Profiling and Convention 108+

Frenay, Benoît; Poullet, Yves

Publication date:
2019

Document Version
le PDF de l'éditeur

Strasbourg, 18 november 2019

T-PD(2019)07rev

CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION

OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING

OF PERSONAL DATA

Convention 108

Profiling and Convention 108+ : Report on developments after the adoption of Recommendation (2010)13 on profiling

DG I – Directorate General of Human Rights and Rule of Law

Section 1: recent developments, technical aspects, players and typology of profiling technologies[1][2]

1.Developments in context and in profiling

Profiling has come a long way since the 2010 recommendation. For one thing, the explosion in technologies such as deep learning has brought analyses that were previously impossible well within reach. At the same time, people have become aware of the opportunities and risks presented by those very technologies for society and individuals. This report begins with a short review of these developments.

1.1 Factors driving the technological changes

The technologies that are currently revolutionising profiling did not just appear from one day to the next. The concept of artificial intelligence saw the light of day in 1956 at the Dartmouth conference, and the term "machine learning" was suggested by Arthur Samuel in 1959. Deep learning can be traced back to the 1980s. There are various reasons why these research-generated technologies have boomed in recent years.

Firstly, research itself has made leaps and bounds in these spheres. Powerful algorithms have been developed to analyse large quantities of data. In their 2012 article "ImageNet Classification with Deep Convolutional Neural Networks"[3], which rekindled interest in deep learning, Krizhevsky, Sutskever and Hinton trained a network of 650,000 neurones with 60 million parameters from ImageNet (1,200,000 images from 1,000 different classes). By way of comparison, the LeNet-5[4] neural network proposed by Lecun, Bottou, Bengio and Haffner in 1998 had only 60,000 parameters and could only recognise digits. This technical prowess made it possible to reduce the error rate in image recognition by 11%. Since then, a whole host of neural network architectures have been put forward.

Secondly, data are currently available on an unprecedented scale. On the one hand, varied collections of images, texts, sounds, raw data etc have been made public by various entities (research laboratories, private companies, public bodies, international organisations etc). Examples include ImageNet[5] as well

[1] This first section was written by Professor Benoît FRENAY, IT specialist, under the supervision and with the assistance of Professor Yves POULLET

[2] Some companies or logos are used to give the reader a clear idea in the framework of this report. This is not a judgement on the importance of these or the quality of the products offered.

[3] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, 1097-1105.

[4] LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2323. https://doi.org/10.1109/5.726791

[5] http://www.image-net.org/

as the open data of cities such as Paris[6], London[7], New York[8] or Namur[9], sites such as Wikipedia[10] which supply enormous quantities of text, the YouTube-8M dataset[11] with its 237,000 video segments published by Google, the International Collaboration on Cancer Reporting[12] which seeks to produce standardised datasets for different types of cancer and makes recommendations along these lines[13] etc. On the other hand, those entities are building their in-house data collections for their own use. It may be a public body seeking to improve the public services it provides, a private company wishing to improve or sell its products, an on-line multimedia service etc. In the second part of this document we will take another look at the great diversity of possible uses of profiling and profilers.

Thirdly, developments in hardware or software technologies have opened up the possibilities for running resource-hungry algorithms on large quantities of data. In 2012, Krizhevsky, Sutskever and Hinton were already using graphics processing units or GPUs, graphics cards initially developed for PCs (video games, graphics rendering etc). GPUs are used extensively to accelerate deep learning computations by several orders of magnitude and to quickly train new neural networks. While the storage of large quantities of data is nothing new (one example being Teradata founded in 1979), other technologies emerged in the 2000s, such as MapReduce (2004) and its open-source implementation Hadoop (2006), and are inextricably linked to the "Big Data phenomenon".

Fourthly, for private stakeholders, the data are an inestimable source of information and, in some cases, constitute their core activity, so it is not surprising that they are investing heavily in this area. Some corporations are actively involved in research into artificial intelligence and new technology development, with considerable budgets devoted to these activities.

In conclusion, the rapid developments of the last fifteen years or so is explained by the convergence of a number of factors: efficient algorithms can now exploit large quantities of data thanks to hardware and software technologies in turn made possible by research and substantial investment by both public stakeholders (including backers such as the European Union with Horizon 2020) and private stakeholders. These developments have paved the way for innovations that are beneficial for individuals and the community but also carry risks at both individual and community level (see section 2).

1.2 Perception and impact of technological change

The development of artificial intelligence, machine learning and deep learning have had a substantial impact on profiling. While profiling did not necessarily require the use of these technologies, they now make it possible to exploit personal data far more effectively.

The rapid progress of smart technologies and the enthusiasm for them have naturally led to them being applied in numerous contexts in recent years. On the one hand, this is enabling us to tackle problems that were previously difficult to resolve, effectively and on a large scale (automatic diagnosis

[6] https://opendata.paris.fr
[7] https://data.london.gov.uk/
[8] https://opendata.cityofnewyork.us/
[9] https://data.namur.be
[10] https://en.wikipedia.org/wiki/Academic_studies_about_Wikipedia
[11] https://research.google.com/youtube8m/
[12] http://www.iccr-cancer.org/
[13] https://www.ncbi.nlm.nih.gov/pubmed/27735079

of skin cancer with a smartphone[14], fraud detection, personalised advice on sales or multimedia platforms etc). On the other hand, the widespread use of artificial intelligence for profiling has highlighted issues connected to information technology, sociology, ethics, law etc: risks run by employees in recruitment or their career, ethical use of profiling, rights and obligations linked to data use, developing more transparent and more easily interpretable algorithms, reducing problems of bias and discrimination, making algorithms more resistant to noise and attacks etc.

In parallel, public and private players, including the media, are exposing the uninitiated to these technologies, as they use them and receive information through them. The claims made regarding these technologies come from both ends of the spectrum. On the one hand, people are told about self-driving cars, artificial intelligence robots playing Go or poker and personal assistants for day-to-day living. On the other hand, they are warned of potential abuses, threats to democracy, discrimination and job losses. It is difficult for people to know what to think, and this is a threat to public debate, which can no longer take place in the right conditions. And when they are bombarded by the hard sell, people also have trouble working out the real capabilities and limits of the systems being sold to them. On a broader scale, the decision-makers within public and private bodies are sometimes equally confused. Moreover, it is not easy to find technical and non-technical staff who are trained in these technologies. There is a real societal problem of education in artificial intelligence and, in particular, the specific issues of profiling.

Society's swift and widespread uptake of artificial intelligence technologies for profiling therefore lays bare unprecedented potential uses, while raising technical and non-technical questions and an urgent need for education at all levels of society.

1.3 Responses to technological change

There have been various responses in recent years to artificial intelligence and its impact on profiling. On the one hand, regulations are applicable and reports have been drawn up at different levels, such as the GDPR Regulation and the "Ethics guidelines for trustworthy AI" report at European level, the "AI for humanity" report in France or AI4Belgium in Belgium. On the other hand, funding has been set up to support research development, such as the recent "H2020 Call on European Network of Artificial Intelligence Excellence Centres".

Training has been put in place at different levels (universities, colleges, training centres etc) to train employees to cope with the challenges of profiling and artificial intelligence. Discussions are under way with a view to educating children, teenagers and the general public, notably in Belgium, Finland, France and the Netherlands. Numerous research laboratories and corporations are looking in tandem at the positive impact that profiling can have on teaching: personalised pathways for pupils with exercises and lessons tailored to their profile, detection and prevention of school drop-out on the basis of the pupil's results and activities etc.

In the world of research, the issues mentioned above are tackled head-on: many scientific conferences include sessions devoted to problems of bias, "interpretability", reliability, ethics, security, preserving anonymity etc. At the same time, corporations have also been made aware of these issues and are

[14] Esteva, Andre & Kuprel, Brett & Novoa, Roberto & Ko, Justin & M Swetter, Susan & M Blau, Helen & Thrun, Sebastian. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature. 542. 10.1038/nature21056.

making efforts accordingly. Initiatives such as the UN "AI for Good" platform promote the beneficial use of artificial intelligence. Despite all these initiatives, it is clear that the existing problems are far from being resolved. At present it is difficult to ensure perfect "interpretability", reliability or security for many of the artificial intelligence systems used in profiling. Research is more necessary than ever, particularly in the IT field, from which the technical solutions will come. But an interdisciplinary approach is indispensable to better meet those challenges.

In conclusion, the technological developments of recent years have prompted numerous responses in the spheres of politics, law, education, research and business. One tendency observed in recent times is particularly noteworthy. Many researchers in the fields of artificial intelligence, machine learning and deep learning are leaving research laboratories for private-sector companies. While this movement is normal and desirable, it is on an unprecedented scale. In the artificial intelligence sphere, we are seeing a real privatisation of research. It is becoming difficult for universities to keep their best talent. An ambitious policy of support for independent research is now needed more than ever so that Europe remains the scientific leader in this area. Meeting the challenges of profiling calls for strong, perennial fundamental and applied research, which reports its findings and progress and accepts peer discussion in return.

Chapter 2. Players

Using artificial intelligence, machine learning and deep learning for profiling makes setting up new projects a more complex business. It is now rare for a project to involve just one player to create the necessary algorithms and adapt, configure and use them. The data also have to be acquired, stored and organised. What is more, the profiling components are usually only one part of a far larger system in which they must be integrated. This section looks at a classic example of using machine learning and the variants tailored to different scenarios involving a variable number of different types of players.

2.1 Base scenario: a single player

The artificial intelligence techniques used for profiling fall chiefly within the domain of machine learning and, in some cases, more specifically deep learning, a sub-discipline of machine learning that is particularly useful when images, video, sound or text are processed. Figure 1 shows the main phases of machine learning: training and prediction.

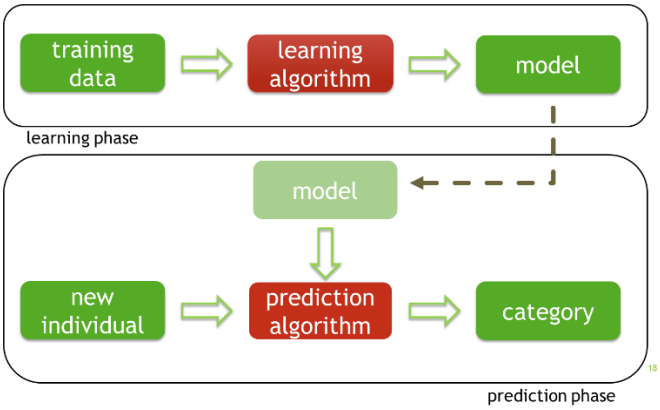


Figure 1: training (learning phase) and use (prediction phase) of a machine learning model.

Training entails exploiting a dataset in order to learn a model using those data. The model uses mathematical abstraction providing a simplified description of the data to resolve the task at hand. For example, an estate agency might want to predict the price of property according to its surface area, the number of bedrooms, the presence of nearby shops, noise pollution indicators, the age of the property etc. In such a case, it will have data on numerous properties it has already sold, for which the sale price will be known: these are the training data, which it will then use to arrive at a formula along the lines of:

$$\begin{aligned} price = {} & parameter_{\,surface\ area} \times surface + parameter_{\,bedrooms} \times bedrooms \\ & + parameter_{\,shops} \times shops + parameter_{\,\mathrm{pollution}} \times \mathrm{pollution} \\ & + parameter_{\,\mathrm{age}} \times age \end{aligned}$$

in which the value of the parameter of each characteristic is initially unknown. The training will involve looking for the best values of these parameters to find the closest possible match to the prices observed for the properties previously sold. The hypothesis is that the model (ie the formula with the best parameter values) will make it possible to correctly estimate the price of properties coming onto the market. A linear model of this kind is too simple to be able to provide a perfect explanation of market prices but it will probably yield an initial estimate that will be accurate enough for the agency's needs.

The prediction phase will use the model taught to the machine using the data to make predictions for newly available property. In practice, a model can predict a number, but also a category, which is more common in profiling. Accordingly, we can use the data available on a large number of customers to teach a customer retention model making it possible to predict whether there is a risk of a customer going over to the competition. The model may be a linear formula, as above, but it may also take the form of logic rules like figure 2 or be rather more complex, such as the Inception-v3 neural network whose architecture is shown in figure 3.

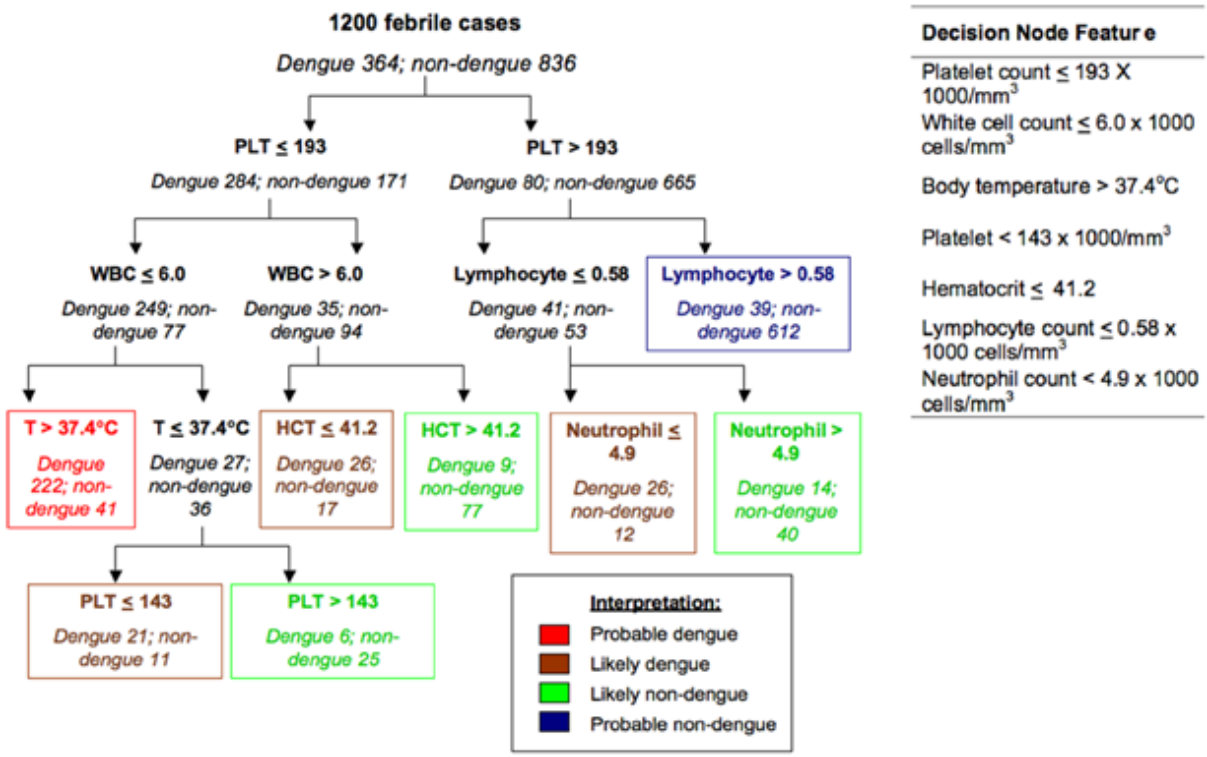


Figure 2: example of a model of categories expressed in the form of logic rules organised in a decision tree[15].

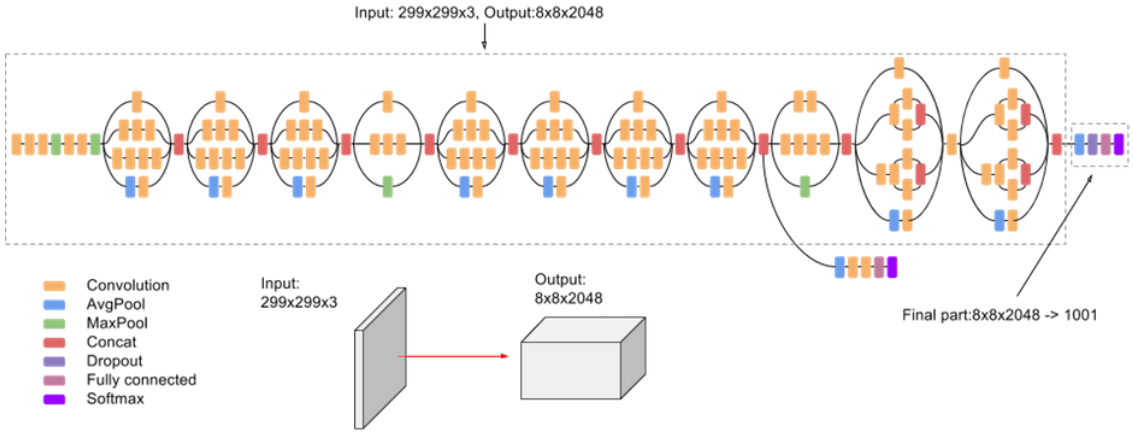


Figure 3: example of a neural network (deep learning) with a parameter count of 23,885,392 (Google's Inception-v3)[16].

In the scenario presented here, only one player is involved: the estate agency which acquires, stores and manages the data, devises and uses machine learning algorithms and then deploys them. In practice, profiling is rarely down to a single player and the situation is far more complex.

Before going any further, the distinction should be noted between artificial intelligence, machine learning and deep learning. Artificial intelligence is an IT discipline which creates new algorithms capable of resolving problems normally requiring human intelligence. This covers a variety of tasks

[15] Figure reproduced from Tanner, L; Schreiber, M; Low, JG; Ong, A; Tolfvenstam, T; Lai, YL; Ng, LC; Leo, YS; Thi Puong,L; Vasudevan, SG; +3 more... Simmons, CP; Hibberd, ML; Ooi, EE; (2008) Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS neglected tropical diseases, 2 (3). e196. ISSN 1935-2727 DOI: https://doi.org/10.1371/journal.pntd.0000196.

[16] Figure reproduced from https://cloud.google.com/tpu/docs/inception-v3-advanced.

such as planning, games, meeting constraints, logical reasoning (notably expert systems) or probabilistic reasoning, processing of the spoken word, text or images etc. Among the artificial intelligence techniques, machine learning has the specific characteristic of being able to exploit the available data to allow the creation of artificial intelligence which learns. More specifically, when the data are images, sound or text, deep learning, a sub-domain of machine learning, is commonly used to create networks of neurones geared to model these types of data.

2.2 Using specialised libraries

Figure 4 takes the previously developed scenario with a focus on training. A single player is involved: the one providing the final service incorporating profiling. Even for a machine learning specialist, it is rare to not at least use support from specialised libraries, as figure 5 shows. A library is a set of ready-to-use functionalities (in other words, implementations of algorithms) for the easy creation of new programmes. In this way, when a development team uses a library for a new project, it avoids having to reinvent the wheel and (re)develop a whole host of commonly used algorithms. For example, experts would not implement a model such as support vector machines themselves; they would instead use the LIBSVM library which provides effective implementation of that model. A model like this has been the subject of thousands of scientific publications and is very tricky to implement.

Many machine learning algorithms are already implemented (ie made available) in free open source libraries such as LIBSVM, LIBLINEAR, scikit-learn, Weka, Keras, TensorFlow, PyTorch or their commercial equivalents. These libraries are extraordinary time-savers, making it possible to achieve competitive results. Effective and reliable implementation of many machine learning algorithms requires substantial expertise and a considerable amount of time, often hinging on decades of research.

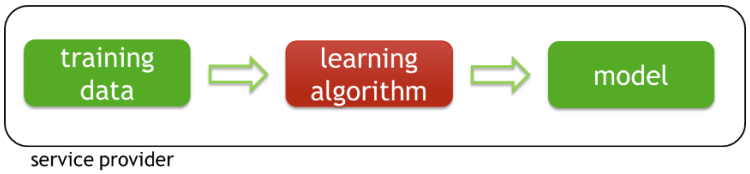


Figure 4: training phase where there is only one player.

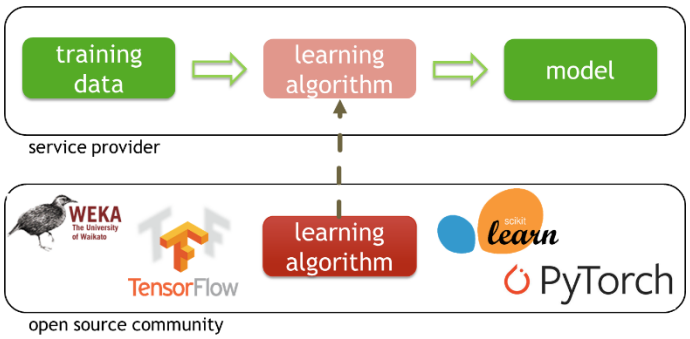


Figure 5: training phase using specialised libraries.

The "libraries" mentioned above are devised thanks to the considerable work carried out by large non-profit communities or corporations. In the case of open source and free libraries, they are available on the Internet and can be downloaded by any actor who wishes to use them. They are usually accompanied by a disclaimer that they provide no guarantees (in the legal sense) and must be used with all due precaution. It is impossible to overstate the importance of free open source libraries on

machine learning, which are very widely used and without which most current-day developments would not exist, particularly in deep learning. Although they are non-commercial in nature, the fact that they have been developed by big communities is often a guarantee of quality. Our recommendations will have to take account of the need for these libraries to be developed, an undeniable factor in innovation, and look to self-regulation in scientific research and researcher communities for safeguards against the risks linked to their use for profiling.

2.3 Subcontracting learning

In this case, the player wishing to use a profiling service does not have the expertise to do so and may therefore call on the services of one or more players who will design algorithms tailored to the needs, thus delegating the training phase as shown in figure 6. Obviously, the players will follow the subcontracting rules laid down.

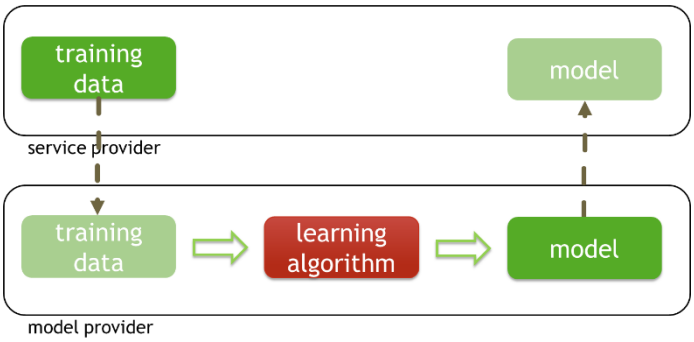


Figure 6: delegating the training phase to an outside provider.

There are a number of possible reasons for delegating the training phase. Even when the technical expertise is there, the sheer quantity of data used in profiling may be such (people often talk of "Big Data", even if the term only really applies to a few major players) that a specific infrastructure must be set up or used, either to store the information or to analyse it. That calls for the deployment of hardware and software resources requiring skills lying outside the field of machine learning. One extreme (but common) case is the use of "cloud"-style platforms where a player pays for access to substantial storage and computing resources to run their own algorithms. And, in some cases, it may simply be more efficient to call on a profiling specialist.

2.4 The special case of deep learning

In the case of deep learning, the problems involved in designing and running algorithms are even more substantial. To train the Inception-v3 network shown in figure 3, Google's researchers had to use 50 GPUs. Each GPU can cost up to several thousand euros, outlay that is well beyond the reach of an SMB. In addition, designing suitable neural network architecture requires sound experience of deep learning. For these reasons, it is difficult for a small player to develop their own deep learning model.

In practice, for profiling activities entailing the recognition of images similar to those found in image datasets, you do not need to create your own model. Some major players have made models they have trained using large collections of images freely available at no cost, enabling smaller players to download and use them straightaway, as in figure 7, without having to bear the training costs. This is common practice in the world of research and teaching, particularly in the study of deep learning or for quickly and cheaply testing out an idea. Obviously, the downside is that only the categories opted

for in the design of the model will be recognised. A network like DeepLoc[17] trained using images of yeast cultures under the microscope will be useless for distinguishing between cats and dogs. Likewise, Inception-v3 could not be used instead of DeepLoc as it recognises only "natural" images (dog, cat, tree etc).

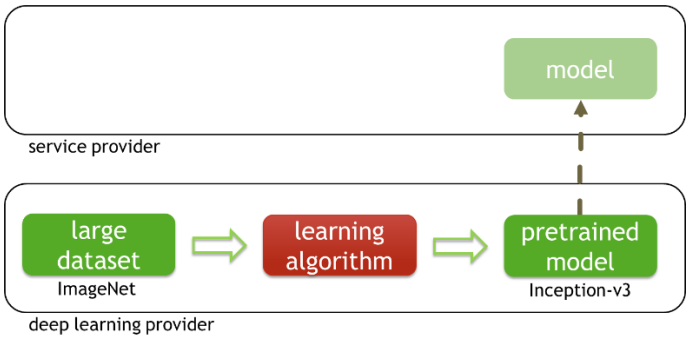


Figure 7: using a pre-trained model.

Numerous free-access networks have been trained using the same set of images regarded as authoritative in the scientific community: ImageNet. As shown in figure 8, this adds in a new player: the entity which devises and makes available the image dataset, without associating any models with it.

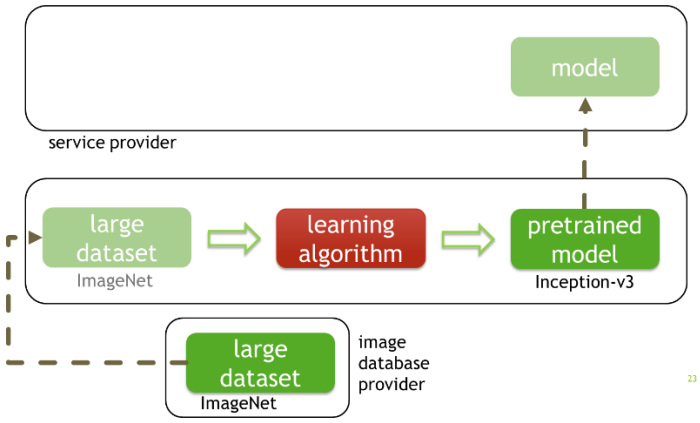


Figure 8: addition of the image dataset provider.

For machine learning to work, the image datasets have to contain images that are sufficiently varied and associated with predefined categories. ImageNet contains 1,200,000 images belonging to 1,000 different classes. However, the conditions in which the images are collected and associated with categories can have a considerable impact on the models that will be developed and used by the other players. Figure 9 shows the breakdown of the geographic origin of the images: there is a clear bias in representativity[18] which explains why certain images are not well recognised by neural networks trained using ImageNet. The same problem has been encountered with a number of commercial facial recognition systems trained using collections chiefly made up of male Caucasians[19] (see figure 10). As

[17] Kraus, Oren & T Grys, Ben & Ba, Jimmy & Chong, Yolanda & J Frey, Brendan & Boone, Charles & J Andrews, Brenda. (2017). Automated analysis of high-content microscopy data with deep learning. Molecular Systems Biology. 13. 924. 10.15252/msb.20177551.

[18] Shankar, Shreya, et al. "No classification without representation: Assessing geodiversity issues in open data sets for the developing world." arXiv:1711.08536 (2017).

[19] Buolamwini, J. & Gebru, T.. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proc. FAT in PMLR 81:77-91

pointed out previously, image datasets like ImageNet are made available without any (legal) guarantee and it is for the users of those images to check that they are suitable for the intended application. ImageNet is described in detail in "ImageNet: A Large-Scale Hierarchical Image Database"[20] and "Construction and Analysis of a Large Scale Image Ontology"[21] and has been used in many studies. In the medical sphere, datasets may be guided by recommendations, such as those of the International Collaboration on Cancer Reporting[22].

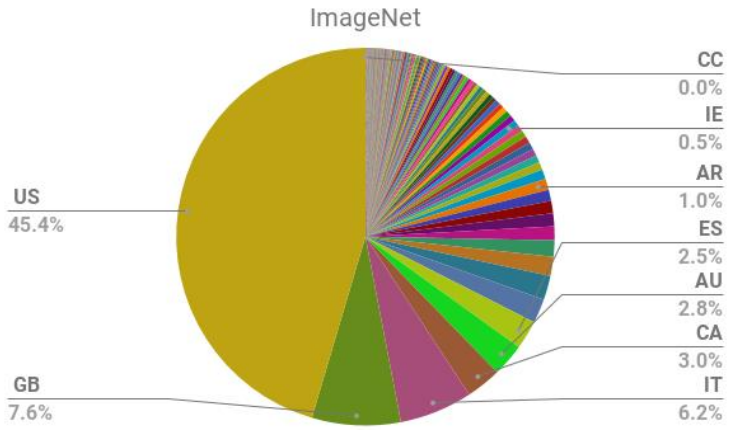


Figure 9: breakdown of the geographic origin of the images in ImageNet[23].

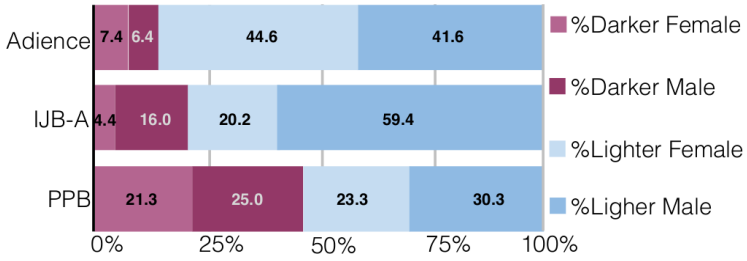


Figure 10: content of three image datasets used to train facial recognition systems[24].

2.5 The pre-trained models revolution

In many cases, it is not enough to use an existing deep learning model without tweaking it. However, it is rare to have the quantity of images needed to train a "deep learning" neural network: a collection of a few hundred or a few thousand images is nowhere near enough but this is very often the size of the datasets available to a player wishing to develop a service using deep learning. The images do not only have to be acquired but also be manually labelled one by one. In a medical context, for example, the process is long and costly because of the expertise and equipment required. Moreover, the number of patients available for such research will be limited.

---

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, **ImageNet: A Large-Scale Hierarchical Image Database.** *IEEE Computer Vision and Pattern Recognition (CVPR), 2009.*

[21] J. Deng, K. Li, M. Do, H. Su, L. Fei-Fei, **Construction and Analysis of a Large Scale Image Ontology.** *In Vision Sciences Society (VSS), 2009*

[22] See the list of publications featured on the site http://www.iccr-cancer.org

[23] Figure reproduced from Shankar, Shreya, et al. "No classification without representation: Assessing geodiversity issues in open data sets for the developing world." arXiv:1711.08536 (2017).

[24] Figure reproduced from Buolamwini, J. & Gebru, T.. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proc. FAT in PMLR 81:77-91.

When a player wants to develop a deep learning model but does not have enough images, there is a simple and often effective solution, which entails simply using the small image dataset to retrain a model that has already been trained with a larger image dataset (see figure 11). In a 2017 study[25] for example, researchers at Stanford took the Inception-v3 network that had been pre-trained on ImageNet and retrained it using 129,450 photos of skin lesions from 18 image datasets that were divided into in 2,032 classes, exploiting what Inception-v3 had already learnt in order to resolve a complex problem, using a more limited number of images. There are plenty of other examples in scientific literature of "transfer learning" techniques (transferring to one problem what the network has learnt about another problem to arrive at a better solution).

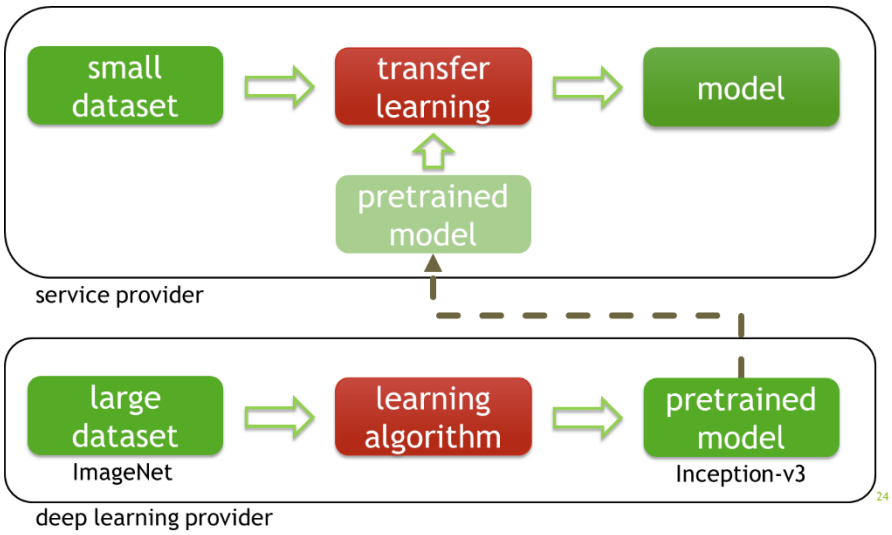


Figure 11: retraining and using a pre-trained neural network by transfer learning.

When transfer learning is used, the biases mentioned above can come from several sources of data: those used to pre-train the model and those used to retrain it. It must be emphasised, however, that transfer learning is often inevitable in deep learning.

2.6 Involvement of players from other disciplines

The scenarios focus on the technical aspects of machine learning but, in addition to artificial intelligence specialists, many other players will also become involved: profiling is by nature multidisciplinary as it requires input from specialists in databases, software engineering, man-machine interfaces, law, ethics etc. The infrastructures to store the data have to be designed, the software systems for profiling have to be painstakingly constructed in a process beginning with the analysis of the needs and culminating in system release, the interfaces for exploiting profiling results (for example for effectively presenting product recommendations to a customer) must be designed, it must be ensured that the profiling system complies with legislation on personal data (among others), checked that the profiling system contains no bias and does not discriminate against any category of individuals etc. All these players must be taken into account in discussion on profiling as their contribution is vital.

[25] https://cs.stanford.edu/people/esteva/nature/

2.7 Conclusion

There are many possible configurations of players in profiling. Within those configurations, each player bears a certain responsibility and has a different impact on the final outcome. The case of open source players is a special one: they have no control over the players who use the data, algorithms and models they provide but they do have a crucial role in innovation where profiling is concerned. It is indispensable, therefore, when considering the legal aspects of profiling, to take account of the specific characteristics of the players and assign adequate responsibility to each of them. In the case of open source tools and datasets, it should be noted that that effort towards documentation has already been undertaken and that, by definition, open source projects are subject to criticism and improvement by their users. Research in the area of artificial intelligence is a source of information on these tools and their limits.

Chapter 3.Typology of profiling, technical solutions and purposes

Numerous types of profiling are possible and can be distinguished by the type of technology used and the purpose of the profiling. A few of these possibilities are reviewed below.

3.1 Types of technical solutions

Profiling uses the data available for a person in order to better understand them or infer other information (risk of developing an illness, consumer behaviour etc). When machine learning must be used to build a model for profiling, the form of profiling must be specified in technical terms in order to choose the right technology.

If we know in advance what purpose profiling is supposed to serve and we have examples of correct responses for a sufficiently large sample of people, supervised machine learning will be used. A **supervised algorithm** is geared to learning the link between the data available for a person and what the profiling entity wants to be able to say about that person. In a psychological study carried out on several volunteers, we might for example look for the link between the data gathered via questionnaires and their level of stress at work. For this kind of profiling, the purpose is clearly established. A supervised machine learning algorithm will be able to use all the questionnaires obtained and see how to best predict stress levels at work based solely on the responses given by each individual.

A number of variants can be distinguished among supervised 'problems', including classification, regression or ordinal regression. **Classification** assigns each individual to one of a predefined set of possible categories. This could relate to illnesses, customer types etc. When the idea is to assign a number to a person, the analysis involves **regression**, for example to predict a person's age on the basis of a photograph, typically by using deep learning, while **ordinal regression** seeks to predict a level of satisfaction or preference.

Systems geared to "recommendations" based on profiling may also use machine learning. This entails predicting an individual's preferences on the basis of other people's preferences, as commonly done on on-line purchasing platforms or in targeted advertising. Collaborative filtering is a recommendation technique which compares the consumer histories of the different users of a service, on the assumption that people with similar habits will be receptive to similar products that they have not yet consumed. Numerous variants exist.

In some profiling scenarios, it is sometimes difficult if not impossible to fully specify the purpose in advance, as the response expected by the profiling system is not yet known. This is typically the case when the aim is to segment the population of a country, the patient community of a hospital or a company's customers. Groups of people will be found automatically by algorithms, the point being precisely to learn new things about the citizens, patients or customers concerned in order to better understand them, develop services better tailored to them, identify sub-populations at risk etc. This **unsupervised profiling** is very common and raises the question as to what extent it is possible or desirable to precisely define the purpose of profiling. Profiling of this kind is generally intended to explore the data and isolate new knowledge that will then be exploited by humans, often in a preliminary phase.

Another form of profiling where the expected behaviour is difficult to predict is **anomaly detection**, which involves detecting "outliers" within a population, ie individuals that are significantly different from the rest of the population. Someone who uses a service in an abnormal way might be detected (case of fraud) or quite simply someone who should be ruled out of other analyses as they would skew the results.

There are also **intermediary situations** in profiling. For example, **semi-supervised** algorithms can be used to carry out supervised profiling, even if the expected response is known only for a limited number of individuals. This means that a large population of individuals can be used to build a classification model, even though the correct category is known only for a low percentage of them. This is a common scenario when the data are easy to acquire but the response (category, number, preference) is costly and difficult to obtain, particularly if it requires intervention by human specialists. This is the case for image processing for example.

In conclusion, there are many different types of algorithms for profiling. Some of those algorithms will themselves help their designers to better understand the people they are studying and more closely define the profiling they are aiming for. It appears important to take account of this variety and the fact that it is particularly difficult to predict what personal data will be useful in the case of unsupervised algorithms. It must also be noted that the aforementioned algorithms are capable of processing not only digitised information but also images, sound, text, sequences etc.

3.2 Types of profiling and their purposes

Part II of this document analyses the purposes of profiling in more detail but it is worth making a few comments here on the basis of the technical discussion above.

The purpose of profiling will be more or less specific depending on whether the profiling is clearly defined to begin with or more exploratory in nature. An additional factor is an important distinction made in machine learning: a **model** (and therefore profiling) **can have a descriptive or predictive aim**. A descriptive model describes in comprehensible form the relation (ie the mathematical formula or logic rules) between someone's personal data and the response sought. Its aim is to better understand this relation, for example when a doctor tries to understand how and why an illness occurs in some

patients and not others. In contrast, a predictive model has the sole aim of predicting the response sought for a specific individual. It is not strictly necessary, therefore, for this model to be transparent, as it must above all be accurate. In practice, some models may be suitable to a certain extent for both description and prediction, such as linear models (see the example of the estate agency) or decision trees (see figure 2). Neural networks are a counterexample: the price to be paid for their predictive power is less transparency. In particular, deep learning makes it possible to process images, sound or text, but without us being able to make sense of the calculations carried out. This problem does not stem from the fact that the calculations are not known (they are known exactly, otherwise the computer would be incapable of running them) but rather from the fact that they are far too complex to be "unravelled" by the human mind in order to arrive at a precise interpretation. There are plenty of tools such as saliency maps that give a vague idea of the area of the image used by deep learning to take a decision but they are insufficient and their reliability is a subject of debate[26].

In the design of a profiling system, one important choice will be the compromise between transparency and accuracy. Most transparent models are less accurate for resolving complex problems. On the other hand, the more accurate models are often difficult to interpret and not very transparent. **The risk to which the individuals profiled are exposed seems to be an important factor in determining where the compromise must lie**. Not understanding the workings of an algorithm that recommends music is probably less of an issue than having no explanation of an algorithm that refuses credit. But in some cases, despite a high risk, it may be difficult to justify the use of a less accurate but transparent profiling system whereas a far more accurate system would be possible. One area where this problem arises is medicine: is it better to have a powerful but opaque model or a less reliable model that we can explain?

Finally, we must stress one specific characteristic of artificial intelligence technologies, machine learning and deep learning that has enabled profiling to come on in leaps and bounds. These algorithms came from scientific research, which is also carried out in certain laboratories owned by large corporations, resulting in a whole host of large open source libraries and datasets that are publicly accessible. The large-scale use of machine learning is made possible by these algorithms and datasets which have been documented and evaluated in hundreds of thousands of scientific publications. The algorithms present in these libraries are often based on the latest scientific developments. In a legal analysis, it is important to take account of this ecosystem in order to preserve it.

[26] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. 2018. Sanity checks for saliency maps. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18), Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, and Nicolò Cesa-Bianchi (Eds.). Curran Associates Inc., USA, 9525-9536.

Section II[27]: Profiling and privacy: from legal considerations to recommendations

Chapter I: CHALLENGES AND DEFINITION OF PROFILING

1. **Introduction**: The term 'profiling' brings together various operations that may pursue a number of purposes and present very different degrees of risk. The profiling operation is not necessarily linked to the use of information systems. Each and every one of us "profiles" other people at some point. We all try to categorise those around us on the basis of their personal attributes whether relevant, objective, permanent or not. In short, we use the subjective or objective data in our possession to categorise others and infer, without doubt with some margin of error in our assessment, other traits or tendencies of which we know nothing. Profiling is therefore in the nature of any human being, as we all try to put a label on the reality around us or, to put it another way, place other people in categories so that we can get a better handle on them and behave accordingly. However, using information systems changes the ways and scope of profiling for various reasons.
   - The first is that present-day information systems, through their interactivity and omnipresence, make it possible to increase - and exponentially so – the amount of data gathered. Whereas data storage and communication capacities were previously limited, today, firstly, they have become pretty well infinite, as demonstrated by the phenomenon of *Big Data* and, secondly, the Internet of things and the multiplication of services that are just a mouse-click away make it possible to capture increasingly trivial aspects of everyday life. Whoever is in possession of those data will be able to 'profile' the individual concerned in ever closer detail.
   - The second is the use of ever more powerful algorithms to analyse that quantity of data. Twenty years ago 'profilers' used the help of algorithms based on a pattern replicating human reasoning, creating what were known as expert systems capable of standing in for (or in all events assisting) the data controller in that they automatically transposed and applied the 'rules' set up by human experts on the basis of their experience. These systems guided reasoning and avoided the subjectivity and risks of discrimination that any human decision-maker would have. These expert systems, which use totally transparent algorithms, are now being superseded by what we call *machine learning* systems or *deep learning* systems capable of working on far more data than could be processed by experts. These systems run correlations between expanding volumes of data using algorithms which feed on the data encountered and gradually refine themselves accordingly. The variety and complexity of the 'models' followed and developed by these

[27] This section was written by Yves POULLET under the supervision and with the assistance of Benoît FRENAY

algorithms are such that their functioning loses some of its transparency, even for the people who developed and/or use them.

2. **Advantages of 'automated' profiling** – So there is a difference between digital 'profiling' and human profiling. Digital profiling presents advantages but also risks for the individual[28]. These risks call for regulation enabling trust on the part of those who are subject to such processing or have its results imposed on them and maximising the benefits linked to the use of these decision-making systems or aids.

   The **advantages** of profiling are quite clear for the **corporations and administrations** which use such systems.

   - For corporations it is all about **optimising** their actions and investment. Profiling will enable corporations to target their clientele, determine their strategies using any number of parameters, better understand the reactions of a given community etc. It is also an aid for making the right choice of location and for employee recruitment or promotion. Finally, it will help them to detect possible cyberattacks, fraud etc.
   - For the public authorities[29], the advantages of these systems lie firstly in having a better grasp of the real situation and then better framing public authority action strategies in areas such as policies on employment, fighting crime or education. Administrations also see them as a means of more effectively applying their regulations.
   - For a **subject** who has been profiled, the advantages are equally well-known. One example is the health sector, where analysis by artificial intelligence of a patient's clinical antecedents and tumour tissue cross-referenced with those of thousands of other patients guides the doctor towards a given course of action in the space of a few seconds and predicts an operation's chances of success. A second example relates to benefits for consumers: someone wishing to buy a lawnmower best suited to their needs and faced with a highly diverse market offer can be gradually guided by an interactive decision-aiding system towards the search engines that will provide answers: this is about optimising consumer choice; a third example is to be found in the service provided by music or film platforms: many of us are very grateful to platform operators for guiding us towards music matching our tastes that we did not know even existed; finally,

[28] We all know the criticisms of the profiling instinctively carried out by a corporate manager who has to recruit an employee or decide on a promotion. Human subjectivity, a bad mood, suspicions of a stitch-up and a lack of quality and quantity of data serving as the basis for their decision will all be sources of doubt hanging over the decision in the eyes of some. Inversely, there is little to criticise in a decision proposed or provided by a machine having worked on plentiful data that appear to be objective (handwriting samples, statistics on a given category of candidates in relation to their studies and curricula vitae, their behaviour during the interview analysed by facial recognition systems etc) and applied without discrimination to all candidates. The neutrality of the workings of the information system, the volume of the data processed and the apparent objectivity are obvious advantages in replacing human assessment with a digital one.

[29] Regarding the use of automated profiling systems in 11 countries of the European Union, see the report: Automating Society Taking Stock of Automated Decision-Making in the EU: A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations, January 2019 available at: www.algorithmwatch.org/automating-society.

in the area of employment policy, public authorities are able to define desirable job profiles and study specialisations, in relation to multiple criteria reflecting the local community, the local economy's current and future needs and numerous other aspects, while pinpointing shortages or surpluses of trainees in the different disciplines. In addition, when made available, the results of this profiling will also be of interest to members of the public who will be able to select their training and orient their job applications in line with demand.

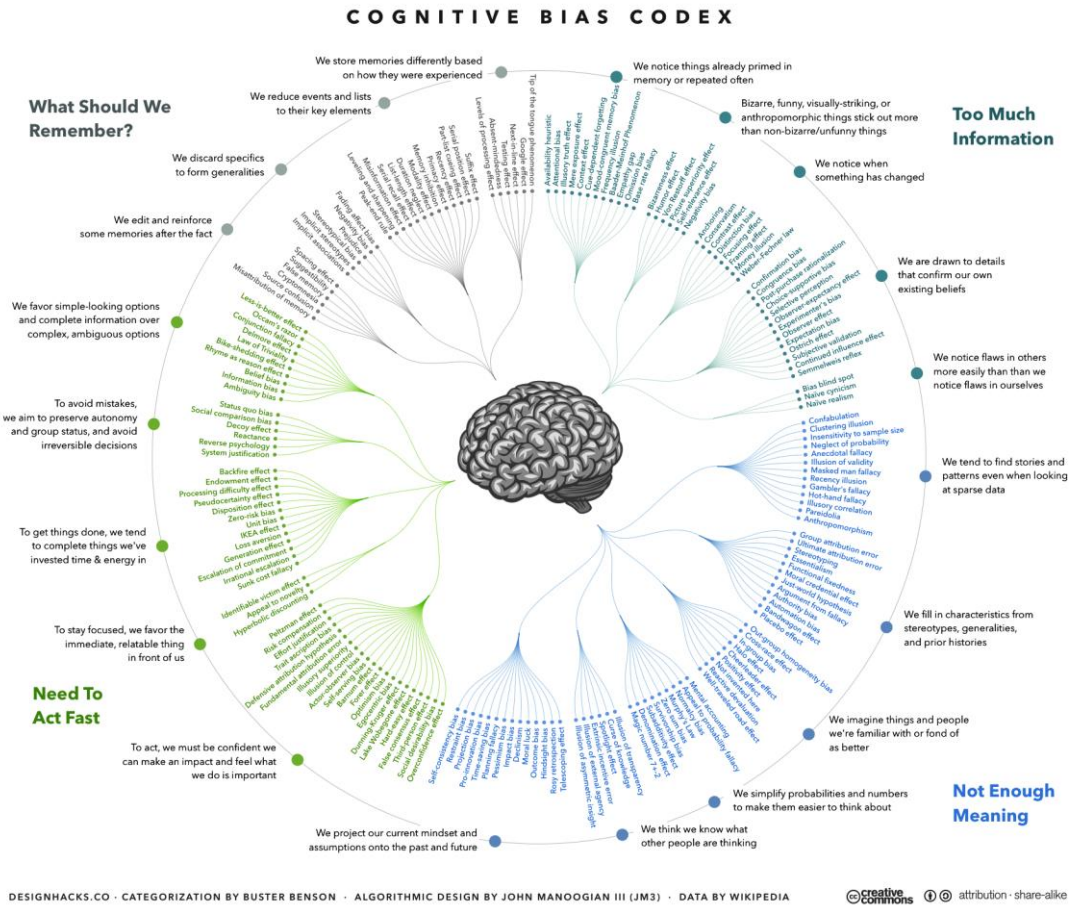
3. **… and risks** - These advantages of automated profiling are to be weighed against **risks** whose seriousness is to be gauged in terms of the consequences and impact of the profiling decisions entrusted to digital systems. The dangers of profiling are to be measured against the purpose sought or discovered by the person setting it up. It must be emphasised that profiling is not just one end in itself but may correspond to a multitude of purposes: medical research, targeting clientele for marketing, framing public strategies, combating fraud or preventing crime etc. That said, because of their characteristics, profiling operations carried out within the framework of *machine learning* information systems generally carry risks inherent in these methods.

    - The first of these is that a large amount of the data gathered is processed out of context.In one example involving the hiring of an employee, a system excluding all graduates who took more than five years to complete their master's degree would rule out a candidate having taken seven years to complete their studies owing to a health problem.
    - The second is that the algorithm may present errors either in its design or in the data used which may be false or of poor quality or a poor match for the problem to be resolved. Worse still, the data or algorithms chosen may contain

a degree of bias[30] yielding misleading results or resulting in discrimination against certain individuals[31] or even groups.

- The third is undeniably the fear of '*Deus ex machina*', ie the trust given *a priori* to the results of the algorithm. Human judgement, though subjective, can be revised and above all contradicted by other human judgements. One could certainly argue that it would be sufficient to reserve the ultimate say for human judgement and, with it, the possibility of revising the 'truth spoken by the machine'. We will come back to this point but we should note that this possibility of review does not always exist and the decision 'suggested' by the computer bears a strong presumption of truth, owing to the qualities of objectivity and neutrality that are acknowledged in or rather attributed to the workings of the information system generating the decision and, therefore, failure to follow the computer's 'suggestion' risks being regarded as an error and, frequently, proof of unacceptable subjectivity on the part of the person not going along with the 'suggestion'. Using such systems, as the GDPR notes in its recitals, incites a real **abdication of responsibility** by the decision-maker. Finally, this manner of taking decisions[32], with no possibility for the individual

30



31 Still on the subject of employee recruitment (see previous note), in this case a computer scientist, taking the gender of top IT specialists into account is a bias: one only has to look at the number of female IT computer scientists to see why there are few people who have the expertise expected

32 As the preparatory work for the RGPD shows, the European legislator concluded that there was cause for concern over such automation as it cuts down the role played by people in decision-making processes: "*This*

to be heard and, sometimes, understand the reasons for the decision taken in their respect, may in some cases be seen as a violation of human dignity, reducing the individual to a mere subject of a calculation.

4. **Initial thoughts on the role of the law** – This brings us to the points targeted by regulatory intervention to correct the risks of digital profiling. Firstly, this entails recognising processing involving profiling as such, defining it. Then it is a matter, when automated profiling systems are designed, of requiring a proper assessment of the risks for those concerned, the risks linked to such processing weighed against the advantages that these systems can offer both for the data subject and for the data controller. In some cases, this weighing up of factors may require a real multidisciplinary debate that is open to consideration of the different interests at stake. Finally, data subjects must have the option, as they used to have, of challenging the 'truths coming from the computer'. Accordingly, such regulation suggests that there should be a distinction between different types of profiling which carry differing risks depending on the purposes sought. We will come back to this after looking at the definitions used in various European texts.

5. **Definitions** - The GDPR defines profiling in Article 4 (4) as follows: "'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 analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements*." Directive (EU) 2016/680 of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data uses the same definition in its Article 3(4).

   This definition does not deviate from the one proposed in Recommendation CM Rec (2010) 13. It should be noted however that the latter recommendation distinguished the notion of profile, resulting from an algorithm capable of serving multiple purposes and being applied to numerous individuals. For example, the fraud suspect 'category' focuses on the abstract characteristics theoretically presented by individuals who may have committed such a crime) while the profiling aspect designates, within an application pursuing a defined purpose, the application of the profile. "'Profile' refers to a set of data characterising a category of individuals that is intended to be applied to an individual". The term 'profiling' in the Council of Europe recommendation hinged on the term 'profile': "*profiling" means an automatic data processing technique that consists of applying a "profile" to an individual, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences,*

provision is designed to protect the interest of the data subject in participating in the making of decisions which are of importance to him. The use of extensive data profiles of individuals by powerful public and private institutions deprives the individual of the capacity to influence decision-making processes within those institutions, should decisions be taken on the sole basis of his 'data shadow'.

behaviours and attitudes". Must the notion of profile be maintained whereas it is dropped by the more recent European texts?

The term 'profile' was and still is meaningful in systems which distinguish operations creating profiles from those that apply them, as would be the case in the definition of an ideal employee profile or a potential criminal. On the other hand, it becomes irrelevant when the functioning of the algorithm no longer allows a distinction between these two phases, as it results directly in an 'action' in relation to an individual (such as the sending of one-to-one marketing material advertising a product or service). In other words, highlighting the 'profile' allows transparency of the criteria to be applied in a second phase by the profiling operation. It also helps to flag up discrimination, which is no longer against individuals but at group level, as it is all the people corresponding to a given category who risk being on the receiving end of a negative evaluation. One could imagine a district presenting a profile of one that is dangerous being systematically stigmatised during operations to track down a criminal or a district merely presenting a profile of widespread illiteracy, where the residents would automatically be assigned a negative coefficient when candidates for jobs were assessed.

In the light of our analysis of the functioning of artificial intelligence systems, we prefer when discussing such systems to use the term 'model' rather than 'profile'. A **model** is a mathematical abstraction providing a simplified description of data to resolve the task at hand, ie the formula with the best parameter values for the solution. A model is not fixed but evolves as it encounters more data. It reflects the fact that profiling, which uses automated learning methods, does not function through causalism which deduces or claims to deduce a rule formulated by experts to establish that a given person belongs to a given category but rather on the basis of purely statistical and evolutive correlation between data. The pattern followed in the case of artificial intelligence is not causal explanation but purely statistical findings.

6. **Generic purposes of profiling: detecting and foreseeing –** But beyond that, how are these definitions to be understood? The definitions distinguish **two generic purposes of profiling: analysis and prevention**. This entails both describing the past in order to understand it and predicting a person's future behaviour. These two purposes are not mutually exclusive. The distinction is clear, for example, when we look at processing carried out by the police: it is one thing for profiling to use past events, such as a crime, to look for a potential culprit (factoring in the place of the crime, the presence of given clues, how the crime was carried out etc) – this is ***reactive*** profiling used to establish the profile of the criminal in terms of a whole host of criteria including the analysis of more or less similar precedents. It is quite another thing, though, to analyse the risks of a prisoner reoffending or anticipating a terrorist attack and potential perpetrators, where a forward-looking ***proactive*** analysis will yield predictions of the behaviour of identified or identifiable individuals. That said, the distinction between these two types of profiling is far from black and white. Obviously, analysing an individual's past may end up pointing to their behaviour in the future.

7. **Profiling beyond personal data** - The definition of profiling in European Union texts envisages processing insofar as it relates to personal data, and the obligations of the

controller relate only to these data. It is noteworthy that the consultative committee of Convention 108 guidelines in respect of Big Data broadens this area of concern. It refers to operations involving data regardless of whether or not they are personal. This is an important point as most of the Big Data on which systems using artificial intelligence for profiling are run bring together both anonymous and personal data. Some are statistical data (for instance, in a database used by the police authorities, statistics on different types of crime by urban sector will be used). Indeed, is the distinction between anonymous and personal still meaningful now that data classified as anonymous can now sometimes be 'deanonymised'? Furthermore, these "anonymous" data are important in most profiling operations, and limiting the obligations of the controller, such as the obligation of information, to personal data alone creates a risk, in our opinion, of having an incomplete view of how profiling-oriented processing works. We will come back to this major concern.

8. **The 2010 recommendation: an avant-garde text** – The Council of Europe's recommendation dates from 2010, just under ten years ago. At the time it was hailed as an avant-garde text. The recommendation explains the importance subsequently attached to profiling-related processing by the European General Data Protection Regulation (GDPR) and it should be noted, in particular, how the points are reiterated in that Regulation, including the controller's obligation to inform the data subject of the 'logic underpinning the processing', to use the terms of the Recommendation or the necessity of a **Risk assessment**.

   So why consider a new recommendation? There are various arguments in favour. **Ambient intelligence (the Internet of things - IoT) and artificial intelligence (AI)**, which are the tools of profiling now and even more so in the future, were little used by the 'profilers' of ten years ago. These two innovations were seen as disruptive as they profoundly modified our environment and our relation to it, creating new risks not only for the individual but also for the functioning of our society as such. The Convetion 108 Committee addressed those risks by drawing up two key documents, one of them being guidelines focusing on the phenomenon of Big Data and the other on AI. It is important, therefore, to take these new risks into account (I) but also the new ideas introduced by the recommendations we have just mentioned (II).

I. The risks posed to our individual freedoms and the other risks linked to profiling in the context of the disruptive innovations of AI and the IoT

9. The notion of 'risk' is central in the Council of Europe Convention and the ensuing Report (see for example and in particular, paragraph 90) but we still have to establish what risks we are talking about. First we will look at those relating to the dangers to our individual freedoms and then go on to analyse other collective risks faced by society and its democratic functioning, risks emphasised by the Convention 108 Committee two recent series of guidelines.

A. **The risks posed to our individual freedoms**

–

10. **Individual freedoms – going beyond data protection** – It should be noted to begin with that the notion of 'individual freedoms' does not stop at the right to data protection but should extend to all individual freedoms that could be jeopardised by profiling. The Council of Europe and the European Union agree on this point: "*As indicated in the Article 29 Data Protection Working Party Statement on the role of a risk-based approach in data protection legal frameworks, the reference to "the rights and freedoms" of data subjects primarily concerns the rights to data protection and privacy but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion*." (Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "*likely to result in a high risk*" for the purposes of Regulation 2016/679 Adopted on 4 April 2017)".

11. **Risks for individual freedoms –** While not an exhaustive list, some of the risks for our individual freedoms are as follows:

- **The risk of reductionism –** It is proving increasingly easy to gather data, as the sources are multiplied by omnipresent technology, storage costs have plunged and transmission capacities facilitate access to the data reservoirs formed or their concentration. All this explains the multiplication of *Big Data* and the possibilities of their use by artificial intelligence systems. Within these data warehouses, individuals are focused on, not as persons but through the aggregation of a certain amount of data concerning them, all regarded as little pieces of 'truth' about each and every one of us in that they represent snapshots of our lives (my presence in a given place, my surfing, the hours I keep and my 'listener objects', my purchase of a given product, my energy consumption) and also through the correlation of data like this with similar data collected from other individuals or anonymous data specific to the entities or groups to which I belong. The truths reflected by data grabbed from life events are further compounded by statistics. It is difficult to deny that someone might be a con-man, the perfect candidate or a person attracted by certain advertising or a political party if their 'profile' or rather the 'model' used shows that 95% of people with the same profile have the same potential or inclinations. In short, in this alchemy of algorithms, the individual becomes an aggregate of data taken out of context and is reduced to being nothing more[33].

  This reality prompts two observations. The first is that the information systems that profile us and, where applicable, decide[34] on a given course of action in our respect perceive us and judge our personalities only through such data and correlations over which there is not always full control. We are recognised through

[33] Specifically regarding the danger of reducing human beings to nothing more than genetic data, read article 2.b of UNESCO's Universal Declaration on the Human Genome and Human Rights (11 November 1997): "*That dignity makes it imperative not to reduce individuals to their genetic characteristics and to respect their uniqueness and diversity*."

[34] Obviously, the system does not decide as such but is the agent to which humans (a corporation, a public body) choose to delegate that power.

"profiles" or "models" shaped by expert systems and 'artificial' intelligence systems and with an eye to the purposes defined by the people who use these data, doubtless without the slightest regard for human dignity. The second concern is that the galloping automation of decision-making processes engenders a pretty well automatic acceptance of the validity and pertinence of those decisions and, in turn, a disengagement and sidestepping of responsibility by "human" decision-makers.

- **The risk of taking data out of context** - Respecting "contexts", ie the areas of trust in which a piece of personal data is transmitted by the data subject, is fundamental in our societies. Use of the same platforms in our various activities (the same social network, for example, the same search engine etc) and the presence of certain platform operators in various activities through their branches, notably the big five (GAFAM - it is noteworthy that these players operate not as competitors but in complementary markets even though they all live off the exploitation of the data they gather through their platforms and the associated services[35], enabling them to cross-reference the data transmitted in different contexts) break down the boundaries which individuals wish or may wish to lay down between the different 'areas of their lives'.

- **The risk of stigmatisation –** Computer memory is or at least seems unlimited, going well beyond the capabilities of human memory. It keeps a trace of our past or of certain events in our past which we would sometimes prefer to forget. This computer memory which is so useful for processing in the area of profiling risks stigmatising an individual for their entire life. An insurer could keep the trace of a customer's illness indefinitely; a bank could keep a note of a bounced cheque; an

35

Economic power of the GAFAM (Source: Wikipedia, v° GAFAM, 2018)

Corporation	Founded	Flagship products	Main source of income (in 2017)[15]	Users[16](billions)	Market capitalisation[17] (billions of USD in March 2018)	Major acquisitions
Google(Alphabet)	1998	Search engine, Advertising, Artificial intelligence	Advertising (86%)	1.42	719	reCAPTCHA, Waze, DoubleClick, YouTube, Android
Apple	1976	PCs	Hardware (81%)	0.85	851[Note 1]	Beats Electronics
Facebook	2005	Réseau social, Advertising, Artificial intelligence	Advertising (98%)	2.13	464	Instagram, WhatsApp, Oculus
Amazon	1994	E-commerce , cloud computing	On-line sales (82%)	0.244	701	Whole Foods Market
Microsoft	1975	Operating system, cloud computing	Software (62%)	1	703	Hotmail, Nokia, Skype, LinkedIn, GitHub

employers' association could keep records of a theft previously committed by a jobseeker or of someone being expelled from or dropping out of school. Is allowing for a person's ability to change and not freezeframing them in their past not an ethical requirement?

- **The risk of the disintegration of the private sphere and the risk of unlimited surveillance** - The fact that technology is everywhere makes individuals transparent as they are increasingly unable to live unconnected to the tools and services of digital society. Their transparency is total, or otherwise partial in the event of them going without some of the services offered by technology or disconnecting. As we have seen, technology records not only the traces voluntarily left[36] by an individual on a social network or sites providing services[37], and not only their movements and motion as well as facial expressions and choices of services, products or information (*surfing*). Using the data captured through AI systems and also *affective computing*, this technology is designed to know or rather guess or predict our emotions and sentiments and, through the examination of our genetic data, our 'identity'. The deletion within our tech-driven society of the distinction between **public sphere and private sphere** is a worrying development: the distinction between the two, which was a mainstay of our right guaranteeing the inviolability of our home, as opposed to the public sphere, has now been done away with. The protection of one's physical abode as inviolable was traditionally viewed, and by the law too, as something that was fundamental for the construction of an individual's personality. The notion of the home, a place where we could be free and out of others' sight, has also been turned upside down by technological developments that are driving the abolition of the distinction between **public and private spheres**.

- **The risk of 'normalisation' linked to the opaqueness of information systems –** In contrast to the transparency of individuals, the information systems engineering that transparency are often not transparent at all. Their opaqueness lies firstly in the functioning of both **terminals** (notably cookies and the RFIDs present in the Internet of things) and **infrastructures** (see the "distributed agents" located throughout information systems such as ambient intelligence systems). This lack of transparency engenders fears of unsolicited and unwanted processing and therefore a desire to conform to what we imagine is the behaviour expected by these new and invisible "places" of surveillance. The dangers and threats stemming from the opaque nature of our information societies where citizens cannot exactly know how information systems function, what data are collected,

[36] Voluntarily but often without being aware of the possibilities opened up for the use of the data confided to the network.

[37] A recent survey by TECHCRUNCH, a media company specialising in the analysis of digital technology, revealed that "*the Israeli company **Glassbox** records everything you do on your telephone when you use a site or application run by one of its clients. This analytics company tries to gain a better understanding of consumers' behaviour and how they navigate in certain applications. Hotels.com, Expedia, Abercombie & Fitch and a great many others use Glassbox to record everything their customers do when using their application, every single keystroke, click and zoom is logged".*

where they are processed or what the intentions are of those processing them, were already highlighted in 1983 by the famous German Constitutional Court judgment in the case of the revised Census Act (*Bundesverfassungsgerichtshof* 15 December 1983, *EuGRZ,* 1983, p.171 ff.). Faced with these opaque systems, citizens are often inclined to adopt the behaviour they imagine society expects and would not dare to express themselves freely, which is harmful for the functioning of our democracies. Furthermore, our life on networks is exposed by the functioning of tools which, in one way or another, formats our knowledge of and approach to the real world and our actions and interactions with others. Our search engines suggest – and we should realise how lucky we are that they do – a ranking of sites in response to our searches and Facebook decides on the priority information we receive.

In short, AI systems help to insidiously set the 'norms' for our behaviour[38] not by imposing them but, in a more subtle manner, by proposing them as the obvious way to make life easier: "*simply click*". These systems operate along the lines of what some call "*libertarian capitalism*". In these systems, the norm is not mandatory but it is suggested that users comply with it; it does not operate transparently but behind the mask of a piece of advice which is presented as meeting your needs[39]. You do not know how that advice is produced other than it is conjured up by the systems which you 'consent' to use … and data, taken from others as well as unknown data that are deemed relevant by the designer and in any case by the system.

- **The risk of manipulation –** The opaque nature of systems' functioning has another consequence: the risk of **manipulation**, and all the more so as artificial intelligence paves the way for what our colleague A. Rouvroy (2014) calls "algorithmic governmentality". As we have already emphasised, the profiles created constitute tools for analysing not only the past but also the 'truth' that these profiles claim to reflect, which, it has to be said, is purely statistical and not exempt from bias. So these profiles are valuable as a means of predicting future behaviour[40].

---

[38] *"They beckon with seductive appeal. Individual citizen-consumers willingly and actively participate in processes of modulation, seeking the benefits that increased personalization can bring. For favoured consumers, these may include price discounts, enhanced products and services, more convenient access to resources, and heightened social status. Within surveillant assemblages, patterns of information flow are accompanied by discourses about why the patterns are natural and beneficial, and those discourses foster widespread internalization of the new norms of information flow. For all of these reasons, a critique of surveillance as privacy invasion "does not do justice to the productive character of consumer surveillance.". Modulation is a mode of privacy invasion, but it is also a mode of knowledge production designed to produce a particular way of knowing and a mode of governance designed to produce a particular kind of subject. Its purpose is to produce tractable, predictable citizen-consumers whose preferred modes of self-determination play out along predictable and profit-generating trajectories. Yet to speak of networked processes of surveillance and modulation in the industrial-era vernacular, as systems for "manufacturing consent" would be too crude*". (J.COHEN, "What Privacy is For?", 126 *Harvard Law Journal*, 2013, (draft of 11 May 2012, p. 12))

[39] 'Recommendation' algorithms used by Facebook for example regarding news items that will interest users.

[40] The boss of Amazon has claimed that "*even before you have placed your order, we have prepared your package*" and the boss of Google has added: "*It will become very difficult for people to see or consume something that has not in some sense been tailored for them.*".

We can see one obvious sign of this manipulation in what are called *nudges*[41]*:* the systems suggest to a driver the best route to take; to a researcher how their H-index might develop; to the head of a municipality the unsafe or no-go areas where policing is required; to a minister of education or a teacher the criteria whereby, in theory, children could succeed at school; to judges the risks of a perpetrator reoffending or the ruling most closely in keeping with the law or rather what has already been ruled as such; to a reader the books that they are bound to enjoy.

To consumers, the computer sends targeted advertising of products or services that are supposed to match your tastes. Is this manipulation reprehensible? Of course not. Salesmen have always practised '*bonus dolus'* without this being thought reprehensible. It might even be seen as a benefit for future 'customers' by giving them more information, enabling them to discover new products or services or even meeting their need for guidance in a market that is ever more complex with an ever wider range of products. Manipulation is reprehensible only if it represents an '**abuse of circumstances'** to borrow the term from the draft Belgian legislation[42] which defines this notion as follows: "*manifest imbalance between the services provided as a result of one party abusing the circumstances linked to the position of weakness of the other party*"[43]. This manipulation is punishable if it constitutes an "***abuse of the weakness of others***", according to the Belgian criminal law of 26 November[44]. The risk of 'abusive' manipulation may be greater when minors, the elderly or disabled persons are involved, but this extension of the law, albeit a vague one, is prompted by the necessity of taking account of the vulnerability of each and every person in our modern society. As for the situations where this risk of manipulation will be applicable, one has the distinct feeling that it will all depend on a '*Risk assessment*' weighing up the interests of each protagonist.

Such manipulation can have far greater ramifications when a profiling system is used for political ends, in order to target voters with the 'right message' which will end up convincing them to vote for a given cause. The 'Cambridge Analytica' scandal shows that this is possible. Here, the risk is not so much individual as collective in that it touches on our notion of democracy.

---

[41] "***Nudge theory*** *(or* ***libertarian paternalism*** *theory*), as Wikipedia explains, is a *concept in behavioural science, political theory and economics drawn from industrial design practices, which argues that indirect suggestions can gently influence the motivation, incentives and decision-making of groups and individuals, at least as effectively if not more than a direct instruction, legislation or implementation*. "

[42] Belgium's current reform of contract law enshrines this concept, in article 5.41 of the draft Code of obligations.

[43] Regarding this provision, its origin and commentaries on it, see the thoughts of H. JACQUEMIN, "Protection du consommateur et numérique en droits européen et belge", in *Vunérabilités et droits dans l'environnement numérique*, Proceedings of the colloquy held in Namur on 14 October 2018, coordinated by H. JACQUEMIN and M. NIHOUL, Larcier, Collection de la faculté de droit de Namur, p. 241 ff; in the same publication, see also F. GEORGE and J.B. HUBIN, "La protection de la personne en droit des obligations", p. 67 ff.

[44] The law of 26 November 2011 introduces, into the Belgian Criminal Code, the notion of abusing another person's situation of weakness. On 7 November 2013, when addressing the complaint brought by some against the vagueness of this legislation, perceived by them as disregarding the principle of 'predictability' of criminal law, the Constitutional Court justified the extension as follows: "*in a democratic society, the protection of people in a situation of weakness constitutes an essential prerequisite for protecting the fundamental rights of everyone*".

- **The risk of dehumanisation –** The ability to reason and take decisions for oneself that men could entrust to machines could result in replacing human reasoning forged by dialogue and consideration of others with an automated mechanism, and this gives rise to major ethical concerns. As can be seen from preparatory work, Europe's law-makers have indeed become worried about this kind of automation and the reduced role played by people and, ultimately, the human being. It has been a question of counter-balancing the risks of human decision-making with everything that comes with it - admittedly subjectivity, over-empathising and poor judgment in the decision process but also the advantages of human decision-making, with its possibilities of correction, dialogue and giving reasons. A further concern is the fact that the galloping automation of decision-making processes is engendering quasi-automatic acceptance of those decisions as valid and pertinent and, in turn, a divestment and abdication of responsibility by "human" decision-makers[45].

- **The risk of discrimination -** Finally, the possibility of automated reasoning being tainted by what is conventionally known as bias has long been argued. This bias, whether deliberate or not, may result in discrimination. The excessive weighting given to one criterion, the fact that such a criterion may conceal another criterion discriminating against a category of the community (black people, women, foreigners, disabled persons, poor people etc) and to a degree that goes far beyond the criteria linked to the categories of sensitive or special data set out in the relevant article of Convention 108+, the refusal to take into consideration a contextual element specific to the data subject, making them a victim of the computer's automaticity, all constitute a final risk both individually, and, where applicable, to an entire group and therefore collectively: an AI system's analysis of districts likely to harbour criminals casts aspersions not only, in individual terms, on the actual people living in that district but also on the district itself and its image and would trigger social consequences (people deserting the district or refusing to go and live there) or possibly increased police surveillance. Therefore, this is very much a risk. The risk of discrimination is exacerbated by the fact that the functioning of the decision-making mechanism appears to be neutral and objective and the opaque nature of that functioning prevents the deciphering of the so-called 'logic' followed.

B. **The "collective" risks and "mutual stakeholder" risks**

12. **Are individual risks the only ones involved?** Many of the risks mentioned above (discrimination, stigmatisation etc) undermine not only the **freedoms of every individual as such but also those of groups**, be they ethnic, philosophical, low-income, district residents etc. Processing of this kind affects other values, particularly **social**

[45] In this connection, the European Commission's AI experts pointed out that *"the results produced by the machine, using more and more sophisticated software, and even expert system, has an apparently objective and incontrovertible character to which a human decision-maker may attach too much weight, thus abdicating his own responsibilities*".

justice or cultural diversity among individuals or among groups[46] and, beyond that and sometimes substantially, the functioning of our societies and our democracies in particular. The abusive manipulation of individuals undermines both freedoms and human dignity in the Kantian sense of the term[47] but while it is a danger to each and every one of us in this respect, it may also affect an entire population or have blanket effects by shaping the political views of citizens for example. Personalising offers of services and excluding certain individuals from the potential benefits of those services has ramifications beyond individuals and impacts groups of people, raising questions of social justice.

Another thought: by definition, Big Data drag us all into a shared adventure. When my surfing data are gathered by my favourite search engine and swallowed up by a vast database together with millions of data items relating to the choices of other web users, the model that will at some point take a decision regarding me, will clearly do so in relation to all the data gathered and differentiation induced by algorithmic searches between my data and those of others. There is also a domino effect whereby a person's choice to log on to a social network for example will ultimately have a major influence on the choices of their friends and family.

Let us take the example of '*one-to-one*' insurances for health care or civil liability: personalising premiums to obtain the closest possible match to the 'risks' that each individual may present, risks calculated on the basis of their profile, puts the sacrosanct principle of risk-sharing, a mainstay of our insurance system, under huge strain. It is quite remarkable that traditional data protection or privacy issues, in the narrow sense of the term, are transcended in this way. We could say the same thing of an artificial intelligence system intended to predict the chances of children doing well at school or the risks of domestic violence against children within the population that would give weightings to certain data. The risk posed by this identification is not just individual but collective too, as there is a danger of stigmatising certain types of communities. The necessity of broadening the concerns of Convention 108 is rightly emphasised in the recent guidelines we have already cited, on Big Data and artificial intelligence. Moreover, while this broadening of the spectrum of risks to be taken into account is not to be found in the European Union's texts, it is present in the recommendations of the OECD adopted a short time ago by the Ministerial Council in May 2019 : "*AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights*". This point is of major importance and raises the question of whether it is possible to combat these risks to collective groups and mutual stakeholders via the data protection texts and the bodies created by such legislation. **Are the scope of those texts and the competence of those bodies not limited to solely protecting against individual risks?** We will come back

[46] Cf. UNESCO's 2003 Declaration on bioethics: "*No individual or group should be discriminated against or stigmatized on any grounds, in violation of human dignity, human rights and fundamental freedoms*."

[47] According to Kantian doctrine which is widely accepted in our European countries, dignity means that human beings must never be regarded as a means but always as an end.

to this important point, which must be taken into account by our recommendations[48].

II. The texts of the European Union and the Council of Europe

13. **Introduction –** We will limit ourselves here to listing the provisions of the European texts and adding a brief comment on how the Council's recent instruments (guidelines adopted by the Convention 108 Committee) point out the necessity of broadening the debate, which is particularly called for where profiling-related processing is concerned.

14. **The GDPR and Directive 206/680 –** In the **texts of the European Union**, the question of profiling is broached by various provisions. We have already cited the definition given by Article 4 (4) of the GDPR or Article 3 (4) of Directive 2016/680 on processing used in police work. In addition, Articles 13. 2 (f) and 14. 2 (g) of the GDPR[49] mention among the information to be provided to the data subject: '*the existence of automated decision-making, including profiling … and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject*'. Article 15.1 (h) stipulates access to the same information. Article 22. 1, under the heading: '*Automated individual decision-making, including profiling',* entitles the data subject to refuse to be subjected to this type of processing if it produces legal effects concerning them or similarly 'significantly affects' them except in the cases provided for in paragraph 2, namely if it is necessary for entering into a contract or if it is authorised by law or by explicit consent. Where such exceptions apply, Article 22.3 obliges the controller to '*implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision'.* Article 35 concerns the controller's obligation to carry out an 'assessment of the impact' of the envisaged processing operations where it is likely to result in a 'high risk' for the data subjects. Paragraph 3 of this article makes this analysis a particular requirement '*in the case of a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling',* when decisions based on that evaluation produce legal effects on the data subjects or significantly affect them. Furthermore, on 3 October 2017, with a view to the then forthcoming application of the GDPR, the so-called Article 29 Working Party issued '*Guidelines on automated decision making and Profiling'*, intended to interpret the Regulation's various provisions. These very substantial '*Guidelines*' were taken on board by the *European Data Protection Board (EDPB)* set up since then under the Regulation, which has inherited the Article 29 Working Party's prerogatives, among

[48] As far as we know, only the recent report by S. DREYER and W. SCHULTZ ("The GDPR and automated decision making : Will it deliver?", *Bertelsman Foundation Report,* January 2019*,* available from the Bertelsmann foundation site*)* looks at this question.

[49] We will not go into the provisions of the text of the directive on processing used in police work in detail here. In the main, they follow those of the Regulation. We will come back to this text when discussing profiling carried out by police authorities or criminal courts (i*nfra*, no 24)

others. We will be looking at a few points of interpretation of the GDPR text featuring in those 'Guidelines'.

Moreover, Directive 2016/680, in this case applicable to the processing of data in the area of criminal offences or penalties, on top of reiterating in Article 11 the principle of the inadequacy of automated individual decision-making enshrined in Article 22, adds in paragraph 3 that " *Profiling that results in discrimination against natural persons on the basis of special categories of personal data … shall be prohibited, …*"

15. **And Convention 108+** ?- The text of Convention 108+ neither mentions nor foresees processing such as profiling. However, we should emphasise Article 9.1 regarding automated decisions (to be compared with Article 22 of the GDPR) : "*Every individual shall have a right: a. not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration".* The importance of an approach geared to risks is emphasised by Article 10, where paragraph 2 stipulates: *"Each Party shall provide that controllers and, where applicable, processors, examine the likely impact of intended data processing on the rights and fundamental freedoms of data subjects prior to the commencement of such processing, and shall design the data processing in such a manner as to* ***prevent or minimise the risk*** *of interference with those rights and fundamental freedoms."* and paragraph 4 allows each State to adopt complementary measures (see also Article 13): *"Each Party may, having regard to the risks arising for the interests, rights and fundamental freedoms of the data subjects, adapt the application of the provisions of paragraphs 1, 2 and 3 in the law giving effect to the provisions of this Convention, according to the nature and volume of the data, the nature, scope and purpose of the processing and, where appropriate, the size of the controller or processor*".

16. **Broadening of the scope by the Guidelines on Big Data and AI** What seems more significant to us are the broader dimensions asserted by the Council guidelines which cover various aspects and must be taken into account when writing any new instrument on profiling, and all the more so in the knowledge that profiling is now increasingly reliant on the resources provided by Big Data and the potential offered by artificial intelligence:

    a. These first of these has already been mentioned: the two guidelines very much set the tone for the debate by taking account of other ethical aspects such as dignity, social justice and non-discrimination, cultural diversity and the joint stakeholder imperatives of democracy and not just the concerns surrounding

individual data protection[50]. The "*Guidelines on Big Data*" of 23 January 2017[51] do not hesitate to spell this out: "*Since the use of Big Data may affect not only individual privacy and data protection, but also the collective dimension of these rights, preventive policies and risk-assessment shall consider the legal, social and ethical impact of the use of Big Data, including with regard to the right to equal treatment and to non-discrimination*". And the general guidance highlighted in the Guidelines on AI follow the same direction: *"The protection of human dignity and safeguarding of human rights and fundamental freedoms, in particular the right to the protection of personal data, are essential when developing and adopting AI applications that may have consequences on individuals and society. This is especially important when AI applications are used in decision-making processes.".*

b. The second involves broadening the circle of players whose role entails certain obligations. While the GDPR and Convention 108+ mention only the obligations of the "controller" of the processing, the two guidelines incorporate the responsibility of other players involved either in the supply of the Big Data or libraries to be used by profiling systems or in the design and implementation of the base algorithms or the tailoring of those algorithms to the particular needs of a given sector or system[52]. This concern comes as no surprise as, when we look at the players' involvement, it is clear that the characteristics of the processing and the extent of the associated risks are far from depending merely on the controller and are in fact the result of the choice of the data often acquired outside (data supplier) and the choice of base algorithm or algorithm

[50] See, along the same lines: "*AI systems should not harm human beings. By design, AI systems should protect the dignity, integrity, liberty, privacy, safety, and security of human beings in society and at work. AI systems should not threaten the democratic process, freedom of expression, freedoms of identify, or the possibility to refuse AI services. At the very least, AI systems should not be designed in a way that enhances existing harms or creates new harms for individuals. Harms can be physical, psychological, financial or social. AI specific harms may stem from the treatment of data on individuals (i.e. how it is collected, stored, used, etc.). To avoid harm, data collected and used for training of AI algorithms must be done in a way that avoids discrimination, manipulation, or negative profiling. Of equal importance, AI systems should be developed and implemented in a way that protects societies from ideological polarization and algorithmic determinism*." High Level Expert Group on Artificial Intelligence in its Ethical Guidelines for a Trustworthy AI (2019).

[51] See also "*Personal data processing should not be in conflict with the ethical values commonly accepted in the relevant community or communities and should not prejudice societal interests, values and norms, including the protection of human rights*. …"

[52] Guidelines on AI: *"1. AI developers, manufacturers and service providers should adopt a values-oriented approach in the design of their products and services, consistent with Convention 108+, in particular with article 10.2, and other relevant instruments of the Council of Europe.*

2. AI developers, manufacturers and service providers should assess the possible adverse consequences of AI applications on human rights and fundamental freedoms, and, considering these consequences, adopt a precautionary approach based on appropriate risk prevention and mitigation measures.

3. In all phases of the processing, including data collection, AI developers, manufacturers and service providers should adopt a human rights by-design approach and avoid any potential biases, including unintentional or hidden, and the risk of discrimination or other adverse impacts on the human rights and fundamental freedoms of data subjects.

4. AI developers should critically assess the quality, nature, origin and amount of personal data used, reducing unnecessary, redundant or marginal data during the development, and training phases and then monitoring the model's accuracy as it is fed with new data. The use of synthetic data may be considered as one possible solution to minimise the amount of personal data processed by AI applications."

tailored by a third party to the needs of the controller's specific application (developers and manufacturers, to use the terms employed by the Guidelines on AI)[53]. Finally, can it not be stipulated, as advocated by the 2010 Recommendation (Section 3.8.) that: "*The distribution and use, without the data subject's knowledge, of software aimed at the observation or the monitoring in the context of profiling of the use being made of a given terminal or electronic communication network should be permitted only if they are expressly provided for by domestic law and accompanied by appropriate safeguards*"?

c. The third point to look at is greater emphasis on the risk-oriented approach. Both guidelines call for prior and full identification of the risks to both individuals and society and for consideration of the different types of algorithms used: *"The risk of adverse impacts on individuals and society due to de-contextualised data and de-contextualised algorithmic models should be adequately considered in developing and using AI applications"*, according to the Guidelines on AI). The Guidelines on Big Data are along similar lines, calling for preventive policies and various risk assessment measures. This approach ties in with the application of the precautionary principle, commonly used in environmental law.

d. The fourth extends the environmental law approach, calling for a participatory, constructive, '*multi-stakeholder*' and interdisciplinary assessment of the risks and solutions to be found. It advocates the creation, where applicable and depending on the extent of the risks envisaged, of an ethics committee (Guidelines on Big Data) *: "If the assessment of the likely impact of an intended data processing described in Section IV.2 highlights a high impact of the use of Big Data on ethical values, controllers could establish an ad hoc ethics committee, or rely on existing ones, to identify the specific ethical values to be safeguarded in the use of data. The ethics committee should be an independent body composed by members selected for their competence, experience and professional qualities and performing their duties impartially and objectively."* This point is echoed in the Guidelines on AI: "*AI developers, manufacturers and service providers are encouraged to set up and consult independent committees of experts from a range of fields, as well as engage with independent academic institutions, which can contribute to designing human rights-based and ethically and socially-oriented AI applications, and to detecting potential bias. Such committees may play an especially important role in areas where*

[53] The same line of thought was taken by the Article 29 Group which, in the context of its '*Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is " likely to result in a high risk " for the purposes of Regulation 2016/679,*" (adopted on 4 April 2017), writes: " *A DPIA can also be useful for assessing the data protection impact of a technology product, for example a piece of hardware or software, where this is likely to be used by different data controllers to carry out different processing operations. Of course, the data controller deploying the product remains obliged to carry out its own DPIA with regard to the specific implementation, but this can be informed by a DPIA prepared by the product provider, if appropriate. An example could be the relationship between manufacturers of smart meters and utility companies. Each product provider or processor should share useful information without neither compromising secrets nor leading to security risks by disclosing vulnerabilities."*

transparency and stakeholder engagement can be more difficult due to competing interests and rights, such as in the fields of predictive justice, crime prevention and detection." and "*Participatory forms of risk assessment, based on the active engagement of the individuals and groups potentially affected by AI applications, should be encouraged.*"

Chapter II: Classification of profiling

17. **What criteria to use?** – The "risk-based" approach highlighted by Council of Europe Convention 108+ and the two Guidelines suggests it is important to make the case for differentiated regulation, based on the degree of risk associated with the different types of processing. This degree of risk is in any case determined by reference firstly to the purpose of the profiling, secondly to the consequences it may have on the data subject or groups concerned and lastly, as has already been pointed out, to the processing method used. So, for example, profiling where the profiler uses only data that have been voluntarily provided by the consumer according to the precise needs of the service they desire will warrant a less tough approach than profiling where platforms make multi-purpose use of an ever-growing volume of data gleaned from the use of their services.

    Profiling is by no means a homogeneous category of processing activity, therefore. In this chapter, we will show that there are numerous sub-categories which present risks of varying kinds and significance. Examining the purposes for which profiling is used in practice provides an indication of these subcategories (I). The second criterion distinguishes processing operations according to the degree of risk involved: we talk about "high risk profiling" activities which we distinguish from other types of profiling in terms of the need for regulation (II). To conclude, we consider the possible implications of this dual classification (III).

I. Classification by purpose

18. **Profiling – what's it for?** – Below we list a series of purposes according to the context of the profiling and look at how data protection instruments classify the different purposes of processing. For example, profiling might take place in a pre-contractual phase of the relationship between the controller and the data subject, or it might occur in the course of their contractual relationship. Profiling can be a service in itself which, whether fed with data from external sources or not, enables profiles to be supplied to third parties. Research, especially (although not exclusively) medical research, is another area where profiling is apt to be employed, or necessary. In the context of processing performed by public authorities, a further distinction will be made between various types of profiling depending on whether they are an aid to general decision-making or whether they are performed for the purpose of applying rules and regulations in individual cases. Specific attention will be paid to profiling carried out by the police or judicial authorities for the purpose of crime prevention or detection. The issue of predictive justice will also be addressed. We could, of course, have classified

profiling in other ways, such as by sector: direct marketing, administration, employment, banking, insurance, policing and justice.

A. **Profiling and the pre-contractual phase of the controller-data subject relationship.**

19. **From targeted advertising to exclusion and adaptive pricing -** Targeted advertising is probably the first application that springs to mind when it comes to profiling. In place of broadcasting messages to audiences in a fairly indiscriminate and relatively neutral fashion (obviously readers of high-end glossy magazines will not be sent the same ads as readers of tabloid newspapers), one-to-one advertising seeks to match the personality and needs of the recipient as closely as possible in order to reach consumers more efficiently. In many cases, the recipient of the message will find such ads useful and may even actively solicit them. Even when it is legitimate, however, the manipulation that is intrinsic to all advertising is exponentially greater here as the message may be communicated in a covert (in the form of information) or misleading way (the recipient is unaware that the message comes from a third party and not from the website they are visiting). The manipulation will be all the more potent, furthermore, in that your profile will "pinpoint" your innermost self, playing on your emotions, inclinations (including sexual ones), race, even your disabilities or opinions[54]. The danger here is particularly acute because in complex AI systems (such as deep learning), it is difficult to make the models transparent and the risk of manipulation is hard to detect, prompting talk of "black boxes".

    Customer targeting can nevertheless have complementary objectives: adaptive pricing combines an estimate of the intensity of demand for a given product with the "profile" and offers different Internet users different prices based on this variable. Another purpose is to select potential contracting parties, by preventing some people from viewing certain messages or eliminating others by making offers that are ludicrous. In one high-profile case in the United States, Facebook was sued when a lettings agency used Facebook's profiling services to screen out enquiries from prospective tenants or buyers belonging to certain groups (African-Americans, people on low incomes, LGBT, etc). In the field of employment, too, it is not uncommon for candidates to be selected on the basis of data that have been analysed by machine, according to criteria which will have been articulated to a greater or lesser degree when using the profiling algorithms.

B. Profiling in the context of a contractual relationship

20. **The needs of the contract** - The purpose of the profiling in this case will be to evaluate the contracting person, but this evaluation will itself have different objectives: to

[54] See in particular the article by Kosinsky et al., "Private traits and attributes are predictable from digital records of human behavior" (2013) which looks at what Facebook can deduce from Likes: "*We show that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.*"

evaluate the contractual performance of an employee, for example, in connection with a promotion or dismissal, or, more broadly, of a customer in the context of a bank or insurance company trying to decide how much to lend or how much to charge in premiums, or to determine customer profitability. Alternatively, there may be financial or other risks to be assessed when evaluating a partner, whether a business or an individual, in order to decide whether it is worth continuing the relationship.

C. **Profiling and activities relating to the "marketing" of profiles**.

21. **The marketing of data or profiles** – Holders of Big Data may be tempted to offer profiling services to third parties either as simple providers of data or in response to requests from the "customer" to develop the necessary algorithms and apply them to their own data; the Big Data holders would then make the persons profiled available, or offer to perform themselves whatever operation the third party requires. The existence of datasets, whether open source or proprietary, and the provision of algorithms free of charge by research laboratories or on a commercial basis by companies offering profiling services were discussed in the first chapter. The "Facebook Fan Page" affair and the Cambridge Analytica scandal are worth citing here. In the former, Facebook provided the administrators of a fan page with information on the profiles of people visiting the site so that they could learn more about the visitors who liked their fan page and design a better strategy. The Cambridge Analytica (CA) case concerned the use of data generated by the social network and supplemented by CA's co-called research surveys. In the Facebook-Cambridge Analytica scandal, 87 million Facebook users were affected when Cambridge Analytica began harvesting their personal data in 2014. The information was used to influence voting intentions in favour of politicians who paid for CA's services. The aim was to ascertain people's political leanings and even predict how they would vote and to market the findings to political operators, so that they could tailor their political marketing strategies. Just recently, the Financial Times (4 September 2019) revealed how Google had been sharing data with commercial companies precisely for the purpose of facilitating their profiling activities[55].

D. Profiling and research activities

22. **Profiling as an aid to research** – In the context of research projects, researchers can use data typically gathered by public or private authorities for primary purposes to develop profiling in order to better understand a particular phenomenon and so improve their knowledge and understanding. Researchers interested in personnel management, for instance, will use corporate data, interviews with relevant actors and data held by the employment or finance ministry to try to understand how certain factors may account for professional success or the lack thereof. In the medical field, artificial intelligence is widely used to construct profiles, including genetic profiles, of

[55] The Financial Times (4/09) revealed how **Google had been undermining its own data privacy policies and circumventing EU regulations which require consent and transparency.** The firm had been labelling Internet users with an identifying tracker. It then invited its advertising clients to log on to a hidden web page containing a unique address that linked it to the users' browsing activity, making it possible for advertisers to send targeted ads.

people suffering from a particular disease and so predict how the illness is likely to evolve and, if necessary, take preventive action. Vehicle manufacturers can likewise benefit from research using data on motorists' use of vehicles and the context in which they are driven.

Scientific research has been the subject of various special arrangements because of the benefits that it brings. As we know, the GDPR takes the view that, subject to appropriate safeguards (accreditation of laboratories, use of pseudonyms, etc), the research purpose is not incompatible with the primary purposes of data processing. More recently, notwithstanding the Database Directive, the Copyright in a Digital Society Directive exempts public-interest scientific institutions from any restrictions on the use of databases through data mining. In other words, the European instruments, while making profiling in the context of research subject to certain conditions, nevertheless take a positive view of such activities.

E. Profiling by public authorities

23. **Profiling for governments and administrations** - Profiling allows public authorities and administrations to pursue a variety of objectives. First and foremost, it is useful in devising **strategies**, whether economic expansion policies, or policies on subsidised housing, mobility, education or employment assistance56. To this end, the authorities consider myriad factors and combine reams of data from public and private sources to produce "predictive models" for determining what impact a particular policy is likely to have. It is easy to see how, if not programmed correctly (bias, poor data quality or errors in the algorithms), activities of this kind could affect certain groups or, at any rate, how decisions based on such forecasts could affect the members of these groups.

    One area where AI and profiling systems can be a ready source of efficiency gains is when it comes to implementing rules and regulations. As part of a philosophy of "benevolent government", where the state plays a proactive role with respect to its citizens, such systems can be used not only to spot or even select people who could benefit from special assistance, or to ensure students receive the best possible advice about education pathways, etc, but also to detect problem families (child abuse) or even social security fraudsters or tax evaders. Predictive justice, which is intended to replace judges, is also worth mentioning in this context insofar as any dispute can be "profiled" according to the many and varied characteristics of the case, analysed in the light of previous decisions.

[56] For information about various examples of decision-making systems in the public sector and in support of governmental strategies, see the report "Automating Society - Taking Stock of Automated Decision-Making in the EU": A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations, January 2019: www.algorithmwatch.org/automating-society.

F. Profiling and criminal investigation and prosecution bodies

24. **AI as an aid to policing and prosecution** - The European Union Agency for Fundamental Rights in its 201757 report identifies two kinds of profiling in this area: "*There are two main purposes of profiling in the context of law enforcement and border management: to **identify known individuals based on intelligence concerning a specific individual**, and as a **predictive method** to identify "unknown' individuals who may be of interest to law enforcement and border management authorities. Both may include conscious or unconscious biases that may discriminate against individuals*." The idea, then, is to act either in response to an offence that has already been committed or proactively so as to prevent possible crimes. In the latter case, an overarching strategy may be devised that will lead to decisions relating either to entire groups (for example, monitoring a particular community) or to individuals.

    Either way, the Agency underlines the risk of discrimination associated with profiling of this kind and points out that according to the European directive: "*Profiling falling under the scope of the Police Directive should abide by Article 11 (3) of the Police Directive. It provides that"[p]rofiling that results in discrimination against natural persons on the basis of special categories of personal data referred to in Article 10* shall be prohibited, in accordance with Union law*". There remains the question of criteria based on data other than the "sensitive" kind. For example, level of education or mobility could also be a consideration when seeking to identify criminogenic settings. One suspects that it is not the nature of the data but the purpose of the processing that creates the risk.

II. Classification by risk

25. **Risks as an important criterion to be considered in the recommendation** – It is our belief that not all profiling should be regulated the same way and that, instead, profiling needs to be understood according to the nature of the risks engendered and the implications of the processing for the data subjects or society. This concern is reflected in Convention 108+, both in Article 10.2 and in the Explanatory Report (paragraph 88): "*Paragraph 2 clarifies that before carrying out a data processing activity, the controller will have to examine its potential impact on the rights and fundamental freedoms of the data subjects*." It is also evident in Article 10.4, which allows the parties to include obligations in addition to those laid down in the other provisions "*taking into consideration the risks at stake*". These risks, moreover, may be assessed at a micro level, ie the data subject, or at a macro level, ie existing social groups, whether organised or not, or at both levels. Lastly, insofar as the profiling involves different actors, each of whose actions has the potential to cause harm, it is important to make it clear how each one is to bear their share of the risk.

    In our discussion, we start from the notion of high risk processing developed by the European Union texts and suggested by the Council of Europe, especially in its

[57] FRA, European Union Agency for Fundamental Rights, *Preventing unlawful profiling today and in the future: a guide*, Publications Office, 2019.

Guidelines on Big Data and AI (A). Attention will then turn to the legal arrangements associated with this concept (B).

A. The concept of "high risk" in European texts

a. The GDPR

26. The requirement to carry out risk assessment - Article 35.1 of the GDPR states: "*Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data*." Article 27.1. of the "Police Directive" employs the same wording. The concept of "high risk" is central to European thinking, therefore. The same article of the GDPR sets out three hypothetical situations which could be considered examples of "high risk" processing. These three scenarios are particularly relevant where profiling is concerned.

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;
b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or
c) a systematic monitoring of a publicly accessible area on a large scale.

Point (a) refers directly to profiling and therefore needs to be taken into account; the second point concerns the creation of Big Data involving sensitive information. When profiling is combined with Big Data of this type, risk assessment becomes even more necessary. The third and last scenario concerns the collection of data for the purpose of monitoring an area accessible to the public (streets, cinemas, shops, etc). There is little doubt that such monitoring, if it is intended to be systematic, will use profiling techniques that rely, for example, on facial recognition, movement analysis, the Internet of Things, etc.

27. **Processing operations deemed "high risk" under the GDPR –** A close look at EU legislation shows that profiling is considered "high risk" in the following instances:

- The evaluation of a natural person through profiling is presumed to be "high risk" if it is carried out in a systematic and extensive fashion. Presumably this is to exclude one-off evaluations, but also evaluations that would lead to simplistic assessments about personality. The interpretation of this second point is ambiguous. Our feeling is that the critical determinant here is not the more or less

sensitive nature of the data (the data may be trivial) but rather the outcome (processing these trivial data may lead to detecting a person's nervousness in times of stress). The processing activity must also have **legal effects on the data subject or significantly affect them.** This echoes the ideas expressed in Article 22 of the GDPR, on automated processing. The Guidelines define both of these notions at length. Examples of a legal effect of a decision generated or suggested by processing would include cancellation of a contract, or the award or denial of a social welfare benefit, or, add the Guidelines, admission (or refused admission) to a country or denial of citizenship. "*Significantly affect"* implies that the damage persists over time, that it leads to exclusion or discrimination, and that it concerns a financial, health or education service or employment. The Guidelines do not exclude certain types of profiling for advertising purposes, depending on the extent of the data collected, the reasonable expectations of the data subject, the techniques used to target the advertising and, above all, how knowledge of the data subject's vulnerabilities is used. The document goes on to state that practices such as dynamic pricing should also in some cases be treated as "high risk" processing operations.

- "Large-scale" processing of the sensitive data referred to in Article 9 of the GDPR, ie "*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*" as well as data relating to criminal convictions is considered "high risk". Here, the main reason for this designation is the risk of discrimination. The latter may be directed against individuals, of course, but it may equally concern entire ethnic groups and this possibility also needs to be considered, notwithstanding the text of the GDPR. We would also point out, and this is another criticism of the EU text, that what should determine whether an operation qualifies as "high risk" processing of this kind is not necessarily the sensitive content of the data, as stated in Article 35, but rather the purpose of the operation, insofar as the revelation of the sensitive character may come about as a result of the processing (see Article 9: "*Processing of personal data revealing…"*). In the Cambridge Analytica case, for example, it was shown that political opinions could be deduced from non-sensitive data linked simply to social media use. The term "large-scale" suggests that only processing involving a large population, based on the geographical context or the population liable to be affected, is covered.

- The third hypothetical situation concerns processing in *publicly accessible areas*[58] and seems to focus only on systematic surveillance processing carried out, say, for security purposes (eg for a department store, to guard against theft or other offences). The reason given is that it is impossible for anyone present in a public place (street, department store, hotel, government offices, etc) to escape such surveillance. We would add that data may be gathered in public areas for purposes

[58] The use of the term "areas" seems to indicate that only physical and non-virtual spaces are covered. Consideration could be given to widening the scope to include communication or news platforms whose services are largely open to the public.

other than "monitoring" alone. For example, a department store may wish to suggest other items for purchase and, where appropriate, share the data collected with companies operating in a different sector such as tourism for profiling purposes. Note that the text does not preclude a broad interpretation of the term “monitoring” and could also, therefore, cover commercial purposes of this type.

b) Council of Europe instruments

28. **What about the Council of Europe?** - The explanatory report to Convention 108+, when commenting on the possibility for member states to provide for additional obligations in the event of higher risks, provides a few criteria: "*Such adaptation should be done considering the nature and volume of data processed, the nature, scope and purposes of the data processing and, in certain cases, the size of the processing entity.*" We would add that the capacity of the persons targeted could also be an indicator since, as has already been pointed out, certain categories such as children and possibly other groups as well, such as patients or employees, are easier to manipulate. It is clear that processing activity which met several of these criteria would undoubtedly qualify as “high risk”. The Guidelines on Big Data suggest a further criterion for "high risk": "*Exposing data subjects to different risks or greater risks than those contemplated by the initial purposes could be considered as a case of further processing of data in an unexpected manner*." This is an important point because, increasingly, our data are blended with others to form pools into which the artificial intelligence systems that make profiling possible can dip in order to pursue a range of goals. The transfer to third parties of profiles or data leading to profiling and/or data sharing should also, in our view, be considered “high risk”.

29. The notion of “special risks” had already been mentioned in the Council’s 2010 recommendation and that this is undoubtedly what inspired the GDPR when it talks about “high risk” processing. Section 9.2 of the recommendation reads as follows: “*Furthermore, in cases of processing that use profiling and entail special risks with regard to the protection of privacy and personal data, member states may foresee either: a. that controllers have to notify the supervisory authority in advance of the processing; or b. that this processing is subject to prior checking by the supervisory authority*.” It is a pity that the concept was not better defined in the recommendation, but the above at least shows the importance attached by the Council of Europe, from as early as 2010, to developing special rules requiring the supervisory authority to monitor or even approve profiling operations that are more than usually hazardous.

Another point made in the Guidelines on Big Data and Artificial Intelligence has been highlighted several times already. Paragraph 2.3 of the Guidelines on Big Data reads: *“Since the use of Big Data may affect not only individual privacy and data protection, but also the collective dimension of these rights, preventive policies and risk-assessment shall consider the legal, social and ethical impact of the use of Big Data, including with regard to the right to equal treatment and to non-discrimination.*” Our position, in line with both Convention 108 Committee Guidelines, is that in order to determine the severity of the risks involved in profiling, consideration needs to be

given to the full range of societal and collective risks, such as the risk of discrimination, or risks relating to social justice, the functioning of our democracies, etc and not just to the risks facing individuals.

B. How best to regulate profiling, in particular the high risk kind?

30. **Towards multiple regimes?** – It is probably hard to devise a single legal regime that would cover all profiling activities. The proportionality rule, ie the need for regulations commensurate with the risks created and for an appropriate response to them, is worth noting here. Regulation that would be excessively costly to implement and miss its mark is to be avoided. We will have an opportunity to demonstrate this is point III, when we review the different types of profiling. For now, though, we will turn our attention to the various ideas expressed not only in the European Union and Council of Europe texts already mentioned but also in other documents, in particular those relating to ethics and artificial intelligence.

Certainly, the first rule, for anyone introducing profiling or, simply, a software tool (profiling algorithm) or database intended for one or more profiling operations, is to identify the risks associated with the planned operations or to which those operations contribute or might contribute. The risk is twofold. While the European texts cited focus on the possible effects of profiling - risk of discrimination, manipulation, forced surveillance or even societal risks associated with a particular processing operation - it is also important to consider, as for any processing operation but bearing in mind the particular features of the system, the risks specific to the system's security: for example, the risks relating to confidentiality, lack of integrity of the raw or processed data or even the risk that the system might be unavailable. The recently adopted OECD recommendations and the recommendations of the European High Level Group on Artificial Intelligence (2018) all emphasise this need for security. For example, Section 1.4 of the OECD recommendations, entitled: "Robustness, security and safety", states:

> "*a)AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.*
> *b)To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.*
> *c)AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias*."

31. **Multidisciplinary thinking –** Any thinking on this subject needs to be collective and multidisciplinary insofar as the risk associated with any processing activity will obviously depend on the quality and appropriateness both of the chosen database (consider the example of databases used for automatically identifying wedding pictures, which overwhelmingly feature western-style weddings or databases that identify families of children abused in a different historical and social context) and of the algorithm used. Of particular note in this regard is the Report accompanying Convention No. 108+ (paragraph 86): "*Paragraph 2 clarifies that before carrying out a data processing activity, the controller will have to examine its potential impact on the rights and fundamental freedoms of the data subjects. This examination can be done without excessive formalities. It will also have to consider respect for the proportionality principle on the basis of a comprehensive overview of the intended processing. In some circumstances, where a processor is involved in addition to the controller, the processor will also have to examine the risks. IT systems developers, including security professionals, or designers, together with users and legal experts could assist in examining the risks*." No doubt a written document will record the key points of this assessment, the persons involved in the exercise and the decisions reached.

    The GDPR requires this "risk assessment" procedure to include certain specific steps where high risk profiling is concerned. Article 35.7 states: "*The assessment shall contain at least:*

    a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;

    b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes;

    c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and

    d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned."

32. **Need for the assessment to be multidisciplinary and open –** The Council of Europe Guidelines on AI state: "*In all phases of the processing, including data collection, AI developers, manufacturers and service providers should adopt a human rights by-design approach and* ***avoid any potential biases, including unintentional*** *or hidden, and the risk of discrimination or other adverse impacts on the human rights and fundamental freedoms of data subjects. AI developers should critically assess* ***the quality, nature, origin and amount of personal data used, reducing unnecessary, redundant or marginal data during the development, and training phases and then monitoring the model's accuracy as it is fed with new data****. The use of synthetic data may be considered as one possible solution to minimise the amount of personal data processed by AI applications."*

Other instruments also emphasise the need for openness in this procedure. One such is the Guidelines on Big Data (see paragraph 7.3) which call for openness on two fronts. Firstly, the procedure must be conducted by an **interdisciplinary group**: "*This assessment process should be carried out by persons with adequate professional qualifications and knowledge to evaluate the different impacts, including the legal, social, ethical and technical dimensions.*" Secondly, representatives of the individuals or groups concerned should be involved in the assessment. The Guidelines call for the **groups concerned to play an active part in the process ("participatory assessment")**. "*AI developers, manufacturers and service providers are encouraged to set up and consult independent committees of experts from a range of fields, as well as engage with independent academic institutions, which can contribute to designing human rights-based and ethically and socially-oriented AI applications, and to detecting potential bias. Such committees may play an especially important role in areas where transparency and stakeholder engagement can be more difficult due to competing interests and rights, such as in the fields of predictive justice, crime prevention and detection.*" This seems to us to be sound advice given the complexity and difficulty of gauging the impact of processing and the fact that profiling affects not only the individual profiled, but everyone else too, whether profiled or not. An individualist approach to protection, moreover, is hardly realistic in a world where the power of data controllers far exceeds that wielded by individuals.

It is important that the Council of Europe make the case for consultation or even negotiation with associations representing the interests of data subjects. The Guidelines on Big Data echo this view: "*With regard to the use of Big Data which may affect fundamental rights, the Parties should encourage the involvement of the different stakeholders (eg individuals or groups potentially affected by the use of Big Data) in this assessment process and in the design of data processing*."

33. **Towards the creation of a multidisciplinary body for "ethical" evaluation of AI systems and in particular profiling powered by AI** – In addition to this internal company procedure or in order to support such a procedure, the Council of Europe could recommend that member states **set up a multidisciplinary body for the "ethical" evaluation of AI systems**[59] and, in the context that concerns us here, of profiling operations. The role of this "**independent multidisciplinary national authority for the assessment of risks related to artificial intelligence and in particular to profiling using machine learning**" would be manifold. Tasked with auditing, testing and labelling AI systems in the private or public sectors, this independent authority would act on a mandatory basis in the case of AI used in public-sector activities and, subject to what may be decided by member states regarding high risk systems, on a voluntary basis for systems operating in the private sector.

    This authority would issue opinions on, firstly, any individual or collective profiling envisaged by administrations or the regulatory authority to support their strategies or

[59] See the Report on Artificial Intelligence written by A. MANTELERO for the Council of Europe; this report served as a basis for preparing the Recommendation on artificial intelligence, in particular page 16 et seq. The author makes the case for a national approach to complement the introduction of procedures at company level.

apply regulations, and secondly, on the assessment of risks associated with private or public policies relating to data sharing and open data and support for the design and delivery of good practice. The authority should make recommendations on the quality of Big Data and profiling algorithms in order to ensure their reliability, transparency and compliance with applicable legislation, in particular on data protection, consumer protection, non-discrimination, competition, etc. Although it would work closely with the supervisory authorities, its assessments would have a wider remit than the latter, as they would also cover collective risks. These opinions and recommendations would be made public.

It is further proposed that this authority, subject to any other labelling or certification bodies which there may be, whether industry-wide or other, should also be able to respond to companies' requests to evaluate and "certify" their own systems. This certification mechanism would be purely voluntary but, through the quality label awarded, would help build much-needed public trust and social acceptance. In the proposed internal or external assessment, particular attention would need to be given to the following:

- The need to combat errors, by examining the database sets used and checking for sets or data that are inadequate, that have not been updated or contain errors and algorithms that are badly designed or unfit for purpose or contain bugs that may lead to undesirable results;
- The need to abide by the principles of "privacy by design" and "privacy by default" when designing the profiling process itself;
- The need to provide for a test phase and evaluation of the results of this test before actually commencing profiling;
- The need to make data subjects aware of the existence of the current or planned profiling with an obligation to ensure easy access (by clicking on an icon?) to a more detailed description of the characteristics of the profiling carried out in respect of them;
- Generally speaking, measures that facilitate understanding of how profiling works, the resources used, the impact on the person, and how to challenge the outcome of any such profiling. The OECD describes the purpose of such measures when they relate to a profiling system using AI as follows: "*AI actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:*
  i.to foster a general understanding of AI systems,
  ii.to make stakeholders aware of their interactions with AI systems, including in the workplace,
  iii.to enable those affected by an AI system to understand the outcome, and,
  iv.to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision."

III. Initial thoughts on applying the principles to a few identified categories of profiling.

34. **Reminder of the principle that regulatory measures must be proportional –** Below we share a few thoughts on both the risks and the kind of measures that could be devised for the types of profiling identified. No doubt there will be other possible suggestions and, perhaps too, some objections but in our view, it is imperative that we conduct this exploratory process in order to show how dangerous and, in many cases, potentially excessive any regulation that sought to capture all profiling activities would be.

A. Profiling in a precontractual context

35. **Profiling for advertising purposes** – Among the profiling operations liable to be carried out in the contractual phase, a distinction will thus be made between those which consist in sending out ads based on categories which have either been specified by the controller, in full or in part (supervised AI systems), or been discovered by the system itself (unsupervised AI systems). In cases like these, we will merely reiterate the need for internal risk assessment focusing on the quality of the data used, with care being taken to ensure that the data subject is informed about the profiling being carried out (via an icon) and has some simple means (eg clicking on the profiling alert icon) of learning about the characteristics of the system and how it works and even the quality label and of objecting, where necessary. The "privacy by default" rule in any case requires that certain service providers which have a dominant position in the market or operate a service that is in the public interest (such as news platforms or social networks) must offer a meaningful and non-discriminatory alternative (eg via the rule that any who refuse must "pay" for access) to profiling for advertising purposes or, at any rate, modulate the ways in which the data gathered can be used, as already happens with some sites.

36. **High risk profiling** – As we see it, there are five circumstances in which further measures aimed specifically at high risk processing are required. Besides the measures already indicated, each of these cases merits a few additional comments. The first is, of course, profiling **young people**, where there is a need to counter the heightened risks of manipulation. One option here might be to have a system of parental authorisation and, in any case, intervention by an internal evaluation committee. The second case concerns profiling that allows **dynamic pricing:** information on this particular feature of the profiling must be given to the data subjects and consumer representatives must be involved in the risk assessment with respect to the pricing margins and also the criteria to ensure they are not discriminatory. Thirdly, there is the case of profiling which may or may not use **sensitive data** but which, directly or indirectly, has the effect of revealing such data. Profiling of this type must be prohibited except in special circumstances, in particular where the data subject has been offered a non-profiling alternative but has explicitly consented to the processing

of sensitive data, in direct connection with the quality of the service offered[60]. In any risk assessment and when evaluating tests prior to the use of profiling, particular attention must be paid to the possibility of bias favouring discrimination based on sensitive data or the exclusion of certain persons, no matter what the criteria.

The last two cases likewise call for comments as they involve taking decisions that have the potential to "significantly impact" the individuals concerned. The fourth situation, profiling for the purpose of **selecting customers**, must be based on **relevant criteria directly related to the proposed transaction**. Clearly, some criteria will be automatically excluded here. It is important to maintain a fair balance between the benefits to companies, in particularly lending institutions, of having a thorough understanding of their future customers, as required by law and as summed up in the phrase "Know your customer", and the need to avoid judging people based on decision-making criteria which go beyond what the data subjects could reasonably and legitimately expect. Lastly, it might be advisable here to include a provision to the effect that profiling may or even must be certified by the authority mentioned above (paragraph 33) or by industry-specific laboratories (covering, for example, banking and insurance, the housing sector, etc) accredited by the authority (?) which could independently evaluate the decision-making systems envisaged. In the case of a quality label, a provision could usefully be inserted to the effect that clicking on the label icon would bring up a description of the criteria used to build the profile. Attention is also drawn to the need for individuals to be able to challenge automated decisions and to access their profiles in an easy-to-understand form after processing.

The fifth possible scenario concerns profiling in the context of staff recruitment. Many of the remarks made regarding customer selection could also apply here: the need to offer a right to contest the "truths" emanating from the computer; the use of criteria that are relevant to the job application, access to profiles, etc. It would be recommended labelling the profiling systems used, preferably following consultation with staff representatives, especially if they are based on emotion recognition techniques[61], given the risks of bias and misinterpretation associated with the use of such techniques.

Lastly, the risk of undue manipulation by exploiting the vulnerabilities of others (see paragraph 11 above) requires that particular attention be paid to the economic, intellectual and social status of the parties involved and the techniques used by the controller to bring about the desired decision.

B. Profiling in the context of contract performance

37. **Customer performance profiling** – Profiling for the purpose of evaluating the "performance" of a customer (e. g. bank customer) or employee is high risk because of

[60] Suppose, for example, I wanted the operator of my favourite music platform to pre-select music to match my tastes. The fact that I come from a particular African culture may be relevant in deciding which tracks to include.
[61] It will be noted that these affective computing systems may also be used in customer selection processing (4th scenario).

the consequences that its use entails. It might determine, for example, whether a person secures a loan or gets a promotion. Once again, there is a need to consider the risks associated with these processing activities, taking into account the various interests involved (consumers or staff representatives); checks will need to be carried out to ensure the algorithms are sound and that there is no bias (risk assessment procedure) and the earlier comments about industry-specific labelling are worth reiterating here.

Some basic information will need to be provided concerning the data taken into account, the approximate weight assigned to each criterion and the consequences attached to each profile and these points discussed. Data subjects will be able to see what data were used, where they came from, an indication of the weight assigned to each criterion, and either the "profile" obtained and the consequences of this "profile", or an explanation in plain language of the model used by the algorithm. Lastly, provision must be made for a right to challenge the computer's decision, with no adverse consequences for the person exercising it, etc. To be effective, this possibility of challenging a decision requires that a competent unit with sufficient expertise to query the "truth" coming out of the computer be set up within the company.

C. Profiling carried out by public authorities (except police and judicial authorities responsible for prosecuting criminal offences)

38. **Profiling by public authorities** - Profiling by public authorities is undertaken or may be undertaken for various purposes, in particular as part of a proactive policy of "benevolent government" or for the purposes of law enforcement by introducing monitoring to verify compliance through the use of expert systems or AI systems. The introduction of any such profiling systems should, firstly, be provided for in unambiguous, proportionate legislation appropriate to the needs of a democratic society and, secondly, be subject to preliminary and regular assessments by the independent multidisciplinary body described above (see paragraph 33) both in terms of the algorithms to be used or already in use (data quality, absence of bias, etc) and system security. **In our opinion, transparency of algorithms and sources is important insofar as it makes it possible to meet both the obligation to have access to public documents and the requirement to state the reasons on which public authorities' decisions are based.** Clearly, it should be possible to waive this rule in cases where transparency might not be in the public interest (disclosing the criteria used to combat benefit fraud, for example, would be counterproductive). At the same time, two precautionary measures would appear to be in order here:

- introduce, from an organisational and management perspective, 1. a genuine supervisory procedure for decisions proposed as a result of profiling; 2. a requirement for public authorities to give advance notice of the use of profiling, along with details of how the system works and the sources on which it is based; 3. a procedure that affords data subjects a real opportunity to express their views and exercise their right to a simulation in order to see how the algorithm would affect them.

- where the profiling results in a decision that is detrimental to the data subject (eg denial of social assistance or triggering of checks for benefit fraud), create a category to make it clear that the person has been identified as such via profiling.

In particular, this body would be involved when profiling is carried out for the purposes of public administration. In April 2018, for example, the AI Now Institute devised a framework for public bodies wishing to implement algorithmic decision-making tools to enable citizens affected by these tools to challenge decisions affecting them. The recommendations are also aimed at the designers of these systems, which are already widely used in government. The AI Now initiative[62] accordingly calls for an end to the use of opaque systems for public decision-making, in order to ensure fair and lawful procedures and to protect members of the public against discrimination of any kind. Such systems must also respect the right to public information, as enshrined in the European directives on access to official documents, and, in addition, comply with the duty incumbent on public authorities to give reasons for their decisions.

To this end, it is recommended[63] that a regular audit be carried out by the authority, which would have an AI systems assessment centre, a test and audit centre that would operate independently of the State. This centre should also ensure that efforts are made to improve the expertise of the organisations that design AI systems used by the authority and its staff. It would set out the organisational and technical procedures whereby citizens could challenge decisions taken on the basis of AI systems. The initiative recommends that government agencies identify and describe any automated decision-making systems, including an assessment of their scope and impact. It further recommends that access procedures be put in place so that researchers, independent experts, associations or journalists can access and evaluate these systems, and that agencies see to it that any private companies supplying them with systems agree to these checks. It also emphasises that agencies need to develop better skills in order to have expert knowledge of the systems they introduce, and in particular to improve the way they communicate with the public, and invites solution providers to give priority to fairness, accountability and transparency in their offerings. That would also allow public bodies to develop procedures for mediation and challenging decisions. Lastly, requiring systems to publish impact assessments of their automated decision-making tools could afford the public a means to assess the tools and the transparency of services.

[62] As G. BRAIBANT wrote in 1971,"(it) should require the public authority, whenever it relies on the results of processing by computer, to make known the data and programmes from which these results have been obtained; these data and programmes may thus be the subject of discussions likely to call into question their results.". (G. BRAIBANT, « La protection des droits individuels au regard du développement technologique », Revue internationale de droit comparé, 1971, 23, p. 812)

[63] See in particular the report by the UK's Select Committee on Communications, *Regulating in a digital world,* 2nd Report of Session 2017-19, House of Lords, published 9 March 2019

Chapter III: Analysis of the application of the various provisions of Convention 108+

39. **Purpose of Chapter III** – In this chapter we seek to draw attention to the fact that the provisions of Convention 108+ as applied to profiling call for some interpretation which may serve as a basis for recommendations.

A. Object and purpose of intervention in the field of profiling

40. **From individual risks to collective ones: where data protection is falling short -** As we have said before, the recent Convention 108 Committee Guidelines place considerable emphasis on this point. As noted in the MANTELERO Report, which introduces the Guidelines on Big Data, "*Of course, data protection per se does not cover all these aspects, which require a broader approach encompassing human rights and societal issues*". This point has numerous ramifications where profiling is concerned. Clearly, there is a vital need to move from a "data protection risk assessment" and a "data protection by design" approach to a more comprehensive one based on "ethical values assessment and design" where concerns about social justice, cultural diversity, human dignity and democracy feature alongside those that relate to data protection alone. That does not mean turning one's back on the Convention or the cause of data protection. As the rapporteur astutely observes, "*Data protection's focus on individuals, an awareness of the social consequences of data use and the link with personality rights may expand the data controller's approach beyond data protection to fundamental rights and collective interests. Regarding its complementary role, data protection helps to reveal the way data are used and the purposes of processing, which represent key elements in a better understanding of the potential consequences for a variety of rights and freedoms*." Among other things, such a stance calls for a widening of the scope of the risks to be considered when assessing profiling, and for attention to be given to other interests, alongside those of data protection representatives. As noted by SCHULTZ and DREYER in the report cited above: "*The GDPR and Automated Decision making : Will it deliver ?*"*: "Yet, the GDPR does not offer great potential when it comes to protecting group-related and societal interest such as non-discrimination, social inclusion or pluralism … there is a need for a complementary approach."*

B. Definitions

41. We will focus here on two terms: "personal data" and "data controller".

a. personal data

42. **The concept of personal data** – The first term calls for the following comments. It is important to note that the notion of personal data now extends beyond the criterion of identifiability, which refers to the possibility for the profiler to link the data collected directly or indirectly to the identity of the data subject, to include contactability and

the possibility of "impacting" the data subject. It effectively marks a shift from identification to individuation. For example, the RFID tag worn by a shopper X moving around a supermarket might not reveal the identity of its wearer but it will make it possible to locate the person within the supermarket and send them an ad for an item in the vicinity. This raises questions about the ability of data subjects, who, although not identified, have certainly been singled out, to exercise their rights. It is important that they be able to exercise the latter (for example, the right to object or simply a right of access) online without revealing their identity. The recommendations must also take care not to confine themselves to personal data only, and to consider any **anonymous data** that go into constructing the profile. To make the data subject aware only of the personal data used to produce the profile, and to leave out anonymous data, would be to provide them with information that was incomplete or even meaningless. Whether data is anonymous or not should be judged having regard to the purpose of all the operations that lead to a person or group being profiled. The need to promote a holistic view of profiling (see the 2010 recommendation) is worth reiterating here. There can be no question of separating the various stages of profiling (data collection, data mining and application to a particular data subject or group) and judging each on its merits. It is from the end stage that all the other stages must be assessed. It seems to us, furthermore, that profiling (see our earlier comments on objectives and purpose) carries with it a major risk of discrimination against groups and even businesses. This brings us back to the debate, one that is no doubt worth rekindling, about the protection of groups, whether ethnic groups or groups of people with a common profile (residents of a particular neighbourhood who are labelled potential criminals, neighbourhoods condemned to long-term industrial decline, etc).

43. **What about legal entities?-** This same concern follows on from a discussion that is still going on at the Council of Europe[64] and has to do with the **protection of legal entities**, for whom profiling is increasingly a reality, sometimes with calamitous consequences. Entrepreneurial freedom is undermined by the information asymmetry that is currently a feature of the market for information. Thanks to digital technology, some platforms, some information service operators, banks and insurers have powerful tools at their disposal, including AIs capable of predicting the future of a company or subsidiary or even dooming it to failure with negative rankings. I believe there is a particular need to protect small and medium-sized companies against profiling of this type, including for the workers employed there and the areas where these companies are located. In saying this, **we are not advocating that Convention 108+ be extended to include legal entities – protecting the latter requires a different mindset, one more akin to competition law than citizens' rights. We do, however, believe that subjective rights of access, rectification and erasure modelled on those conferred on individuals, protection against bias, etc should be granted to legal entities, not within the framework of the recommendation that concerns us here but rather in a regulation with a more economic focus. Such an instrument, we believe, would be useful – essential even - in an era when predictions are being powered by AI.** Worth mentioning in this context are the new rights bestowed by the recent European Union Regulation 2019/1150 of 20 June 2019 promoting fairness and transparency for

[64] J.P. WALTER, Le profilage des individus à l'heure du « cyberspace » : un défi pour le respect du droit à la protection des données.

business users of online intermediation services (OJEU 11.7.2019, L 186/57), which aims to protect the companies using platforms' services, in particular where their rankings are concerned.

b. **The concept of data controller**

44. **Multiplicity of actors and arrangements and qualification of the various actors –** In our introduction we drew attention to the variety of arrangements under which a profiling system may be put in place, especially when it is based on AI technologies. Data can be derived from various sources, each with its own controllers, while the datasets, on which algorithms have already been running, may come from somewhere else; some basic algorithms are available on the web in open source while others are purchased from their designers under private licences. The task of adapting these different components to the needs of a given sector or company is often performed by specialised players. Last but not least, the task of applying profiling to individuals or groups may be entrusted to a third party, usually as a "turnkey" service provided by a platform. The latter, after all, will already have reams of data relating to the persons concerned and in many cases will already be using profiling algorithms for its own purposes. As we have pointed out (see paragraph *** above), it may make such algorithms - or even the results of these algorithms – available to third parties. The fact that all these players are involved raises the question of what status should be accorded to each one. The position of the authors of the two Convention 108 Committee Guidelines has already been established. As we have pointed out, both instruments seek to extend the obligations incumbent on each player directly or indirectly involved in the design, implementation and operation of profiling activities. There are two ways in which this widening may occur.

For some suppliers of commercially available components needed to operate a profiling system (a basic algorithm, a dataset, a database), what are required, naturally, are product quality, a description of the limitations of the product and, where appropriate, co-operation in the risk assessment and during the test phase. Should we perhaps go further and consider some of these actors to be processors or even joint controllers in respect of the profiling in question? Ruling in the “Facebook Fan Pages” case, the ECJ concluded that, in some circumstances, those involved in profiling may be deemed joint controllers. In the case in question, a non-profit-making association running a singer’s fan club had wanted to “profile” its “clients”, ie readers of the singer’s blog. To this end, it obtained (in return for payment) data relating to the blog readers from the administrators of the platform by which readers accessed the blog, namely Facebook. In short, the former, the association, determined the purpose but the latter, the platform, determined the means. The latter was accordingly found to be jointly responsible and, to quote the Council of Europe convention, a co-decision-maker sharing with the association “decision-making power with respect to data processing”: “*When assessing whether the person or body is a controller, special account should be taken of whether that person or body determines the reasons justifying the processing, in other terms its purposes and the means used for it. Further relevant factors for this assessment include whether the person or body has control over the processing methods, the choice of data to be processed and who is allowed to*

access it." In some instances, the term "processor" might be preferred over "joint controller" in which case it will be important to ensure that the process of selecting the processor is accompanied by safeguards, that a contract is concluded between the controller and this/these processor/s regarding the use of the data, and their security, and that certain obligations are imposed on the processor such as notification in the event of security breaches, risk assessment, etc.

C. Legitimacy of processing and data quality (Art. 5)

45. **Basic principles - Proportionality of processing in relation to the legitimate purpose pursued and a fair balance between all interests concerned are recognised as fundamental**. First of all, it should be stressed that profiling is not an end in itself but rather a technique for achieving one of the many purposes already described in the proposed classification. It is with these aims in mind, therefore, that those responsible for profiling should be guided by the two principles of legitimacy and proportionality, and this, insists Article 5.1., from the design stage and throughout the profiling operation. These principles concern processing and so take us back to the issue of justification for the use of profiling and the ability of the data subject to choose in such matters: is the profiling really necessary and shouldn't the data subject be allowed to choose between non-profiled and profiled access, or even between anonymous or identified access? Of particular relevance in both of these respects – the right to anonymity and the right not to be profiled – is Section 3.7 of the 2010 recommendation, which reads: "*As much as possible, and unless the service required necessitates knowledge of the data subject's identity, everyone should have access to information about goods or services or access to these goods or services themselves without having to communicate personal data to the goods or service provider. In order to ensure free, specific and informed consent to profiling, providers of information society services should ensure, by default, non-profiled access to information about their services.*" Aside from this initial point, it is important that the data subject be able in certain circumstances, in particular profiling for advertising purposes, to determine the purpose for which profiling may be performed and so reduce the scope of the data that will be used. For example, just because you use an online music service does not mean that you are amenable to having your musical tastes profiled, but if you want the provider to recommend or offer you music that matches your tastes, then that is what is required. You should be able to choose the various purposes and, where applicable, the recipients who will make it possible to achieve those purposes. So, staying with the example of online music services, it may be that I wish to be able to receive advertising from third parties about the release of a song by my favourite artist. Equally, however, the reverse may be true. In other words, the fact that I consent to the data controller carrying out profiling for advertising purposes does not necessarily mean that I consent to profiling by third parties, or to my data or profile being released to third parties.

46. **The need for assessment** – These principles likewise call for **participatory, multidisciplinary and if possible multi-stakeholder risk assessment** and require that any such assessment be ongoing. Continuous assessment is all the more necessary as AI systems that support large numbers of profiling operations may, as new data are

gathered and new possible correlations discovered, have their purposes altered by those who use them. For example, a profiling activity that was originally designed for one-to-one advertising may, when fed with additional data and new algorithms, develop into software for excluding or evaluating "clients". Prison administrations, for instance, can use data relating to prisoners' status, behaviour and environment to develop profiling that will be of assistance in making decisions about early release. Capturing location data from our mobile phones certainly makes it easier for users to find nearby restaurants or hotels (the primary purpose) but such data may then be sold to third parties or used by the same operator as the one which performed the initial processing to, say, carry out advertising for department stores or other providers of goods and services.

47. **IA systems and the principle of purpose** – Complying with the principle of purpose limitation poses a challenge, therefore, in a world of **complex, constantly evolving information systems** which, at the whim of some AI, can be co-opted to perform new tasks linked to machine learning opportunities. This applies not only to profiling and raises the question of how such a fundamental principle as purpose limitation can be observed. Article 5.4 introduces a degree of flexibility in terms of how the principle is applied in that it does not prohibit purposes which are "compatible" with those of the original processing operation. The 2017 European Union Guidelines contain a reminder of the criteria by which compatibility is to be assessed, as set out in Article 6.4 of the GDPR[65]: "*the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia: (a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing; the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller; (c) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10; (d) the possible consequences of the intended further processing for data subjects; (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.*"

This long list of criteria can be used to inform the controller or controllers and must be taken into consideration when continuously evaluating profiling and, in any case, throughout the development phase, ie from the initial design, in the choice of algorithms, datasets, test planning, etc. Suppliers of components such as databases will also be required to pay attention to these criteria as part of their duties to provide information and advice. It should be noted that the generic purpose of some of these databases (eg a set of photos enabling facial recognition algorithms to work in different

[65] In the case of public authority processing including activities involving the police, we believe that the only circumstance in which such "slippage of purpose" may be justified is when performing public tasks, under the supervision of the body described above. The Council of Europe, furthermore, concedes that, subject to "complementary safeguards" (rules on access to data, pseudonymisation, system of accreditation for research laboratories, non-commercialisation of results rule, etc) data collected for primary purposes (eg for treating patients) may also be processed – in this case used in profiling - for the purposes of scientific or historical research or for statistical purposes (eg cancer research).

contexts)[66] requires specific attention of this kind. As noted in the Guidelines on Big Data (paragraph 2.5.), therefore, it is with regard to each application that the person responsible for any datasets marketed commercially must *''identify and evaluate the risks of each processing activity involving Big Data and its potential negative outcome on individuals' rights and fundamental freedoms, in particular the right to the protection of personal data and the right to non-discrimination, taking into account the social and ethical impacts*".

48. **What about platform operators?** A special point needs to be made with regard to platform operators. The activity linked to visitors' use of the various sites made accessible or in any case accessed by these operators generates data that the latter consider as their own. The processing of these site usage data (and not just data relating to the choice of sites visited), however, generally exceeds the purpose pursued by the Internet user, who, although they certainly go through the operator's platform, will not necessarily want the operator to make use of this fact for the purposes of profiling or marketing to third parties. There may, of course, be situations in which the Internet user will be willing to allow, or even actively intend for, the platform operator to use their data in this way because they want the operator to analyse their profile in order to select sites that match their personality, or even, albeit less commonly, to share information about their behaviour and personality traits with third parties. As a general rule, however, such uses are not legitimate and should be permitted only with the data subject's consent[67].

---

[66] These databases may be used in processing for a variety of applications developed by the data holder or by third parties (eg strategy development, customer profiling, scientific research, etc).

[67] In this connection, but with a different focus, the concern this time being to protect business users of platforms' services, the recent EU Regulation 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (OJEU 11.7. 2019, L 186/57) states (recital no. 34): "*In the same vein, it is important for business users to understand whether the provider shares with third parties any data which has been generated through the use of the intermediation service by the business user. Business users should in particular be made aware of any sharing of data with third parties that occurs for purposes which are not necessary for the proper functioning of the online intermediation services; for example where the provider monetises data under commercial considerations. To allow business users to fully exercise available rights to influence such data sharing, providers of online intermediation services should also be explicit about possibilities to opt out from the data sharing where they exist under their contractual relationship with the business user.*" Article 9 thus specifically provides for an obligation on the part of the operator to be transparent vis-à-vis business users, in particular : "*Through the description referred to in paragraph 1, providers of online intermediation services shall adequately inform business users in particular of the following:*

(a) whether the provider of online intermediation services has access to personal data or other data, or both, which business users or consumers provide for the use of those services or which are generated through the provision of those services, and if so, to which categories of such data and under what conditions;

(b) whether a business user has access to personal data or other data, or both, provided by that business user in connection to the business user's use of the online intermediation services concerned or generated through the provision of those services to that business user and the consumers of the business user's goods or services, and if so, to which categories of such data and under what conditions;

(c) in addition to point (b), whether a business user has access to personal data or other data, or both, including in aggregated form, provided by or generated through the provision of the online intermediation services to all of the business users and consumers thereof, and if so, to which categories of such data and under what conditions; and

49. **Principle of minimisation or proportionality of the data processed -** The question of proportionality or data minimisation in relation to the purpose pursued throws up other challenges. Machine learning systems rely on statistical correlations based on arbitrary combinations of data. So, to take a purely fictitious scenario, the tax authorities might conclude from trawling through vast data banks that managers of companies with more than 200 employees and fewer than 400, who own a red car registered between year X and year Y, who take all-inclusive package holidays to the Mediterranean, live in a particular type of neighbourhood in cities with more than 50,000 inhabitants and have a child and a dog are potential fraudsters. This example highlights the difficulty, in theory at any rate, of determining what elements will go into building a profile. Reaffirming the principle that data must be appropriate or even necessary in relation to the purpose pursued would demand that such work be done but, unless the criterion is interpreted broadly, users of AI systems will object to a narrow construction on the ground that the requirements imposed hinder innovation and prevent profiling from being fully effective. Consider the example of tracking down criminals. Sometimes unexpected correlations are what makes it possible to identify offenders. The Guidelines maintain the principle: "*Controllers must make sure that they are complying with the data minimisation principle*, ..." and see pseudonymisation as a solution.

    Without repeating all the requirements laid down in Article 5.4 (fairness, specified purposes, need for data to be relevant, accurate and kept up to date, storage period, etc), we will simply flag up a few issues that will need to resolved, initially through a sectoral approach, involving multi-stakeholder discussions, and then, if necessary, through the intervention of the data protection authorities or even lawmakers.

    -To what extent is it acceptable for an insurance company to make use of data relating to insured persons for the purpose of providing personalised services?

    -To what extent is it acceptable for a bank or credit institution to profile its clients, citing its responsibilities as a lender?

    -To what extent is it acceptable for an employer to use affective computing systems to select job applicants or manage the careers of existing staff?

50. **Processing time and data quality –** Two further points on the subject of processing time and data quality. With regard to the period for which data may be kept, it would surely make sense to include a provision that would prohibit the use of some data beyond a certain period (eg data relating to surgical operations should cease to be used once the patient has been in remission for a given period), and the keeping of records

---

(d)whether any data under point (a) is provided to third parties, along with, where the provision of such data to third parties is not necessary for the proper functioning of the online intermediation services, information specifying the purpose of such data sharing, as well as possibilities for business users to opt out from that data sharing."

There is surely a need for this information to also be made available to Internet users, who should be able to ascertain to what extent data gleaned from their interaction with a particular site accessed via a platform are used by the platform itself or possibly even sold to third parties!

of accidents if no further accidents or risks of accidents (eg failed breath test) are reported. In addition, the person profiled should surely be able to insist that the controller disregard any data concerning them once a certain period has elapsed, which may vary according to the type of data and the purpose of the profiling.

The quality of data when they come from Big Data held by a third party should be able to be documented by that third party or even certified by an accredited body. "*In addition to safeguarding privacy and personal data, requirements must be fulfilled to ensure high quality AI systems. The quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may reflect socially constructed biases, or contain inaccuracies, errors and mistakes. This needs to be addressed prior to training an AI system with any given data set. In addition, the* ***integrity*** *of the data must be ensured. Processes and data sets used must be tested and documented at each step such as planning, training, testing and deployment. This should also apply to AI systems that were not developed in-house but acquired elsewhere. Finally, the access to data must be adequately governed and controlled*."

51. **The issue of consent to profiling is crucial** – Consent, of course, confers legitimacy and, as has already been noted, profiling may be considered to be of real benefit to data subjects looking for advice on how to navigate the maze of goods and services on offer. It is difficult to see in this case, however, how consent will differ from consent to a contract insofar as the data subject recognises that the processing of their own and other data is necessary for the profiling service required. Some might argue that the consent is revocable, but the contract that allows the controller to profile a client may be for an unspecified period and be revocable *ad nutum*. Also, the idea of contractual terms and conditions being negotiated collectively between consumer groups and the controller where it is proposed to carry out profiling on a large scale is, we believe, preferable in terms of protecting the privacy of data subjects to individual consent, where individuals are left to fend for themselves in the face of the economic might of the controllers[68].

Having said all that, we should stress that the attributes of consent stipulated in Article 5.2. will seldom be met. They presuppose, subject to yet more requirements, that consent is freely given, ie without manipulation (see above), and that it is a genuine choice and not simply a matter of clicking “I agree”; consenting to profiling must not be a precondition for access to services typically regarded as essential for participation in social activity and, except in case of necessity due to the service itself (eg a distance tailoring service or distance learning), a non-profiled service should be offered by default; Article 5.2. further requires that any consent given be informed, which raises a number of issues (see our comments on Article 8). At the same time, consent presupposes - and this condition is clearly stated in the Council of Europe report[69] - that the other tests of legitimacy set out in paragraphs 1, 3 and 4 of Article 5 have been met. That means, especially in the case of high risk profiling, compliance with the procedures and conditions already stipulated and, in the case of all profiling activities,

[68] On all these points, see, Y. POULLET, « Consentement et RGPD : des zones d’ombre ! », DCCR, 2018, p. 3 – 39.
[69] “*Paragraphs 1, 2, 3 and 4 of Article 5 are cumulative and must be respected in order to ensure the legitimacy of the data processing*” (Explanatory report, paragraph 41; see also paragraph 44).

risk assessment and mandatory compliance with the principles of proportionality, purpose and security, even if construed broadly.

There are other issues, too, surrounding this "consent"': under what circumstances will it be deemed permissible to consent to profiling in exchange for "remuneration"? As we see it, there is no room here for dogmatism. Everything hinges on the extent of the "remuneration" offered by the controller, the data in question, the purposes of the profiling envisaged, the restrictions placed on the scope for profiling and, above all, negotiation with not only the data protection authorities but also consumer representatives.

D. Special categories of data (Article 6)

52. Preliminary considerations – We will merely point out that the sensitivity of any processing operation is, of course, often intrinsic to the data themselves, but perhaps also, in the context of the last indent of Article 6.1., directly related to the purpose of the processing, where the data are not sensitive in themselves but processing them reveals racial or ethnic origin (eg selecting all names ending in "SKI" in order to trace the Polish origin of the persons concerned), political opinions (inferred from a person's presence in the street during a political demonstration), trade union membership, religious beliefs, health status or information about a person's sex life. It should be emphasised that numerous conscious or unconscious biases[70] can thus affect the operating tools used in profiling, both in the selection of source data and in the choice of algorithms. It will also be noted that such profiling is "high risk" insofar as the risks of discrimination and restriction of religious, trade union or philosophical freedoms attached to the data are significant (see Article 6.2.). Last but not least, any profiling of groups (and not only individuals) carried out in this way must likewise be regulated.

E. Data security (Art. 7)

53. **Security: a broadly defined concept** - The use of AI systems in profiling underscores the need to define the integrity of processing as the search for and elimination (obligation of means) of any biases that might affect that processing. Once again, the use of third party algorithms[71] will require the controller, in line with the "reasonable

[70] J. GRIMMELMANN and D. WESTREICH, "Incomprehensible Discrimination", *California Law Review*, 2017, vol. 7, pp. 170–176. See the comments by L. EDWARDS and M. VEALE, "Enslaving the algorithm: from a 'right to an explanation' to a 'right to better decisions'?", *IEEE Security and Privacy Magazine*, 2018, n.° 16, p. 52: For instance, biases may cause discrimination in automated decision-making on the basis of skin colour, or other legally protected attributes. Such biases may directly rely on those attributes, or indirectly. In the second case, AI tools detect correlations between diverse non-protected attributes to produce the same discriminatory results as if they were taking into account legally protected ones (proxies). An example of this could be the use of a person's address (ie in an area where there is a high concentration of people with black skin) and his average wage to refuse credit, whereas a person with the same attributes but living in an area where most people have white skin would receive the credit. What would be your attitude if the problem were not the colour of people's skin but simply the address?

[71] The issue arises when the algorithm is "open source", supplied free of charge by a research body, for example. Placing onerous obligations on such designers would be counterproductive and would impede the flow of "products" capable of being improved upon by those who use them, thus stifling innovation.

coder" principle[72], to insist on seeing the results of tests undergone by the algorithm producer or even a quality label issued by an independent body, if the processing, at any rate, involves significant risks for individuals or society[73]. It has been observed that very often these algorithms are developed by scientific research centres, whether private companies or universities, and will have featured in publications that have been the subject of extensive discussion and in many cases criticism from third parties (see Chapter 1). We should point out that in some sectors, in particular health care, algorithms are certified by industry experts.

The process of assessing "security" in terms of integrity, confidentiality and availability should start at the design stage of the profiling system and be continuous (see the 2010 recommendation). There should be a report identifying the risks of external adverse effects on individuals and society and containing an analysis of the technical risks and risks of bias associated with the choice of data and operating system or arising from the environment (risks of external attacks). Arguments must also be presented to justify the choices made. Another issue to be addressed is whether any negative or unlawful consequences of the system's operation are capable of being reversed. The principle of security requires that the "ethical values by design" principles be applied as far as practicable. As noted in OECD Recommendation No. 5 on the security of processing using AI: "*AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.*"[74]

---

[72] P. TERZIS, "The reasonable coder", article available on *Academia edu*, 2019

[73] See the 2010 Recommendation on "'*Profiling*", Article 9.2 : "*Furthermore, in cases of processing that use profiling and entail special risks with regard to the protection of privacy and personal data, member states may foresee: a. either:*

a. that controllers have to notify the supervisory authority in advance of the processing; or

b. that this processing is subject to prior checking by the supervisory authority."

[74] See also the Policy Document prepared by the High Level Group of Experts on AI for the EU: "*Trustworthy AI requires algorithms to be secure, reliable and robust enough to deal with errors or inconsistencies during all life cycle phases of the AI system, and to adequately cope with erroneous outcomes. AI systems need to be **reliable**, secure enough to be **resilient** against both overt attacks and more subtle attempts to manipulate data or algorithms themselves, and they must ensure a **fall-back plan** in case of problems. Their decisions must be **accurate**, or at least correctly reflect their level of accuracy, and their outcomes should be **reproducible**. In addition, AI systems should integrate safety and security-by-design mechanisms to ensure that they are **verifiably safe** at every step, taking at heart the physical and mental safety of all concerned. This includes the minimisation and where possible the reversibility of unintended consequences or errors in the system's operation. Processes to*

F. **System transparency (Article 8)**

54. **Information on what: the logic involved?** The information specifically provided for on the basis of Article 8.1. requires little comment. The second paragraph of the article, however, provides for additional obligations in respect of information where such information is necessary "*in order to ensure fair and transparent processing of the personal data*". It goes without saying that anyone who is being profiled must be made aware of the existence of such profiling via an icon that allows them to access the basic and additional information which controllers are required to provide. In the case of high risk profiling, they should also be told about the procedure regarding participatory assessment (access to the report may additionally be required by the data protection authority) and on how profiling is likely to affect the person being profiled. For that reason, too, the GDPR further requires that information be provided regarding the "logic involved" in the system, the very terms used in the Council of Europe's recommendation of 2010.

55. **A step further** – some thoughts regarding the expression "logic involved":
- the term "logic involved", while it is appropriate when dealing with traditional expert systems, reflecting the thinking of the experts in question in given types of situations or decisions, is not appropriate when using machine learning systems that rely on statistical correlations rather than causal logics. In such cases, there should, at a minimum, be a requirement for the system's operation to be "explainable". The High Level Group of Experts of the European Commission, for instance, has a very broad "explainability" requirement, in that not only the functioning of the algorithm but also its role in the organisational process and the justification for the use of AI must be explainable: "*Linked to this, **explainability** of the algorithmic decision-making process, adapted to the persons involved, should be provided to the extent possible. Ongoing research to develop explainability mechanisms should be pursued. In addition, explanations of the degree to which an AI system influences and shapes the organisational decision-making process, design choices of the system, as well as the rationale for deploying it, should be available (hence ensuring not just data and system transparency, but also business model transparency)*"[75].

- What kind of information is meant by the requirement to make individuals aware of the logic involved or, in the case of machine learning, the requirement to make the

clarify and assess potential risks associated with the use of AI systems, across various application areas, should be put in place."

[75] The OECD recommendation also talks about explainability with regard to the requisite transparency in the case of AI systems, taking a fairly broad view of it: "*AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:*

i.to foster a general understanding of AI systems,

ii.to make stakeholders aware of their interactions with AI systems, including in the workplace,

iii.to enable those affected by an AI system to understand the outcome, and,

iv.to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision."

workings of the system explainable? Article 9.c of the Convention refers to the right to obtain knowledge of the "**reasoning underlying data processing**". This wording is vague and erroneous in that, in AI, it is not possible to talk meaningfully about reasoning, as that would imply a causalist view, where the system operates on the basis of correlations that are, to a greater or lesser degree, supervised[76]. Is it necessary to attend a course on machine learning in order to be able to explain the kind of calculations that are performed? Do the calculations behind a particular decision need to be explained? Is it necessary to give very specific details or will a vague indication of the weight assigned to each item of personal information in the decision suffice? Is there proportionality in the requirement for transparency depending on the nature of the decision and its impact on the data subject? Whatever the case, the data subject must be placed in a position where they can understand how the system works and why a particular unfavourable recommendation, decision or prediction about them has been made[77]. Clearly that means providing information about anonymous or

[76] With regard to the expression "logic involved" used in the GDPR, see the *"Guidelines on Automated individual decision-making and Profiling*", published by the Article 29 Working Party and since endorsed by the EDPB, on 3 October 2017 and revised on 6 February 2018, WP251rev.01 (text available on the EDPB site: https://edpb.europa.eu/edpb_en: "*Data controllers should find simple ways to tell the data subject[s] about the rationale behind, or the criteria relied on in reaching the decision[s] [… but] not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision*".

[77] By way of comparison, as regards the ranking of websites by online intermediation services, Article 5 of EU Regulation 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (OJEU 11.7. 2019, L 186/57) likewise requires providers of these services to be transparent vis-à-vis business users of their services about the ranking of sites by the online intermediation service provider (platform operators): *"Providers of online intermediation services shall set out in their terms and conditions the main parameters determining ranking and the reasons for the relative importance of those main parameters as opposed to other parameters.*

2. Providers of online search engines shall set out the main parameters, which individually or collectively are most significant in determining ranking and the relative importance of those main parameters, by providing an easily and publicly available description, drafted in plain and intelligible language, on the online search engines of those providers. They shall keep that description up to date.

3. Where the main parameters include the possibility to influence ranking against any direct or indirect remuneration paid by business users or corporate website users to the respective provider, that provider shall also set out a description of those possibilities and of the effects of such remuneration on ranking in accordance with the requirements set out in paragraphs 1 and 2.

4. …

5. The descriptions referred to in paragraphs 1, 2 and 3 shall be sufficient to enable the business users or corporate website users to obtain an adequate understanding of whether, and if so how and to what extent, the ranking mechanism takes account of the following:

(a)the characteristics of the goods and services offered to consumers through the online intermediation services or the online search engine;

(b) the relevance of those characteristics for those consumers;

(c)as regards online search engines, the design characteristics of the website used by corporate website users."

Article 6 a) of the draft directive recently approved by the European Parliament on better enforcement and modernisation of EU consumer protection rules (Position of the European Parliament adopted at first reading on 17 April 2019 with a view to the adoption of Directive (EU) 2019/… of the European Parliament and of the Council

unprocessed data and where they came from, how the algorithm works and, although probably without going into specifics, the weight assigned to each type of data in the basic algorithm. It is surely reasonable to suppose that the "logic involved" in any decision requires that the data subject be told about everything that went into producing the result, not only the data but also the ways in which the system combines and relates them. This information should be displayed on a website accessible via the icon that warns the data subject they are being profiled. We also believe that the need for transparency should be paramount and take precedence over any arguments put forward by those controllers who would seek – wrongly - to oppose **any** communication about the logic or explainability of the chosen profiling system on the grounds of copyright, patents or trade secrets[78]. The principle must be asserted that trade secrets and intellectual property rights are but a partial defence against disclosure and may be relied on only to the extent strictly necessary to protect the interests of their holders.

-the use of profiling by administrations must be subject to more stringent rules on transparency, except where it is justified by overriding reasons in the general interest or where the purpose of the processing is to identify fraudsters or other criminals. The idea of having systems, in particular AI systems, assessed by a supervisory body, especially in order to avoid bias and discrimination, and of publishing the algorithms involved in the operation, a corollary of the obligation on public decision-makers to give reasons for their decisions, would seem to us to be essential. This obligation has implications for public procurement, especially for suppliers or rather designers of algorithms.

-Last but not least: as a result of this transparency, civil liberties or civil society groups or even the prosecution service could litigate on behalf of individuals who have been excluded or injured as members of discriminated groups, on behalf of the groups themselves or in the public interest. In the context of one-to-one insurance, which poses a threat to the principle of mutualisation, one could imagine, for example, representatives of the public interest or consumer groups acting on behalf of persons affected by the system, even if they are not necessarily data subjects. Whatever the case, the insistence on transparency remains one of the key requirements necessary to ensure confidence in the functioning of our systems. We believe, therefore, that the Council of Europe should encourage research in this area and urge companies to make their algorithms available to all on an open-source basis.

…as regards better enforcement and modernisation of EU consumer protection rules) takes a similar line, focusing this time on the information that operators must provide to consumers who use their platforms, where ranking sites is concerned.

[78] Compare recital 63 of the GDPR which reads: "*This right (of access to the logic involved) should not affect the rights and freedoms of others, in particular, not adversely affect trade secrets or intellectual property or the copyright protecting the software. However, the result of those considerations should not be a refusal to provide all information to the data subject.*"

55. Rights of the data subject in automated decision-making - The rights of data subjects "*not to be subject to a purely automated decision significantly affecting them without having their views taken into consideration*" call for the following comments:

-the expression "significantly affecting" has been extensively discussed in similar terms by the European Union's Article 29 Group. The idea is that consideration should be given to the context, the data subjects and the duration of the consequences (for example, denying someone a visa after they have applied for immigration cannot be equated with using profiling to set prices);

-the term "purely" poses a problem: it will need to be determined to what extent the human being responsible for acting on the proposal generated by the computer possesses, within the framework of their organisation and the procedure prescribed therein, a genuine capacity, in terms of time and competence, to deviate from the proposal and actually exercises that capacity;

- "without having their views taken into consideration" requires that the person be duly informed of this possibility of expressing their views, that specific provision be made within the organisation for the procedure to be followed in such matters and that, as far as possible, the person in question be able to get help and support.

To be effective, this last requirement demands under Article 9.1.c) that the person be able to obtain a written explanation of the criteria used to justify the decision and how these apply in their specific case. Obviously, exceptions[79] must be permitted in cases where access to this reasoning runs counter to the higher interests of the controller, as enshrined in law[80] (eg in the fight against fraud) (see Article 9.2., which mentions,

---

[79] In our view, Section 5.4 of the 2010 recommendation is worth repeating here: "*If there are any grounds for restricting the rights set out in this section in accordance with Section 6, this decision should be communicated to the data subject by any means that allows it to be put on record, with a mention of the legal and factual reasons for such a restriction. This mention may be omitted when a reason exists which endangers the aim of the restriction. In such cases, information should be given to the data subject on how to challenge this decision before the competent national supervisory authority, a judicial authority or a court.*"

[80] Needless to say, this law must be specific, unambiguous and pursue objectives consistent with the requirements of a democratic society. Sweeping references to exceptions such as "consent" or "the existence of a contractual relationship" are inadvisable, in our view. Section 5.5 of the 2010 recommendation reads:

"Where a person is subject to a decision having legal effects concerning him or her or significantly affecting him or her, taken on the sole basis of profiling, he or she should be able to object to the decision unless:

a. this is provided for by law, which lays down measures to safeguard data subjects' legitimate interests, particularly by allowing them to put forward their point of view;

b. the decision was taken in the course of the performance of a contract to which the data subject is party or for the implementation of pre-contractual measures taken at the request of the data subject and that measures for safeguarding the legitimate interests of the data subject are in place."

See also recommendation 6 which widens the scope to include the principles of legitimacy and transparency of profiling: "*Where it is necessary in a democratic society for reasons of state security, public safety, the monetary interests of the state or the prevention and suppression of criminal offences, or protecting the data subject or the rights and freedoms of others, member states need not apply the provisions set out in Sections 3, 4 and 5 of the present recommendation, where this is provided for in law*."

however, the need for "*suitable measures to safeguard the data subject's rights, freedoms and legitimate interests*[81]") or where such access is impossible owing to the complexity of the correlations performed by the machine, making the machine's operating "model" (deep learning system) intelligible, even to the system developer. In this last instance, the controller will nevertheless provide all the information in their possession as part of the "explainability" requirements (see our comments on Article 8 above). In this respect, the High Level Expert Group on AI appointed by the European Commission calls for the various decisions taken by the system to be "traceable": "*The **traceability** of AI systems should be ensured; it is important to log and document both the decisions made by the systems, as well as the entire process (including a description of data gathering and labelling, and a description of the algorithm used) that yielded the decisions*".

56. **Rights of the data subject** - The following recommendations (Sections 5.2 and 5.3) from the 2010 text are worth citing here: "*Data subjects should be entitled to secure, as the case may be, correction, deletion or blocking of their personal data, where profiling in the course of personal data processing is performed contrary to the provisions of domestic law which enforce the principles set out in this recommendation. Unless the law provides for profiling in the context of personal data processing, the data subject should be entitled to object on compelling legitimate grounds relating to his or her situation to the use of his or her personal data for profiling. Where there is justified objection, the profiling should no longer involve the use of the personal data of the data subject. Where the purpose of the processing is direct marketing, the data subject does not have to present any justification*."

G. Additional obligations depending on the characteristics of the profiling (Article 10)

57. **The principle of accountability** - Articles 10.1 and 10.4 enshrine the principle of accountability. It is for the controller(s) to show that they have complied with the obligations stemming from any profiling performed by them, having due regard to the risks associated with these operations. The quality of the data will need to be checked, therefore, to ensure not only that they are accurate and up to date, but also that there is no bias in the way they are used in a given application. The same caution extends to the algorithm(s) used, whether developed by the controller themselves or elsewhere. The next step will be to document the various operations involved in processing, and keep logs of the decisions made. Lastly, the controller will stipulate the organisational procedures necessary to ensure that the right to have a human involved in the decision is respected and that data subjects are genuinely able to express their views and, should they do so, to have them taken on board. Clearly, all these obligations could be

[81] For example, in profiling carried out for the purpose of investigating crimes, any suspects identified in this way would need to be placed in a clearly separate category from suspects identified through traditional investigative methods.

the subject of certification by the AI systems supervisory body mentioned above or by other bodies accredited by it.

This obligation on the part of the data controller does not mean that the other players operating on a commercial basis and offering one or more components (datasets, algorithms) are absolved of their responsibilities. Accordingly, depending on the "foreseeable" risks associated with their operation, the algorithm provider will document the product they are marketing, describing the applications for which it is designed or, on the contrary, those for which it must not or should not be used, and the fact that it has already been applied or tested, and will collaborate during the test period, etc. It is, of course, conceivable that these products might also be certified for a particular sector or for general use.

Article 10.2 introduces a duty for controllers and, where applicable, processors to carry out a risk assessment and to implement organisational and technical measures to reduce these risks in the case of any processing. It goes without saying that it is the internal responsibility of the organisation that uses profiling, or develops tools for the purpose of profiling, to train those involved in carrying out the processing or producing the tools to detect and combat the risks associated with profiling. In addition, these bodies will, where appropriate, promote the interdisciplinary and open discussion necessary for such detection and consideration. We have talked at length about what this obligation means in the case of high risk profiling and would remind the reader that this notion ought to be explored further, not only through sectoral approaches (medicine, retailing, platforms, banking and insurance, policing, etc) but also according to certain types of relationships (see the issue of profiling in the context of labour relations). Any risk assessment of this kind requires consultation with external groups and associations. Data protection authorities should also play a prominent role here.

H. Role of supervisory authorities (Article 15)

58. **Some roles for the supervisory authority with regard to specific aspects of the use of AI in profiling** - Multiple roles should be assigned to the supervisory authorities, working together if possible. Collaboration with the independent interdisciplinary authority responsible for testing, evaluating (mandatory evaluation at least for AI profiling performed by public authorities, voluntary for others but with a requirement for those responsible for high risk processing to send the risk assessment report to the authority and to inform it of the steps taken to mitigate this risk), identifying, where appropriate, best practices and issuing labels certifying compliance with the requirements of the Convention. We would also draw attention to Section 9.2 of the 2010 recommendation which reads: "*Furthermore, in cases of processing that use profiling and entail special risks with regard to the protection of privacy and personal data, member states may foresee either: a. that controllers have to notify the supervisory authority in advance of the processing; or b. that this processing is subject to prior checking by the supervisory authority.*" In our opinion, this competence requires a certain degree of expertise on the part of the **supervisory authorities** with

regard to the sometimes highly sophisticated data processing techniques or, at the very least, a basic knowledge of how they work, their limitations and potential. The aim will be to strengthen the links between AI academics and supervisory authorities. Academics can also play a role in advising the Convention Committee of Convention 108.14.3.b. It is essential that academics specialising in AI be involved in every standardisation process lest it become "artificial" (sic) or completely "disconnected" from technical and/or technological reality.

Member states should mandate the independent supervisory authorities to ensure compliance with the domestic law implementing the principles set out in this recommendation. To this end, they must have the necessary powers of investigation and intervention, in particular the power to hear claims lodged by individuals. Furthermore, in cases of processing that use profiling and entail specific risks with regard to the protection of privacy and personal data, member states may, particularly in the case of high risk processing and profiling by public authorities, either require controllers to notify the supervisory authority of such risks in advance or insist that this processing is subject to prior checking by the supervisory authority.

59. **Two rather sensitive issues** deserve attention. The first concerns the duty on the part of the supervisory authorities, on the one hand, and the consumer protection and competition authorities on the other, to engage in **co-operation**, with regard to the assessment and supervision of profiling; there is also a duty to co-operate with the institutions responsible for ensuring equal opportunities or promoting democracy. The second point is more tricky. We have observed that, in addition to those matters that are unquestionably within the realm of data protection, there are wider issues at play, relating to social justice, group protection and the threat to democracy. It is evident from the texts that the focus is on protecting individuals without considering the challenges that digital technology presents for society, even if, in practice, the authorities are not afraid to act on a broader front. This tendency on the part of the supervisory authorities should be monitored. In my view, it would be dangerous to exclude data protection authorities from such debates when the Convention provides effective instruments for assessing risks (see the risk assessment procedure[82]), for informing data subjects and, above all, for challenging decisions taken on an automated basis. We suggest that the authorities broaden their field of inquiry to include collective and societal risks. They will ensure that their opinions mention such risks and that they are factored into their decisions. Where appropriate, they will initiate debate on the subject. They will draw the attention of member states to the importance of broadening their remit in this area. Until their powers are extended,

[82] It is surprising to note that the Article 29 Working Party's Guidelines endorsed by the EDPB: *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, wp248rev.01 (*available at: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236) mention only the risks incurred by the data subject, without looking at the impact this processing could have on other individuals, groups or society: "*The GDPR does not require a DPIA to be carried out for every processing operation which may result in risks for the rights and freedoms of natural persons. The carrying out of a DPIA is only mandatory where processing is "likely to result in a high risk to the rights and freedoms of natural persons" (Article 35(1), illustrated by Article 35(3) and complemented by Article 35(4)). It is particularly relevant when a new data processing technology is being introduced."*

they might not be able to impose sanctions in this area, but there is nothing to prevent our authorities from exercising supervision, conducting investigations and initiating public debate on the subject and stigmatising any practices that are incompatible, if not with data protection legislation, then at least with democratic values, in particular dignity and social justice. In this context, the supervisory authorities should receive and investigate complaints from associations which concern the collective interests of a particular group or the public interest at large. Where appropriate, the authorities will make recommendations on this subject. They should inform the public of the application of the legislation implementing the principles set out in this recommendation.

One final role that should be assigned to the supervisory authorities is information and public awareness raising about both the benefits and the drawbacks of profiling. That includes educating young people in schools about these issues.

BIBLIOGRAPHY:

ACCESS NOW, "*The Toronto Declaration: Protecting the rights to equality and non-discrimination in Machine-Learning Systems, 2018*", https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

AI NOW Institute, « *Litigating Algorithms : Challenging Government Use of Algorithmic Decision Systems*", 2018, https://ainowinstitute.org/litigatingalgorithms.pdf

M. BRKAN, "Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond", *International Journal of Law and Information Technology*, 2019, eay017, p. 15S.

M. BRUNDLAGE, J. CLARK, G. C. ALLEN, G. C., FLYNN, S. FARQUILHAR, S., R. CROOTOF, et J. BRYSON, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.Information Society Project. Future of Human Institute*, 2018,. Disponible à l'adresse suivante: https://www.repository.cam.ac.uk/bitstream/handle/1810/275332/1802.07228.pdf?sequence=1

J. BURELL, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms", *Big Data & Society*, 2016, vol. 3, pp. 3–4.

A.CAMBON-THOMSEN, "Acteurs et outils de la prédiction génétique: l'éthique au coeur de la gouvernance", *Journal international de bioéthique,* 2016, Vol. 25 -2, p. 159 à 168

C.CATH, S.WACHTER et al., "Artificial Intelligence and the 'GoodSociety': the US, EU, and UK approach", *Science and Engineering Ethics, 2018,* 24 (2), p. 505-528.

CNIL, « *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle.»*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une république numérique, https://www.cnil.fr/sites/default/files/atoms/.../cnil_rapport_garder_la_main_web.pdf

Comité économique et social européen, " *L'éthique des mégadonnées (Big Data): Équilibrer les avantages économiques et les questions d'éthique liées aux données massives dans le contexte des politiques européennes* », 2017, étude disponible sur le site du Comité à l'adresse suivante : https://www.eesc.europa.eu/sites/default/files/resources/docs/qe-04-17-306-fr-n.pdf

Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, "Artificial Intelligence for Europe", Brussels, 25 April 2018, COM(2018) 237 final, p. 17.

Conseil de l'Europe (CoE), *Algorithms and Human Rights*, Study on the human rights dimensions of automated data processing techniques and possible regulatory implications, 2017, disponible à l'adresse suivante : https://rm.coe.int/study-hr-dimension-of-automated-dataprocessing-incl-algorithms/168075b94a

Conseil de l'Europe (CoE). *Technological convergence, artificial intelligence and human rights*. 2017, disponible à l'adresse : http://www.assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=24236&lang

Conseil de l'Europe, « *Lignes directrices sur les mégadonnées* », disponibles à l'adresse : https://rm.coe.int/CoERMPublicCommnSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a .

Conseil de l'Europe, « *Rapport sur l'intelligence artificielle* », rapport établi par A. MANTELERO, T-PD(2018)09Rev., Strasbourg, le 25 janvier 2019, https://rm.coe.int/intelligence-artificielle-et-protection-des-

donnees-enjeux-et-solution/168091f8a5 et « Lignes directrices », T-PD(2019)01, https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40

D.DESAI and J. KROLL, "Trust But Verify: A Guide to Algorithms and the Law", *Harvard Journal of Law & Technology*, 2017, vol. 31, pp. 46–47.

F. DOSHI-VELEZ and K. MASON, "Accountability of AI Under the Law: The Role of Explanation", *Berkman Klein Centre Working Group on Explanation and the Law, Berkman Klein Centre for Internet & Society*, 2017, pp. 2-3.

R. DWORKIN « What is equality? » Part 2: equality of ressources, *Philos Public Aff*, 1981, 10, p. 283-345.

EDPS Ethics Advisory Group, "*Towards a digital Ethics*". Le rapport est disponible sur le site: https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

L. EDWARDS and M. VEALE, "Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For", *Duke Law & Technology Review*, 2017, vol. 16, p. 38 et s.

European Group on Ethics (EGE). 2018. "Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. European Group on Ethics in Science and New Technologies." Brussels: European Commission. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

A.FERRETTI, M. SCHNIEIDER and A. BLASIMME, "Machine Learning in Medicine", *European Data Protection Law Review*, 2018, vol. 4, p. 326.

L. FLORIDI et al. (2018), "AI4People —An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds and Machines*, 2018, 28 (4) p. 689-707.

L. FLORIDI, 'Tolerant Paternalism: Pro-Ethical Design as a Resolution of the Dilemma of Toleration.', *Sci. Eng. Ethics*, 2016, 22(6): 1669-1688.

GOOGLE, "Perspectives in Issues in AI Governance", January 2019, p. 8.

I.GRIMMELMANN and D. WESTREICH, "Incomprehensible Discrimination", *California Law Review*, 2017, vol. 7, pp. 170–176.

Groupe européen d'éthique des sciences et des technologies nouvelles, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. 2018, Bruxelles, Commission européenne, disponible à l'adresse suivante : https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

M. HILDEBRANDT, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Edward Elgar Publishing, 2016.

E.HIRSCH et F. HIRSCH (sous la direction de), *(Les) Les nouveaux territoires de la bioéthique, Traité de bioéthique,* Vol. IV, Collection Espaces éthiques, érès, Toulouse, 2018, ISBN 978-2- 7492-6083-9

Th. HOEREN et M. NIEHOFF,"AI in Medical Diagnoses and the Right to Explanation", *EdpL*, 2018, n°3, p. 308 et s.

IEEE *Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, version 1, IEEE, 2016, disponible à l'adresse suivante: https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead_v1.pdf.

IEEE, *General Principles, 2nd version of Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 2017, document disponible sur le site: https://ethicsinaction.ieee.org/

M.KOSINSKI, D.STIWELL et T.GRAEPEL, « Private traits and attributes are predictable from digital records of human behavior », *PNAS*, Avril 2013, Vol. 110, p. 5803 et s.

J. KROLL, J. HUEY, S. BAROCAS, E. FELTEN, J. REIDENBERG, D. ROBINSON, and H. YU, "Accountable Algorithms", *University of Pennsylvania Law Review*, 2017 vol. 165, pp. 660 – 706.

D.Le METAYER et J. Le CLAINCHE. « From the Protection of Data to the Protection of Individuals: Extending the Application of Non-Discrimination Principles". In S. Gutwirth, P. De Hert, & Y. Poullet (Eds.), *European Data Protection: In Good Health*. Springer Dordrecht, Heidelberg London New York, 2012, p. 315 à 329

T.LINNET, L. FLORIDI et B. van der SLOOT (dir.), *Group Privacy: New Challenges of Data Technologies*, Springer International Publishing, 2017

G. MALGIERI et G. COMANDE, « Why a Right to Legibility of Automated Decision – Making Exists in the GDPR?", *International Data Privacy Law*, 2017, Vol.7, n° 4, p. 243 et s.

A.MANTELERO, Artificial Intelligence and Data Protection: Challenges and Possible Remedies – Report, Comité consultatif de la convention n108 du Conseil de l'Europe, 3 décembre 2018, T-PD(2018)09Rev

A.MANTELERO « AI and Big Data: A blueprint for a human rights, social and ethical impact assessment », in *Computer Law & Security Review,* 2018, https://doi.org/10.1016/j.clsr.2018.05.017.

A.NAUDTS, "Fair or Unfair Algorithmic Differentiation? Luck Egalitarianism as a Lens for Evaluating Algorithmic Decision-Making", *article à paraître*

J. NEW and D. CASTRO, "How Policymakers Can Foster Algorithmic Accountability", *Centre For Data innovation*, May 2018, pp. 20-22.

G.NOTO LA DIEGA, "Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information", *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 2018, n°.9-1, pp. 11–16.

OCDE, *Recommendation on Artificial Intelligence (AI) – the first intergovernmental standard on AI* –, adopté par le Conseil des Ministres de l'OCDE, le 22 mai 2019.

PARLEMENT EUROPEEN, Résolution du Parlement européen du 14 mars 2017 sur les incidences des mégadonnées pour les droits fondamentaux : respect de la vie privée, protection des données, non-discrimination, sécurité et application de la loi (2016/2225(INI)), disponible : http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//FR

G.PASQUALE and D. CITRON, "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, 2014, vol. 91, pp. 5-6.

Y. POULLET, *La vie privée à l'heure du numérique - Essai,* Larcier, Cahier du Crids, n° 47, 2019.

F.ROSSI, « *Artificial Intelligence: Potential Benefits and Ethical Considerations* », Parlement européen, Département thématique C « Droit des citoyens et affaires constitutionnelles », 2016, note d'information PE 571.380, disponible à l'adresse suivante : http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_ EN.pdf..

A.ROUVROY, « *Des donnés et des hommes* » Droits et libertés fondamentaux dans un monde de données massives, 2016, https://rm.coe.int/16806b1659

A. ROUVROY, "Privacy, Data Protection, and *the* Unprecedented Challenges of Ambient Intelligence" (September 11, 2007). disponible sur le site SSRN: http://ssrn.com/abstract=1013984

A.ROUVROY, « Face à la gouvernementalité algorithmique, repenser le sujet de droit comme puissance », 2012, p.42.

A.ROUVROY et Y. POULLET, “The right to informational Self-determination and the value of Self-development: Reassessing the importance of Privacy for Democracy”, in *Reinventing Data Protection?,* Springer, Dordrecht, 2009, p. 159 et s.

D.SADIN, *La vie algorithmique : critique de la raison numérique*, Paris, l’Echappée, 2015

C.SANDVIG, K. HAMILTON, K. KARAHALIOS, and C. LANGBORT, “Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms”, presentation given at the conference *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, Washington, 22 mars 2018

A.SELBST and S. BAROCAS, “Big Data’s Disparate impact”, California Law Review, 2016, vol. 104, pp. 691-692.

P. TERZIS, « The reasonable coder », article en projet disponible sur *Academia edu*, 2019

J. Van DIJK, “Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology”. *Surveillance & Society*, 2014, 12 (2), p.197-208.

C.VILLANI, “Donner un sens à l’intelligence artificielle : pour une stratégie nationale et européenne”, French public report, March 2018, p. 145.

M. WACHTER, B. MITTELSTADT and C. RUSSEL, “Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR”, *Harvard Journal of Law and Technology*, 2018, vol. 31, pp. 845–846

J.P. WALTER, Le profilage des individus à l’heure du « cyberespace » : un défi pour le respect du droit à la protection des données,