14

BIG DATA IN HEALTH AND OTHER SECTORS. ETHICAL QUESTIONS

Louis Dubertret, Alain Bravo, France

Summary and Recommendations on Big Data of the Ethics Commission of the Academy of Technologies of France[261]

14.1. Introduction

The rapid development of technologies changes every day a little more the relationship of man with himself and with his environment. This development is fuelled by three main motivations: the will to innovate, the search for power and profit and the will to improve living conditions. The balance between these three motivations varies in time and space.

It is also necessary to take into account a law, constantly verified in the everyday life: a new tool will always be used in the most unpredicta-

[261] Louis Duberret/Alain Bravo, Synthèse et recommandations, in Académie des Technologies, Big Data: Questions éthiques, Oct 2019, 9-20. Translated from French by the editors, https://www.academie-technologies.fr/publications/big-data-questions-ethiques. | DOI: 10.58863/20.500.12424/4276027 |

ble, even the most harmful, somewhere in the world, what whatever the rules of its proper use. The security specialists computers experience this on a daily basis.

The Ethics Commission of the Academy of Technologies is committed to develop an ethical questioning on the modifications of life human and the environment brought about by new technologies. This identification makes it possible to compare these modifications with the values chosen benchmarks set out in the Universal Declaration of human rights, voted by the United Nations on December 10, 1948. It can thus issue, when necessary, recommendations, or even alerts.

These recommendations avoid, as much as possible, formulating "prohibitions" as these generally prove to be ineffective. They Big Data — Ethical Issues rather incite to encourage technological research towards development securing processes and directing society towards the definition of the rules of good use that must accompany the updating availability of new tools. Ethical reflection can thus become a a powerful engine for innovation, as constantly illustrated by the research intended to improve, inter alia, motor vehicle safety, that of drugs, or that of interpersonal communications, especially on the inside. In this context, the Ethics Commission of the Académie des technologies has centred, here, its reflection on the so-called "Big Data" technology and we present the synthesis.

The control of information systems has always been an issue of power capital for better and for worse. The ability to collect, process, collect and make available a very large number of data in all fields, raises major ethical questions, in particular good use and safety.

14.2 Collection and Storage of Data

14.2.1 The collection of individual data

This collection is done in two very different contexts:

a. Major epidemiological studies or major surveys of opinion. The goal is to study a population with no direct impact on participating individuals. The ethical recommendation is then to find an effective anonymization technique.
b. The collection of individual data with consequences, after analysis, for the individual who provided it. We are then in front of the ethical necessity of a prior agreement. This agreement must specify the use what will be done with this data and explain the consequences possible for the issuing individual. This consent cannot be enlightened only if given specifically for clear use: health, finance, leisure, geolocation, food, social networks… it results in the ethical recommendation to control the fusion of information about these different types of activities. This fusion increases considerably the value of the information collected, but makes more and more illusory any informed consent, hence the importance of a regulation.

14.2.2 Ownership of information collected

This information is only raw material whose value commercial or scientific depends on the quality of the treatment, which is applied. Thus, collected for a purpose, they can be resold and used for another purpose. A system of traceability a priori and has posteriori is therefore recommended, again to clarify the initial choice.

14.2.3 The rights of those who are affected by this information: right to secrecy, right to rectification, right to be forgotten

To make it possible to exercise this right, it is recommended to put in place a technology allowing easy consultation, by each individual, of all data the related to the person. It is then necessary to put at its provision of the technological means allowing it to easily exercise this right to secrecy, rectification or oblivion.

14.3 The Treatment of Data

3.1 The first stage of data processing: checking their accuracy

Whoever says information immediately also says disinformation. Again, a technological research and ethical behaviour are essential. The research is very active and relies in particular on consistency checks and likelihood. One of the major problems is that veracity is not only accuracy and that the omission of data may cause toggle all analysis.

14.3.2 The second step: questioning the control of quality and relevance algorithms used

The choice of analysis algorithms determines the robustness of the results. Most often, the users of the results of the analyses do not have the skill allowing them to assess the relevance of algorithms used to get them. This is true in many areas without forgetting biology where the researcher often uses automatons whose settings he does not know. This highlights the problem the reliance placed on the results of the data processing by algorithms (whether classic or the result of "AI"). An approach rustic statistics, being concerned only with the central part of the curve of Gauss2, can be sufficient within the framework of an approach advertising, for example. On the other hand, when faced with events rare, such as a climatic catastrophe, or unique, such as the adaptation of a treatment to a patient, the stakes are major and these are the extremities of the Gaussian curve[262] to be taken into account. The adapted algorithms are very different and their bad use could have serious consequences.

The recommendation is never to apply the results obtained by a statistical approach, like that of certain algorithms, by example, to an individual or exceptional situation, without going through an expert capable

[262] Or any other relevant statistical law.

of translating information into individual benefit from population studies.

This seems relatively easy when it comes to simple algorithms, but becomes more and more difficult with algorithms of increasing complexity and, in particular, of those used in Artificial Intelligence. It even becomes impossible with the programs who reconfigure themselves.

14.3.3 The third step: questioning the processing of the data collected by the Artificial Intelligence techniques.

The increasing complexity of artificial intelligence systems and, in particular, the development of learning capacities, gives them increasing autonomy. We go from automation to autonomy. This raises the ethical question of the purpose of their programming, of course, but, more and more, of their control and responsibility for actions taken. A simple recommendation could be that of the "red button" allowing the emergency stop. Unfortunately, as soon as several artificial intelligence systems are connected, the sudden stop of one of them can have collateral consequences difficult to predict and potentially serious. Man controls the machine, but the machine also controls man and ethical and technological research is essential for manage these interactions. The recommendation is that man not ever lose never control[263] and therefore the responsibility.

14.4 Use of Results of Data

The use of the results obtained by Big Data technology presents important ethical issues:

[263] In the broad sense: understanding of functioning, possibility of interacting with it.

14.4.1 Fairly distribute the benefits of data analysis

For some, the benefit of data analysis must come first place to those who are at the origin of these data, at the origin of the matter first. For others the main benefit of data analysis must go back to those who bought them to process them. If we refer to the universal declaration of human rights, the priority beneficiary is the one providing the information. The temptation of confiscation profits from Big Data is permanent. A vigilance process should to be placed.

14.4.2 Fairly distribute the benefits of data use

The question is crucial as shown by the few examples below.

- In agriculture, the collection of massive data allows certain seed industries to acquire a detailed vision of the world agricultural production. This can lead to situations of dominance and sometimes monopoly and could even lead to certain loss of sovereignty at the state level. Farmers who will benefit from highly effective decision support tools delivered by these firms, subject to being able to make the necessary investments, will thus gain in precision and efficiency at the risk of sliding into remote-controlled agriculture, making the cultivator a mere performer. This will likely be accompanied by an acceleration of the decrease in their number and the increase in the average area of farms.
- Thanks to the exploitation of Big Data, medicine will be increasingly effective in the fields of prevention, diagnosis and in certain aspects of therapeutic adjustment (medicine known as "precision"). But the risk is to consider that the conclusions from statistical analyses or processing algorithms data, can be applied directly to an individual, by definition unique and different. Precision medicine cannot be confused with personalised medicine. It will always persist on the one hand a "scientific" medicine based on the analysis population statistics and on the other hand the art of caring, which is the art to apply this knowledge to this unique and different patient. However

the temptation is very great, for regulatory, economic and legal reasons, to take refuge behind the algorithms of "scientific" medicine and the increasingly more restrictive practices, and to forget the patient. As for many other sectors of activity, the use of Big Data in the field of health raises many ethical questions, more particularly those relating to respect for individuals and of their freedom, but also to the technical and economic mastery of these new tools, starting with the validation of the algorithms data extraction.

- In all commercial relations between sellers and customers, whether it is banking or internet commerce, it is extremely tempting to use Big Data for the benefit of the seller more than for the benefit of the customer who thereby becomes, more and more, in a situation of dependency.
- In the field of security, the dilemma is particularly precise: How far are you willing to sacrifice your privacy in order to to increase your security? It should be noted in this regard that the relationship between information and security is not linear and that a deletion, total privacy will never allow total security.
- In the field of finance, Big Data are useful tools for control/size the risks taken by both individuals in their choice of savings products, by banks in their market activities and by insurance companies in their pricing. But Big Data can also encourage penalising the insurance of the worst risks and reduce that of the best risks, thus untying the bond of solidarity and weakening the principle of pooling. They can still contribute to disconnect flows of the real economy and transform occasional crises and limited in systemic crises.
- Any new technology has consequences for the world of work and employment. A major ethical recommendation would be to always study and manage these repercussions upstream, as soon as we are considering the implementation of a new technology, with a special attention for the most fragile, that is to say the less qualified.

As we can see, Big Data technology is already modifying and will modify more and more all aspects of our daily lives, our way to communicate with ourselves and with others, and even some aspects of our way of thinking.

14.5 The Recommendations

Intended to turn these technological possibilities into progress, they can be summed up by emphasising the importance of developing in a way very high priority for regulatory tools along the following lines.

14.5.1 It is up to each individual to remain the owner of his privacy, each evaluating the perimeter of the latter as he wishes.

That implies:

- a fully informed prior agreement on the use that will be made of his data and the consequences of this use about his personal life;
- strict control of the merger between information files concerning different areas of activity: health, finance, leisure, geolocation, etc.
- clear traceability of the use of data from their collection until their use, the only way to clarify the agreement prior;
- developing technologies that allow easy consultation by each individual of all the data concerned in order to allow the exercise of the right to secrecy, the right to rectification and the right to be forgotten.

14.5.2 Quality control of algorithms

This includes:

- checking the accuracy and representativeness of the information processed;
- the adaptation of the algorithms to the purpose of the research: simple classification using the central part of the Gaussian curve, or taken into account exceptional or individual events;

- the obligation never to use the results obtained by a statistical approach, to an individual or exceptional situation, without going through an expert capable of translating into benefits/ individual risks of information from data analysis data from population studies;
- the permanent reminder, in particular during the training of the users, that Big Data technology only allows the observation of correlations and that correlation and causation are different concepts.

14.5.3 The implementation of vigilance systems on ethical use of results obtained by Big Data technology.

This concerns in particular:

- the temptation to confiscate these results for the benefit of the most powerful and wealthy;
- the risk of favouring, as we have already observed, situations of global monopoly thus escaping the regulation of the States, resulting in confiscation of property for the benefit of a few;
- the risk that the collection and processing of data does not benefit not first of all to those who are the origin of it. This is particularly important in the fields of finance, health and agriculture;
- the impact of Big Data and Artificial Intelligence on the world work. It is essential, as in all areas of use information and communication technologies, anticipate and manage changes in the distribution of tasks that will result and to put in place, upstream, the strategies which will make it possible to avoid any instrumentalisation of men.

14.5.4 Monitoring the autonomy of tools using artificial intelligence

The use of Big Data and AI must not undermine the fundamental principle of responsibility and must therefore always be possible to determine which is the human who is, in the end, responsible for the direct or indirect impact of these tools on other humans.

Many regulatory structures already exist, first and foremost of which, as far as France is concerned, the Cnil[264], to deal with these important ethical considerations. It would be very useful to coordinate them and unite efforts within a national network, promoting interactions between the ethics committees around the questions asked by all technologies and not only those involved in the development of life sciences. This network should have the possibility, in addition to its advisory and regulatory functions, to highlight areas where technological innovations are needed to respond to the ethical problems posed by technologies.

We will not forget the fact that leading companies in terms of Big Data and information networks are global and therefore, the national reflection and regulatory bodies must imperatively coordinate with each other and do as much as possible stand together.

[264] Commission nationale de l'informatique et des libertés.