neural networks. The reason for that is that even though deep learning methods yield very complex models there is a technical specificity to them that allows a more detailed analysis.

SHAP and LIME are very similar, both aim at giving insight on a particular prediction of a model. They do so by doing multiple little modifications to the input relative to that said prediction. They are then able to approximate locally that model and derive the importance of each feature constituting the input in the prediction. Where LIME and SHAP are different is in the way this approximation is made. They allow one to answer the question “Which variables caused the prediction?”. Depending on the features used for the prediction, any non AI expert can get some insight on the model’s decision.
Here is an example of LIME in profiling, where the machine learning model has to predict whether someone is an atheist or a christian based on an email.

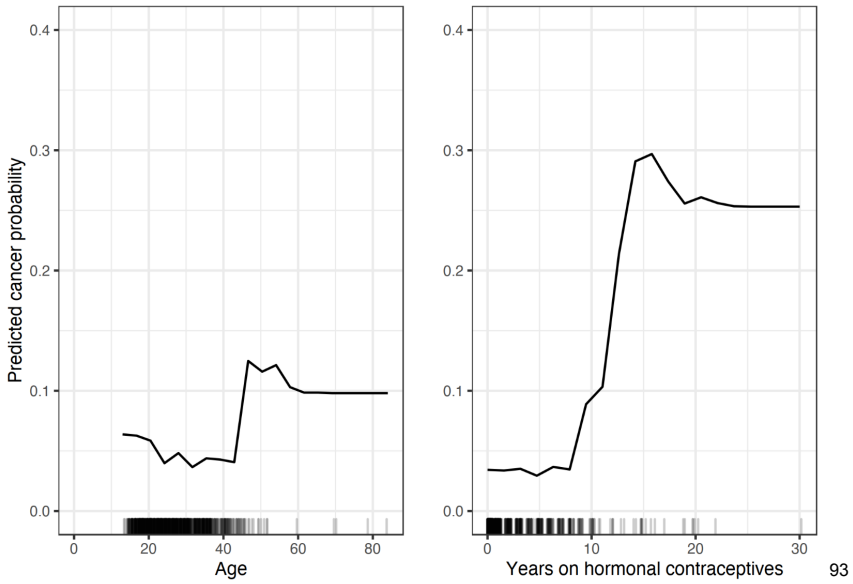
92

Words highlighted are considered by the model as the most significant words in the decision process. The model predicts atheism for that specific person, but anyone could tell from the LIME analysis that the prediction is not well motivated. Indeed, the most important word for the predictions aren’t relevant. It is important to note that in this example, the features are words which are understandable by any English literate person. Obviously, this is not the case for all applications, most of the time features are complex and understandable by experts only.

PDP is a rather old (2001) technique that aims at analyzing the global behavior of machine learning models. It answers the question “what is the relationship between the variable and the general prediction?”. Unlike LIME and SHAP, PDP does not inform on the direct reason

[88] Ribeiro et. al’s, Why Should I Trust You?(2016)
[89] Scott M. Lundberg and Su-In Lee, A Unified Approach to Interpreting Model Predictions (2017)
[90] Jerome H. Friedman, Greedy function approximation: A gradient boosting machine (2001)
[91] https://towardsdatascience.com/picking-an-explainability-technique-48e807d687b9
[92] https://homes.cs.washington.edu/~marcotcr/blog/lime/

of a specific prediction. PDP analysis provides transparency of the overall model, it gives insight on the black box itself and not its output. It is useful for evaluating the legitimacy and degree of bias of a model as the absence or presence of a relationship between one of the input's features and the prediction is often expected. For example, there shouldn't be any link between a person's sex and their ability to take a loan. As such PDP is more directed at the application's field expert as they have the intuition on the issue.
Here is an example of PDP analysis on a machine learning model aiming at predicting cervical cancer among women, only two variables are featured in this analysis but more were used:



93

From these relationships any non-expert could deduce that the older a woman is and the more years on hormonal contraceptives a woman has the more likely she is to develop a cancer. But is this conclusion right? The link between age and cancer likeliness is fairly obvious, but this is not the case for the one between contraceptives and cancer. Only a person with enough knowledge in the field evaluates the veracity of this relationship and assess the model's quality.

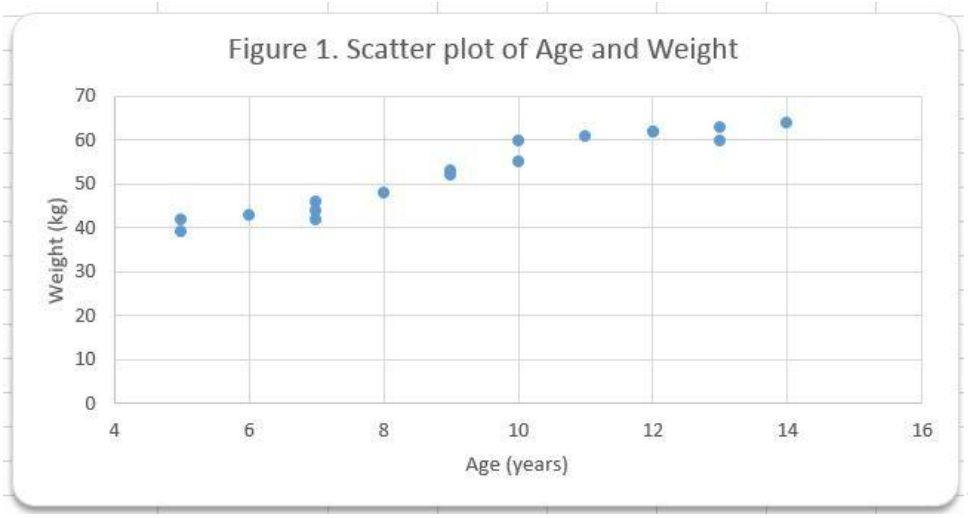
LIME, SHAP and PDP do not state why a prediction was made, but more what influenced that prediction. It is left to the user to interpret the "why was that prediction made" and its legitimacy. Telling someone that has been refused a loan by an automated service that his civil status and address played a big part in the decision isn't really achieving explainability or transparency. This is why a competent supervisor is needed in order to interpret the result and evaluate the legitimacy of the prediction.

[93] https://christophm.github.io/interpretable-ml-book/pdp.html

Correlation and causation

Before going further, it is interesting to discuss the concept of correlation since it is particularly exploited by profiling systems. Correlation is a statistical metric computed from two or more variables. This metric is used to analyze the relationship between those variables. On the other hand, it is said that two variables have a causal relationship if a variation in one variable causes a change in another.

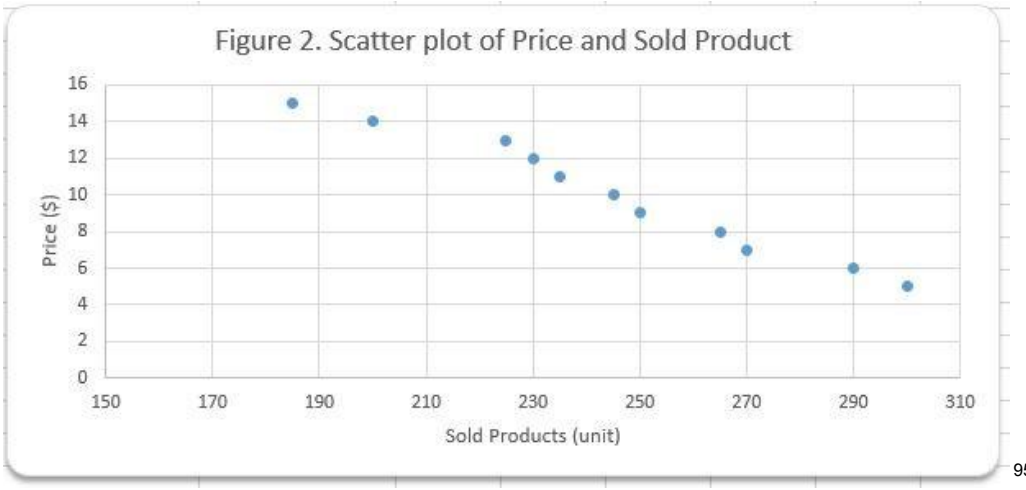
Multiple measures of correlation exist in statistics but the ones that are mostly used, partly due to their simplicity, are linear correlations. With those metrics, two variables can be either positively correlated if they increase or decrease together, negatively correlated if when one increases the other decreases or uncorrelated if no link exists between the variables. To illustrate this here are some examples of correlations:



[94]

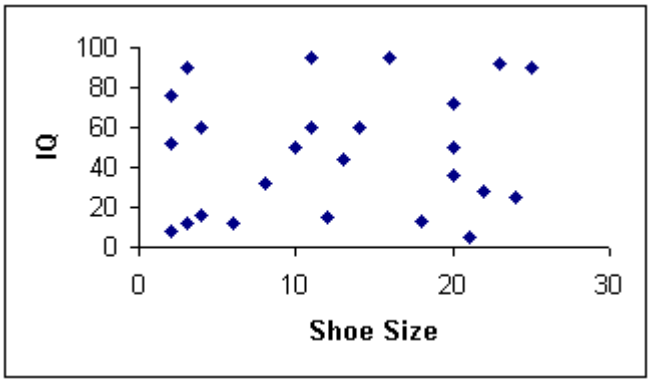
The figure above is an example of positive correlation. Each point corresponds to a young person's age and weight. Older people have greater weight, which is perfectly expected.

[94] https://worldsustainable.org/analysis-of-correlation/



[95]

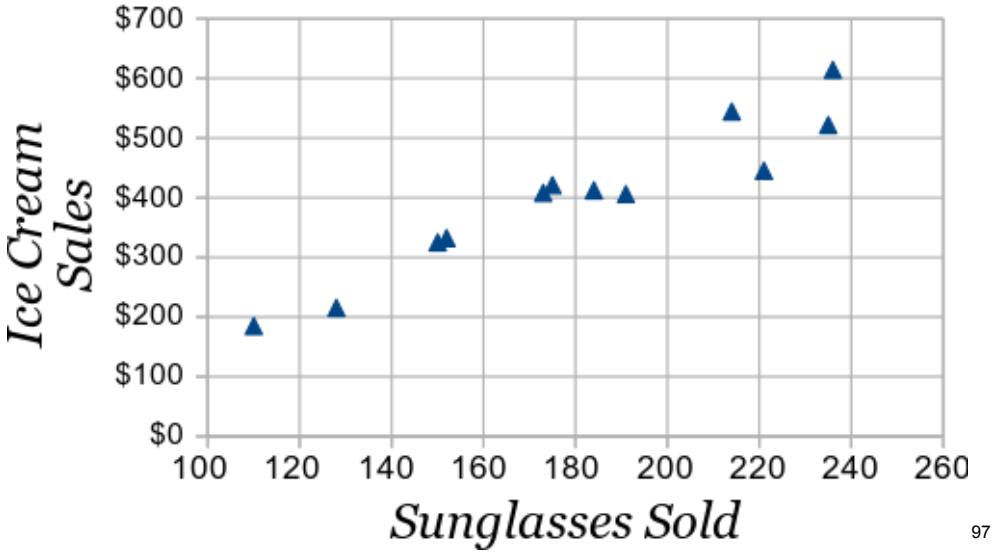
This graph illustrates a negative correlation between the two variables. Each point corresponds to an item price and it's number of units sold at that price. The lower the price is the larger the number of sales, which is natural since more people can afford the product.



[96]

And finally an example of uncorrelation. This figure shows the relation between the IQ of people and their shoe size, each person corresponds to a point. The graph looks like a cluster of dots without any real structure. No relationship exists between the variables.
In these examples correlation seems tightly connected to causality. But this is not generally the case, there are numerous reasons as to why a correlation can appear without being directly linked to causality. One of the most common reasons is that both variables share a common cause. Here is an example of that.

[95] *Ibid*.
[96] https://www2.palomar.edu/users/rmorrissette/lectures/stats/correlation/correlation.htm



[97]

On this graph each point corresponds to the number of sales of ice cream and sunglasses for a day. This graph shows a positive correlation between the two variables. Are those variable causally related ? Certainly not but they do share a common causal source. The brighter the sun shines, the more people will buy sunglasses and the hotter it will be. The hotter it will be, the more likely people are to buy ice cream. In this case, ice cream and sunglasses sales are connected by the intensity of the sun but they still positively correlate. This phenomena of shared causal source is the most common one and is sometimes very difficult to notice and has been used to construct deceptively pertinent arguments.

Correlation does not entail causation and at least for now there exists no mathematical way to deduce the causality between variables. This concept of causality is inherently connected to the context and the meaning of the variables. From pure numbers it is impossible to infer a causal link between variables. If you wish to see more examples of this, here is a site dedicated to illustrate weird correlations: https://www.tylervigen.com/spurious-correlations.

This is specifically important for profiling tasks. Most automated decision algorithm uses in profiling are based on statistics or machine learning techniques, which are just in essence smarter statistical methods. Those algorithms don't have access to the context or direct meaning behind variables thus they rely on correlations. This can potentially lead to some problems such as an unjustified bias in the decision making process. For example, in the case of PredPol, the number of African American people and crime in a place correlates. Can we say that there is a causal link between African american populations and crime ? No, as said earlier correlation doesn't entail causation. This correlation exists because of a common causal source between variables. African American tend to habit poorer suburbs and poorer suburbs have the highest crime rate. This reflection cannot be made alone by an automated decision algorithme. It will establish a causal link between african american population and crime rate and thus a bias is created.

[97] https://medium.com/@seema.singh/why-correlation-does-not-imply-causation-5b99790df07e

Profiling system's limitation

Let's discuss a brief overview of the current profiling systems' limitations and how they can be overcome.

Firstly, in order to have the perfect profiling system an infinite amount of data is necessary and even then its accuracy won't be perfect for complex applications, it will be very close to but not perfect. And this perfect profiling system is unachievable in practice thus those they are doomed to be imperfect, even without their potential biases. Depending on the applications, errors can have devastating consequences and therefore need human supervision. This human supervision must be aware of the flaws of the profiling system, highlighting the importance of transparency.

Secondly, out of distribution evaluation is a serious issue in profiling. It can happen when the context of the dataset used is somewhat different then the context of the application. In this case the profiling system performances are completely undetermined. This problem can be very subtle and often happens in medical applications as in this field dataset are rare and heavily restricted. This issue can also be the source of bias, specifically when the representation of some feature in the dataset isn't faithful to its representation in the application. This was in fact the issue with the famous Amazon problems, in their initial dataset of resume, employed women were underrepresented and this led to an out of distribution issue. To avoid such things the usage of data augmentation is possible, which consists in adding data fabricated to some degrees in the dataset used. Unfortunately, this method can be complex and impossible in some applications.

Finally, as illustrated previously, biases can appear when creating a profiling system. And this is, in regards to the law, unacceptable. But no system will randomly decide to discriminate against a group of people on its own. It always comes from the data. Thus in order to prevent this issue an analysis of the quality of the data must be done. Correlations methods and others can be used to detect potential bias in data but most of the time they aren't enough. To really be sure that no unfair discrimination can be drawn from the data an exhaustive analysis is necessary but with the size and complexity of the dataset used nowadays, it is most often impossible.

Therefore undetected biases may appear in the profiling system. It is then a critical issue for them if they exist. The basic approach for evaluating the performances of any predictive system is to split the initial dataset into two parts, one dedicated for training and another dedicated for testing. But since both parts come from the same origin, if a bias exists in the training set it will most likely than not exist in the testing set and thus the said bias will not be detected by the system evaluation. In order to highlight biases a more complex analysis is needed. And for that transparency methods can be used as discussed previously. One good way to also detect possible discrimination is simply to test the system on real situation, but this is sometimes not doable.

Once a bias is identified, it must be eliminated. And this task is maybe the most complicated or even impossible. To do that either the initial data needed to be modified or some kind of mechanism must be added to the design of the profiling system. Modifying the data can be

simply done by removing the source of the bias, for example: removing the gender for a profiling system with sexist bias. But most often then not this is not sufficient as the bias exists within hidden relation in the data. Regarding the other method, modifying the design in itself of the system is fairly new in the field and is still in progress but some mechanisms do exist such as the paper[98] : Mitigating Gender Bias for Neural Dialogue Generation with Adversarial Learning. This explains the construction of a chatbot with mitigated gender bias.

[98] Haochen Liu et al. Mitigating Gender Bias for Neural Dialogue Generation with Adversarial Learning (2020)

Conclusion

In this report, we have explained in a general way what profiling is from a legal point of view and how profiling systems work in practice, as well as several problematic case studies which highlighted gender discrimination, alleged race discrimination and lack of transparency on the part of administrative or private actors but also political manipulation coupled with privacy issues. And also but not the least, the infringement of our fundamentals freedoms.

To respond to these problems, the European Union has intervened by adopting the GDPR which incorporates fundamental principles of users' protection. The importance of creating a framework around "risk" is because the specific characteristics of some AI systems can create new "risks".
Recently, the European Commission has proposed the AI Act before the Economic and Social Committee to reinforce legal certainty in societies and thus overcome the lack of trust of citizens[99]. Therefore, in order to avoid patchwork responses, the AI regulation could allow us to trust profiling systems[100]. We then approached the technical feasibility of such restrictions. The major question that we are currently asking ourselves is how the world will react when the AI act comes into force?

From what has been mentioned before, we can conclude that the two Regulations can be placed side by side for optimal use.

[99] A Europe fit for the digital age : Commission proposes new rules and actions for excellence and trust in artificial intelligence, EU. A Europe fit for the digital age: Artificial intelligence (europa.eu)
[100] *Ibid.*