Research article

Towards an ethical framework about Big Data era: metaethical, normative ethical and hermeneutical approaches☆

Ariel Antonio Morán-Reyes *

Institute of Research on the University and the Education, National Autonomous University of Mexico, University City, 04510, Mexico City, Mexico

ABSTRACT

The main ethical challenges that arise for Information Sciences (with the daily use in different areas of Big Data applications) are not about the reliability of its professionals to carry out tasks in the organization area in an impartial way or about the obligation to train themselves technologically in the area of Data Science. The most important problems are related to the concept of *moral responsibility*, especially from a metaethical perspective, in line with the reflection of the implementation of technology with respect to human autonomy. In this paper it is stated that the challenges of Big Data go beyond the individual spectrum of responsibility of a professional in Information Sciences (specifically, due to the negative social consequences), so that the changes brought about by massive data sets are essentially problems of a group ethics, so they require approaches from the theoretical postulates of these disciplines. In addition to this, the moral challenges in dealing with Big Data are usually approached from applied ethics (such as information ethics), but in this article it will be approached as a problem of metaethics and normative ethics (as a foundation for its application in professional codes), and also from some ideas of digital hermeneutics and the philosophy of technology.

1. Introduction: the Big Data reality framework

Big data are present in multiple actions of daily life, such as those related to the search for information, the personalization of Internet advertising, the recognition of data patterns, and the predictive keyboard, among others. Currently, one of the main features of the dominant digital ecosystem resides in the flow and increase of content that supposedly comes from diverse sources, including the mass media, blogs, social networks, and so on. In this same sense, the volume and variety of digital information produced, together with the growth of the ubiquity of mobile devices and the rapid advancement and diffusion of computer processes, are characterized by the constant use of algorithms and different forms of machine learning.

This reality directly affects the tasks of governments, companies or researchers, and reaffirms the importance of the operations of collection, analysis and representation of information, associated with techniques based on calculation and quantification, and where they acquire greater importance the intersection between media, technology and society. To understand this phenomenon, it is necessary to delve into the analysis of empirical cases in the generation of a conceptual framework to organize,

interpret and theorize this question, and in the application of critical perspectives that help us to interpret it (Tang et al., 2020).

Several academic, governmental and private institutions have begun to venture into lines of work and institutional projects of an administrative nature that involve the use of Data Science and Big Data technologies. The main lines are based on the following problems:

- The Big Data exceed the capabilities of the tools available in the academic and administrative areas for its realization.
- The Big Data involves the use of some specialized software (frameworks) and High-Performance Computing platforms (clusters), intended for scientific research.
- Big Data methodologies require specialized personnel (something very scarce).

It is true that many of these institutions have not ignored their ethical repercussions, but they have approached them only from certain deontological aspects (Holloway, 2020), and not from moral facts or principles of morality: "The point of morality is not to mirror the world, but to charge it; it is concerned with such things as principles of action, choice,

---

responsibility" (Williams, 2012, p. 33). Among the main strategies that are beginning to be implemented in various companies and organizations to use principles of Data Science are:

- The acquisition of frameworks and clusters dedicated to data processing and storage.
- The acquisition of self-service data analytics software.
- The acquisition of cloud storage tools.
- The contracting of DSaaS ("Data Science as a service").
- Training and recruitment of personnel.

Information Sciences have tried to understand this reality from documentary theories focused on information, but above all with the implementation of metric studies, aimed at collecting and analyzing data on a large scale in terms of software development (with specific methods to analyze the information, e.g. with the analysis of links or with perspectives such as altmetrics). Another documentary effort is the standardization of practices to represent data, among which the integration of the Semantic Web and Linked Open Data can be highlighted. The work of Information Sciences is based on linked documents and not so much on linked data (as Data Science does). This leads to some privacy problems, since the most popular procedures (such as anonymization) tend to escape the most recent techniques of linked data (in the case of personal data or sensitive information, for example). Big Data can be generated and processed exhaustively and at high speed, practically in real time, which is why they offer the possibility of re-identifying and representing those people registered in anonymized databases. This would be an epistemic injustice (of the testimonial type), since people must define to what extent they declare to the social environment the experiences of their identity, including their personal data.

Data has attributes that make up its meaning in a given context. The representation of these attributes through an effective description of metadata schemas encourages their linkage in accordance with the attributes they manifest. Therefore, data linking is a process that involves the analysis of its attributes, which will allow establishing links between data with similar attributes. The dissemination of this data requires that we reconcile the principles of privacy, protection and intellectual property, in relation to the general treatment of all the information that accompanies the alphanumeric data and that is transmitted through a documentary range.

2. The ethical challenges for Information Sciences

In many academic and administrative areas, computers, servers and information systems based on a relational and business intelligence approach are available, but they do not have the technical characteristics for their use in Data Science or Big Data tasks. Faced with this situation, it could be thought that a professional challenge for librarians, archivists and documentalists would be to acquire expertise in these topics to extract knowledge from the information, no matter how complex and voluminous it may be (Hu and Zhang, 2018). While the above is technical training that can be important, it should be remembered that the professional relevance of librarians is related "to promote the enhancement and preservation of documents and information". This stems from the development of its ethical framework (its notion of *informative value*) and its epistemological framework (the definition of its object of study):

[...] in many ethical codes for librarians and other library employees adopted by national library or librarians' associations or implemented by government agencies, 'informational entities' are considered to have a moral value and deserve respect [...] Indeed, even ideal, intangible, or intellectual objects are acknowledged to have a minimal degree of moral value, no matter how humble, and so are entitled to some respect (Floridi, 2013, p. 123).

One could also think that the institutional challenges of companies, universities and scientific networks would be to have access to high-performance clusters, specialized software and other computing resources for the development of projects, within an efficient, quality and relevance scheme (Hiriyannaiah et al., 2020). However, the most decisive challenges and resolutions go beyond these practical complications.

For example, a general problem would be: In what way do the dynamics of Big Data contribute to position the large information consortiums more? Today there is a large and intense circulation of information flows, but much of the content comes from a few information sources:

[...] information is no longer informative but deformative, and communication is no longer communicative but cumulative [...] Even the largest accumulation of information, Big Data, possesses very little knowledge. Big Data is used to find correlations [...] Correlation is the most primitive form of knowledge, being not even capable of ascertaining the relationship between cause and effect [...] thus nothing is *understood*. But knowledge is understanding. Hence Big Data renders thought superfluous (Han, 2018: 1–3).

While librarians can identify nodes and relationships in a huge swarm of data (which is often characterized by repetition and replication), their greatest expertise lies in content analysis, creating thematic categories suitable for that large amount of data, so that these can be deconstructed and critically weighted (for example, to identify the obsolescence or authenticity of the information). Precisely, the correlational character highlighted by Byung-Chul Han is an important aspect to understand the phenomenon of massive data sets (enormous amounts of data) in the information age:

Seeing "data" almost as an adjective, as a relational property (like being the youngest child in a family), makes apparent why what counts as data changes over time: as the frontier between nature and knowledge evolves, so do the data that inhabit this moving frontier [...] How are objects related to data? We say, without thinking, that we "collect" or "assemble" data, as if they were shells on the beach or a drawer full of random Lego pieces (Strasser and Edwards, 2017: 330–331).

In this paper it is stated that the challenges of Big Data go beyond the individual spectrum of responsibility of a professional in Information Sciences (specifically, due to the negative social consequences), so that the changes brought about by massive data sets are essentially problems of a group ethics, so they require approaches from the theoretical postulates of these disciplines. In addition to this, the moral challenges in dealing with Big Data are usually approached from *applied ethics* (such as information ethics), but in this article it will be approached as a problem of *metaethics* and *normative ethics* (as a foundation for its *application* in professional codes), and also from some ideas of digital hermeneutics and the philosophy of technology. The intention is "pushing beyond legal concerns to address often messy ethical dilemmas" (Hesse et al., 2019: 563).

As a philosophical discipline, it can be said that ethics, or moral philosophy, is divided into three parts: normative ethics, applied ethics and metaethics. Basically *normative ethics* establish the conditions to follow a desirable life (answers to the question "What would be a correct action?"), especially through our obligations to others. With the normative language of ethics, some actions are not only described but also evaluated, and a change in behavior is indicated. The point is that this prescriptive power must come from some principle (Fraser, 2014). On the other hand, *applied ethics* studies specific problems in everyday life, difficult situations we find ourselves with, moral dilemmas, such as in bioethics, public service, politics, business, or social media. *Metaethics* investigates more theoretical, more abstract problems. Within moral philosophy, metaethics is the study of the meaning of moral terms, but above all on the nature of justification of fundamental moral principles (Miller, 2013).

The task of metaethics is not to solve the problems we encounter in our daily lives (such as establishing how we should live or the way in

which we should act in certain situations), but to give philosophical explanations on various questions about that part of our lives. Metaethics tries to answer some questions such as the existence of moral knowledge and moral facts, and, if there are any, about the possibility of the existence of moral truths. These truths would be those that we should all agree on. But metaethics not only answers questions, but also raises questions such as: Is ethics relative?, is it subjective?, or are there universal principles that apply to all cultures (that is, to all human beings regardless of the circumstances)?, and, therefore, could we talk about a group ethic? (Van Roojen, 2015).

A very important philosophical problem has been to verify if these three areas of ethics (applied ethics, normative ethics and metaethics) are logically and necessarily linked or if they are independent of each other. This means that the results in one of these areas would have to lead to results in another, and from there we would have to apply what was elucidated in those previous phases in order to solve specific moral problems. For example, if to solve certain problems of a normative ethics we would have to have solved some theoretical problems through a metaethical analysis. If a group of people think that there are no moral truths, or if this same group resorts to a relativistic ethics, then this will influence their resolutions on a problem (that is, what are they going to recommend to people to do in certain situations).

For some authors, there is a relationship between moral judgments, conative states and our mental states, but this in no way implies that moral judgments can be reduced to a deterministic component. Some positions assume that moral properties do not exist independently of human opinions and responses (Tiefensee, 2016). Despite this, the results offered by ethics would have to be at least consistent with those of sciences, since a continuity between the methods of both efforts is presupposed (in this case, the metaethical status of moral judgments based on problems posed by the Big Data and its applications).

3. Approaches from metaethics

In the case of Information Sciences, it has been necessary for their *normative ethical* positions to also be articulated with a *metaethics*, or rather, to develop an *informative metaethics*, where librarians, archivists and documentary makers explore the moral nature of their judgments and principles, and that these base what is prescribed in normative ethics, and that, in turn, solve specific problems in applied ethics. In the case of moral challenges in the handling of Big Data, a recurring problem is that these are usually approached from applied ethics, such as information ethics (which is actually a "transversal applied ethics"). Of course, the data pose challenges in specific settings, such as Biomedicine (e.g. informed consent, privacy and data protection, intellectual property in data set analysis, data classification criteria, fiduciary relationships, distinction in "academic" and "commercial" uses of data, data access rights and so forth) (Mittelstadt and Floridi, 2016). In this paper, though, moral problems in the Big Data era are addressed as a problem of metaethics and normative ethics.

Digital ethics (and its link with digital rights) could reinforce, in the first place, the link between *normative ethics* of Information Sciences with *applied ethics* (as happened with information ethics), but it would be necessary to constitute also a *metaethics* through the reflection of digitality:

> The shift from *information* ethics to *digital* ethics highlights the need to concentrate not only on what is being handled as the true invariant of our concerns but also on the general environment (infosphere), the technologies and sciences involved, the corresponding practices and structures (e.g. in business and governance), and the overall impact of the digital world broadly construed (Floridi et al., 2019a: 11).

The rapid deployment of digital technologies and their adoption by society has modified our relationships with ourselves, with each other and with our environment. As a result, our individual and social well-being is now intimately connected to the state of our informational environment —and to the digital technologies that mediate our interaction with it— raising pressing ethical questions about the impact of technologies on our well-being that must be approached.

The subject of metaethics has an initial challenge, which is to discover and analyze why normative ethics is the main criterion on the basis of which the practical and moral value of professional actions in Information Sciences is judged. Traditionally, the claim that moral values are "objective" is contrasted with the claim that they are "subjective". This problem (about whether values are subjective or objective) is generally taken as the metaphysical problem of locating the status of morality in the world. The fact that a metaethical theory defends one or another thesis qualified it as objectivist or subjectivist, respectively. By this, metaethicists refer to the question of how our moral behavior develops (Tiefensee, 2016). An example of this is honesty in privacy matters. One way of posing the question about its status is if we value it because our subjective attitudes lead us to think that it is valuable or is it because we recognize that it is valuable independently of them (because personal data represents someone's individual identity).

A person with a subjectivist approach would defend this first option because for him the moral properties (such as "valuable" or "obligatory") are simply the result of other properties or psychological relationships, such as the desires, emotions or approvals of others. Honesty would be valuable because we approve and desire it, since our emotions lead us to think so. It is important to note that although psychological properties are characteristically individual, a subjectivist does not necessarily have to commit that morality is individual, as it can be the result of the attitudes of a group of people, such as a community. On the other hand, a person with an objectivist approach would take the second option because this individual would defend that moral properties do not depend on what people consider (Van Roojen, 2015). For this person, honesty is valuable in itself, since moral properties do not depend on any other property. In addition, it can be said that even if the moral properties are objective, we would have to know them, in some way, through our subjective psychological conditions, and statements such as "stealing personal data is morally wrong" would depend on them. Considerations like these could lead us to think that the initial characterization is problematic, since there are cases in which we cannot distinguish if a statement is about something "subjective" or "objective" (Miller, 2013).

In the following section it will be develop in greater detail the statement that Big Data challenges go beyond the individual spectrum of responsibility of an Information Science professional (so the changes that Big Data have brought about are essentially group ethical problems), so they require approaches and reformulations from the theoretical postulates of these disciplines (for example, from their epistemological commitments). This also implies that ethical approaches within Archival Science, Library Science and Documentation must articulate, even more, an axiological development with the deontological formulation of its principles, towards an ethical framework with a view to introducing professionals to identify *consistent moral principles*, so they can participate in creation of new guidelines for conduct (Morán-Reyes, 2019).

As it happened with these disciplines, when approaching information ethics, which: "provides a general frame for moral evaluation, not a list of commandments or detailed prescriptions (compare this to the similar complaint of 'emptiness' made against deontological approaches)" (Floridi, 2013, p. 125). Despite being a problem of group ethics, the task of asking how we contribute (from our individuality, not only as professionals) to solving or accentuating a global challenge should not be avoided (Taylor, 2016), since by contributing minimally (in one sense or another) we have some responsibility and correlative duties: "in such a way that you will be ultimately morally responsible for whatever you do choose. Even if you believe that determinism is true, and that you will in 5 min time be able to look back and say that what you did was determined, this does not seem to undermine your sense of the absoluteness and inescapability of your freedom, and of your moral responsibility for your choice" (Strawson, 1994, p. 10).

Given this scenario, the deepest ethical problems that arise for Information Sciences (with daily use in different areas of Big Data) are not about whether we trust librarians to carry out tasks in the information organization area in an impartial way or about the obligation of archivists to educate themselves technologically in the area of Data Science. Some of the most important problems are related to the concept of *moral responsibility*, especially from a metaethical perspective, and "this may require rethinking what constitutes the meaning of ethical research, and who and what are responsible" (Hesse et al., 2019: 577). Moral responsibility is outlined by an agent's sense of his own freedom, within his experience of decision (including the socio-cultural conditions surrounding his acts of choice). This is important, especially because some practices of attribution of responsibility in the professional field sometimes seem to resort to desiderative criteria and in others they seek absolution or dilution of responsibility, without further argumentative justification: "the specification of their moral responsibilities will be effective [...] only insofar as it will rest on an ethical framework able to reconcile the different ethical views and stakeholders's interest" (Taddeo and Floridi, 2017, p. 15).

Certainly, a specific problem for Information Sciences would be "what is the source of credibility attribution for a librarian or an archivist (as an epistemic agents)?", taking into account that an agent can be understood to be morally responsible if:

a) His actions demonstrate his evaluation of what you consider valuable in life (that is, if you properly value other people through your actions); or,

b) Has the *moral agency capacity* to avoid acting in a certain way.

In other words, your *moral responsibility* is related to your freedom of choice:

> The problem for moral responsibility begins when we recognize that the agent's freedom is a necessary condition for the agent being morally responsible for the act [...] In short, if I am to meaningfully blame you for spilling the milk, I have to assume that you were free to take better care when pouring it. If your spilling of the milk was no more than a natural process playing itself out through your limbs, then it would make no more sense to blame you for the spill than it would be to blame a river-bank for bursting (Cowley, 2014, p. 14).

In order to conceive a less confusing idea of *moral responsibility*, it should be considered that this type of responsibility "begins when an observer is inclined to experience and express reactive attitudes towards the agent for some aspect of the agent's relationship with his environment, namely what he did, intentionally or unintentionally, or failed to do" (Cowley, 2014, p. 3).

On the other hand, the attribution of credibility on the responsibility of a professional usually seems something positive (to be worthy of social trust), although this is not always the case. In some situations the incorrect assignment of credibility towards a professional (for example, a doctor) can turn into an epistemic injustice (either excessive or insufficient credibility) (Fricker, 2007: 18). Although it remains to be defined what conditions must have a correct credibility assignment. It is clear that an incorrect assignment of credibility towards a professional exceeds the dimension of responsibility of that professional (or falls short). But what conditions must a correct credibility assignment have? It could be argued that an attribution of credibility is correct if and only if it is proportional to the degree of reliability of the agent, but this statement would face the problem of sustaining that a necessary condition for the digital ecosystem to function must guarantee reciprocity in credibility (above all because, it is thought, that in some issues, some agents have greater credibility than others, because they are experts).

It is true that, in some cases, reliability plays an important role. But it must be pointed out that a *morally responsible agent* does not depend on questions of reliability, especially if an ethical challenge goes beyond the

spectrum of responsibility of an individual. The trust that an epistemic agent grants towards information that it consults depends on the credibility of the informant source. In this sense, trust is often understood as a fiduciary relationship in which one of the parties decides to grant credibility to another to perform a certain action, based on a reliability assessment. This is defined as the set of beliefs that an agent has about the capabilities and performance of a potential agent, and the probabilities that they assign to those beliefs. But trust is not a source of epistemic justification, nor is it a source of attribution of responsibility (although it may raise additional questions about the justification for someone's decision to grant their trust, for example on the basis of reputation).

4. Approaches from normative ethics

Information Sciences must articulate their axiological approaches of their discipline with the deontological propositions of their codes of ethics (with the aim that they can formulate, through them, *consistent moral principles*). The relevant question here is whether a code of ethics can help us to continue developing those moral capacities and to become morally virtuous agents. A code can help by giving us certain guidelines to do so, but its effect will be limited; it should not be expected that a code of ethics will radically influence the moral decisions of people. In any case, it is something that has to be part of a more comprehensive strategy.

The content of many codes, in general, tends to be more focused on issues related to the interest of organizations over the interest of society: "The collection of data on citizens through digital portals is viewed by organizations as an opportunity to create value, leverage competitive advantage, and maximize productivity and efficiencies in service and product delivery" (Jurkiewicz, 2018: 46). The main themes of many library codes have to do with conflict of interest and use of privileged information, that is, its focus is concentrated on issues related to the protection of institutions, and not on more general aspects of corporate responsibility (more focused on the institution's relations with society. If the content of a code is more focused on the protection of the institution, then it will have gaps about many other aspects of the conduct of its members, and this will not help decision-making based on the code. In fact, on many occasions, a code of ethics is not exactly a guide for professionals to adopt values. Rather, these codes are a tool of dissemination that expresses the values and postulates that underlie disciplinary education.

Precisely, people often wonder how effective they are in guiding individual or collective decisions. Some of the people who are charged with monitoring compliance with codes of ethics, such as those who participate in ethics committees, complain that the codes suffer from two problems:

a) Have loopholes or normative holes.

b) Tend towards overregulation.

In the first case, the codes speak of some values, but leave out many others; they include rules about certain situations, but not about others; they do not have clear rules for particular cases or specific sanctions, among other points. How to sanction a conduct that seems wrong to us if the code of ethics does not say anything about it? If the cases multiply, then it seems that the code is left over, because finally we will end up acting at discretion, according to the criteria and the moral principles (or prejudices) of those who decide. Given these shortcomings, many consider that the code should be much more specific in its values and rules. But this gives rise to the reverse problem: the presence of too many rules.

In this second case, the overregulation occurs mainly in government contexts. Now, as the possible situations of inappropriate behavior are literally infinite, then it would seem to follow, in principle, that the rules can be multiplied until the code is bulky enough to become difficult to

use and even unmanageable, thus moving the code away from the everyday use by people and thereby loses some of its appeal. In this situation, the presence of an ethical expert (an ethicist) may be required to help apply the code, as in the case of law with lawyers, who are experts in the knowledge and application of legal codes).

It is common to think that only people with this training (to correctly apply values and moral principles in everyday life) would be able to resolve moral dilemmas in the face of which the code does not say anything, or that they will be able to make a reasonable decision in those cases in which there are regulatory gaps. For example: "bona fide organisations must have a Data Protection Officer, who is overlooking the responsible processing and management of personal data and, at least in certain cases, the equivalent of an Ethics Advisory Board or at least access to such, which may be an external Ethics Committee" (Floridi et al., 2019b: 366). However, in the face of these two problems (that of loopholes and that of over-regulation) the solution seems to lie in finding a middle ground in a code that considers the most common cases of lack of ethics within the institution, without attempting to regulate all of them. The question is if that is possible and if the codes that are currently being generated or those that are already in use manage to find that right medium, and if that point does not coincide with common sense.

From a certain point of view, there are reasons to think that, due to their own characteristics, codes do not help to resolve moral dilemmas (understanding by "moral dilemma" the case in which two moral values conflict, as when we are subject to two obligations morals that we cannot meet at the same time, breaching any of them) (Aberbach, 2021). You can think of countless cases in which moral values can conflict with one another. What should I do, for example, if I am in a situation where the protection of personal data conflicts with its misuse to generate informal job sources for people who need to work? What if the value of impartiality conflicts with that of obedience to an institution? What happens if my loyalty to the institution, on the one hand, and my loyalty to society, on the other hand, conflict (e.g. in cases where irregularities that occur within the institution and that affect the society)? And what if I can't satisfy both values at the same time, as is often the case with genuine moral dilemmas? What can the code of ethics tell me about these dilemmas? This type of conflict of values is not exclusive to codes of ethics, values conflict in different areas and it is not always easy to find the answer. In the judicial context, it is common for rights to conflict, and judges must decide at their discretion on issues that are not specified in the laws (we must remember the case of Cambridge Analytica). At other times, values such as law and justice come into conflict and then what should be done, apply the law or do justice? These are issues that are not usually specified in codes of ethics either.

To remedy the problem of moral dilemmas, it could be proposed to establish a hierarchy of values, which would tell us, in order of priority, which are the most important values for the institution and which are secondary. Nonetheless, this can also be problematic. Consider, for example, a code modeled after a utilitarian theory, according to which the promotion of the interests of the majority of the people involved is what has priority for the institution (Zuk, 2015). But what if, to promote the interests of the majority, I have to sacrifice those of a minority? Is that ethically correct? Furthermore, the promotion of this priority value would even seem to justify the performance of some acts that, from the point of view of everyday morality, could be incorrect. Someone could, for example, lie or even betray their own principles if this promote the interest of the majority.

To many it will seem that it is not correct or convenient for codes of ethics to hierarchize values, especially because, in any case, values will continue to conflict with each other, and the answers that a hierarchical code gives us will not always satisfy our everyday moral intuitions and will sometimes even collide with them. Moral values (but also other types of values) have more or less weight depending on the context in which they are involved and it is a question that agents have to judge: determining when a value should have more weight and when not. This is not something that a code of ethics should establish in advance for all possible cases. But, then, it seems that it would be the moral agents, in particular contexts, who will end up deciding, and not always by referring to the code, but by appealing to their own moral consciences.

If we think that people should always act according to moral reasons —especially when these are opposed to economic, political or even reasons of mere personal convenience—, then we should be able to justify that. But we cannot hope that the code will help us to resolve these conflicts (and even less to resolve the question about the weight that morality should have in the face of other types of reasons, that is not its function). All this is something that we have to do independently, that is, without resorting to code of ethics.

It is ingenuous to expect that a code of ethics will help us resolve conflicts of values (either the moral values established within the code or with values external to it) but this is the type of conflict that the person who uses the code will face (Morán-Reyes, 2019). Although, when resorting to the code, it is likely that this person, surely concerned about the resolution of a certain ethical conflict, in many cases ends up solving it in the same way as they would without the code: when they do not do so by imposing a personal point of view already predetermined, in the best of cases will seek to balance the different aspects of the situation, giving more importance to some aspects over others, evaluating situation by situation, until a satisfactory result is achieved.

It could be answered that the code of ethics fulfills a technical and not a moral function in deliberation. It will be enough that the deliberation refers to the code to accept a point, and it will not be necessary to justify a certain point every time from scratch. If the code indicates that conflicts of interest must be avoided, and that one way is to make them public in advance, then it is not necessary to argue in favor of this in the deliberation each time it has to be applied, but it is enough to refer to the code. This can make deliberation more efficient. In fact, although the code of ethics fulfills a technical function, it is necessary that the ethics of the Information Sciences are not reduced to their respective codes, but that the code is only an expression of something more profound, and that the norms and rules represent an axiological reflection and a metaethical analysis. But there is not only responsibility in librarians, archivists and documentalists; in this case, some technological challenges of designers and users must be addressed, which means "we need to address professional responsibilities and deontology of those actors involved in coding, maintaining, and updating the Internet's infrastructure, including its applications and platforms" (Floridi et *al*., 2019a: 13).

5. Technological challenges in the era of Big Data: hermeneutical approach

It is true that the production of massive data is a daily reality today, and that Data Science technologies make it possible for its management and exploitation: "In particular, machine learning and deep neural networks have been hugely successful in recent years. While not a fundamentally novel technology, recent successes of machine learning have been made possible by the availability of large data sets for training and testing purposes and the affordability and availability of large amounts of computing power" (Ryan and Stahl, 2021: 63). For example, in the case of the use of mobile applications (based on Artificial Intelligence algorithms) it is possible that they allow:

- Gathering strategic information from academic communities, including in real time.
- The use of smart devices.
- Analyzing clickstream with Big Data techniques, to address previously unsolvable problems in relation to learning or decision-making.

Some of these applications and devices compete with some services provided by academic libraries and with their information search processes. The extraction of knowledge from the data that is generated every day in each of the academic and administrative areas of the universities is strategic for designing new personalized learning services. The emergence

of new and better analytical tools, and the appearance of innovative low-cost services in the cloud are part of a computing culture that is not closely related to the processes in institutions such as libraries or archives. The data analysis carried out from Data Science involves the development of complex algorithms and the identification of business opportunities with statistical techniques. The analysis carried out from the Information Sciences turns out to be more qualitative and less syntactivistic; Library Science is more of an "Applied Philosophy of Information", namely: "A Philosophy of Information approach to the foundations of Library of Information Science may be expected to work on the ontology of its 'objects', on a substantial theory of information dynamics, and on an ethical approach to the domain of information" (Floridi, 2004, p. 659).

Until a few years ago, the main limitations to use Data Science technologies and Big Data techniques and applications on a regular basis were:

- The costs and access facilities to the platforms and technological resources required.
- The complexity of the software tools.
- The shortage of sufficient specialized personnel.
- The lack of training programs.

In recent times, this trend about practical problems has begun to change thanks to:

- The significant reduction of costs in the hardware and software required to do Data Science.
- The beginning of new and better analytical tools.
- The emergence of innovative low-cost cloud services.

Nevertheless, beyond the pragmatic benefits that the design and application of disruptive technologies entail, in this scenario, philosophical problems of various kinds appear, such as on the moral plane, for example that "data mining is hardly perceived as an 'intervention' in the life or body of the individual" (Kuyumdzhieva, 2019: 55). This could bring us closer to the "designer fallacy", in which it is believed that it is the designer who has control over the functionality and effects of a technological device, without taking into account the interrelationships with material reality, the will of the users, cultural contexts, and so on (Ihde, 2002, 2008). In this vein, "Crucial challenges include the moral responsibility and accountability of both designers and data scientists with respect to unforeseen and undesired consequences and missed opportunities; the ethical design and auditing of algorithms; and the assessment of potential undesirable outcomes (e.g., discrimination or the promotion of anti-social content)" (Floridi et al., 2019a: 12).

One of these challenges is related to the issue of whether or not it is ethical to design technologies that have the ability to interpret reality by themselves, without the participation of human beings (which disrupts not only self-understanding, but also the autonomy of the humanity). Linked to this, another relevant problem is that of understanding and interpreting the data. Keep in mind that "one accumulates information and data, yet does not attain *knowledge* [...] Big Data is used to find correlations [...] Correlation is the most primitive form of knowledge, being not even capable of ascertaining the relationship between cause and effect [...] But knowledge is understanding" (Han, 2018: 2–3). This leads to an important topic: the "active/passive" role of human beings as preponderant agents in the interpretation of data.

There are ways through which the development of technology and, at the same time, that of human interpretive autonomy can be balanced (without technology displacing human agency, and without imposing restrictions on the possibility of future technological applications). The way proposed here is that of hermeneutics, as a possibility for people to investigate certain experiences and redefine their actions in the world, and so that they can reconstruct their identity (since the question of hermeneutics is in the interpretation of the way of existing).

In *traditional hermeneutics*, the activity of collecting the world's data and its subsequent interpretation fell to the agency of the human being. In the stage of *material hermeneutics* —faced with the expansion of science— the human being required technological devices to collect the data, but the human being still remained as the interpreter of these data (although conditioned by their technological tools). These technologies appeared as a kind of "extenders" of human capacities (the telescope or the microscope on the human eye), but, in the end, the human being was the one who interpreted these data (with his cognitive faculties). In the current era, with the use of artificial intelligence algorithms, many everyday technologies are responsible for collecting data from the environment, but it is they themselves who interpret them (even better than humans). One example is biometric analysis technologies (even in mobile phones, such as fingerprint reading) that carry out the entire process, from scanning to the interpretation of the patterns itself (Hongladarom, 2021). With digital hermeneutics, the possibility of a new situation appears, in which, thanks to algorithms, smart devices share with humanity the task of interpreting, analyzing and making sense of the data from the environment: "digital hermeneutics can aspire to universality as far as it aligns itself with digital traces and data, the presence of which is henceforth as great as the human and social reality [...] can be actually defined as a general problematization of the supposed homology between (social) reality and its digital representations" (Romele, 2020, p. 73).

With all of the above (which is an undeniable reality), these technologies must have an operational limit, according to which they should never replace the interpretive role of the human being. They may have to share with human intelligence those processes of interpretation about the phenomena of the world, but never replace it completely, not even in certain aspects (although their presence reconfigures, in fact, some central ideas for current societies, such as that of the privacy or identity). Of course, there are also epistemological implications that result from the idea that smart technologies are artifacts that correct our cognitive myopia (such as "augmented reality", which is no longer a fictitious reproduction of "what is real", but a more direct access to the real world).

A *digital hermeneutics* does not consider or defend a substitution of this type (not even partial), as *computational hermeneutics* has proposed. As mentioned above, not only on the subject of design, but on that of moral leadership, human intelligence must be behind any digital application, to prevent these technologies from being used for purposes that go against life itself:

> The digital revolution provides huge opportunities to improve private and public life, and our environments, from health care to smart cities and global warming. Unfortunately, such opportunities come with significant ethical challenges. In particular, the extensive use of increasingly more data —often personal, if not sensitive (Big Data)— the growing reliance on algorithms to analyse them in order to shape choices and to make decisions (including machine learning, AI, and robotics), and the gradual reduction of human involvement or oversight over many automatic processes, pose pressing questions about fairness, responsibility, and respect of human rights (Floridi et al., 2019a: 9).

Excessive reliance on algorithmic criteria or Big Data techniques is dangerous, especially if one takes into account that there are platforms that are designed to deceive and commit fraud, or when these technologies are implemented for the purpose of social surveillance, even with explicit user permission: "As a condition of accessing digital media, individuals implicitly agree to allow the collection of data they generate while on a site, as well as content on the devices used to access the sites, unless steps are taken to limit such access" (Jurkiewicz, 2018: 46). Although —it must be said— they can also be used to strengthen the scope of respect for data protection laws (Floridi et al., 2019b).

The task of many disciplines (including Information Sciences) is twofold: to think about *digitality* and, at the same time, to be approached

by it. This means that a digital hermeneutic must cover any field that digitality has disrupted and, as a result of this redefinition, must deal with how the digital code is structured, but above all how it is interpreted, in contemporary societies (and not only certain common digital manifestations). In this sense, such an effort would consist of trying to understand the problems surrounding the current situation in which human existence and thought are located (such as the question of how complex hybrid life systems are built, or how the constitution of personal identity is affected by digitality).

For this reason, *digitality* should not be understood only from a mere physicalist criterion (as "physical things" or "fixed objects"), but as an opening to the world and self-knowledge), or as a *digital event* (in the that the expressiveness of experiences in digital media flows, and there is a consequent interpersonal exchange of intentions), in addition to taking into account those social conditions that configure their potential for representability (Ricœur, 1994).

6. Conclusions

Although there are practical challenges that are important because they allow the operation of an institution to be dynamic, there are other challenges that disrupt the professional practices of librarians, archivists, and documentary makers. An operational challenge would be for these professionals to be trained as specialists who support academic, government and business communities in the development of their Data Science and Big Data projects (Taylor, 2016), and who advise the implementation of these technologies in other local, regional or national instances (Hu and Zhang, 2018).

Big Data not only exceeds the capabilities of currently available tools, but also proposes the creation of services based on the search for information (which compete and displace some processes of institutions such as academic libraries). Nonetheless, as disciplines, this does not mean that Data Science displaces or represents an "epistemological obstacle" (in its Bachelardian sense) for Information Sciences, but theoretical approaches that deeply contemplate the phenomena derived from Big Data are required. Although the principles of Data Science can offer a technical training that can be important for Library Science, it should not be forgotten that the expertise of professional librarians is in the *knowledge organization*, and not in the *knowledge extraction*.

However, the main problems are related to the concept of moral responsibility, especially from a metaethical perspective, in line with the reflection of the implementation of technology with respect to human autonomy. The main challenges for Information Sciences in the era of Big Data are primarily of an ethical nature. But not about the reliability or impartiality of its disciplinary actions, but on how the dimension of moral responsibility far exceeds the individual spectrum of its professionals. Therefore, it is a problem of group ethics, so it requires approaches such as the articulation between its axiological reflection, the metaethical analysis and the deontological prescriptions of its codes of ethics (with the objective that they can formulate, through them, consistent moral principles).

There are reasons to be skeptical of these codes, such as regulatory loopholes and overregulation, but also their inability to help resolve moral dilemmas. Those who formulate codes of ethics would have to take into account these limitations, especially those that have to do with dilemmas, because only a more comprehensive strategy in which other types of tools are contemplated for the moral development of the members of the institution can help them deal with the dilemmas that are often not addressed by a code of ethics.

Only a more comprehensive strategy could help the moral development of professionals subject to a code; a strategy that includes not only the formulation of a code, but also aspects such as training courses or moral development, the figure of a consultant on ethics issues, the possibility of anonymously reporting irregularities, the creation of autonomous ethics committees, among others. Nonetheless, empirical studies should be carried out in the future to see the impact that these types of

more comprehensive strategies have on the moral behavior of people in the institutions where they are implemented. The issue of combating the lack of ethics within an institution and the moral development of its members is not simple, and requires more complex responses than simply formulating a code of ethics. We must be aware that we are faced here with a very complicated and much broader issue (to an extensive social scope), which takes us beyond the formulation of codes of ethics and the policies that a given institution can implement to improve the moral conduct of its members. This would have to decant into a dynamic of "elaboration, maintenance, and refinement of factual narratives: personal identity, ordinary experience, community ethos, family values, scientific theories, common-sense constituting beliefs, and so forth" (Floridi, 2011: 7).

Another reason to warn that the ethical challenges in the era of Big Data are of a group ethic, and not for an individual spectrum, is that the technologies used to process them turn out to be more invasive. The moral problems derived from the use of information and communication technologies in the professional field of librarians and archivists could be approached in the individual spectrum from the approach of an applied ethics (information ethics), because these are operational problems. But the use of technologies with artificial intelligence brings to light problems that jeopardize the interpretive autonomy of human beings. It is recommended that the development of the next disruptive technologies in the area of Data Science also be the object of study of a digital hermeneutics, to ensure that technological progress continues without undermining the cognitive capacity of humanity; so that both developments go hand in hand (that of technology and that of human knowledge), and that the era of Big Data is envisioned as "a new worldview, a new imaginary, and dominant symbolism, with tangible effects in the way in which individuals make and practice science and undertake the knowledge of their milieu" (Romele, 2020, p. 60).

Declarations

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Data availability statement

No data was used for the research described in the article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Aberbach, D., 2021. The Environment and Literature of Moral Dilemmas: from Adam to Michael K. Routledge, NY.
Cowley, C., 2014. Moral Responsibility. Routledge, London.
Floridi, L., 2004. LIS as applied philosophy of information: a reappraisal. Libr. Tr. 52 (3), 658–665.
Floridi, L., 2011. The Philosophy of Information. Oxfrd University Press, NY.
Floridi, L., 2013. The Ethics of Information. Oxford University Press, NY.

Floridi, L., Cath, C., Taddeo, M., 2019a. Digital ethics: its nature and scope. In: Öhman, C., Watson, D. (Eds.), The 2018 Yearbook of the Digital Ethics Lab. Springer Nature, Cham, Switzerland, pp. 9–17.

Floridi, L., et al., 2019b. Key ethical challenges in the European medical information framework. Minds Mach. J. Artif. Intell. Philos. Cognit. Sci. 29 (3), 355–371.

Fraser, B., 2014. Moral error theories and folk metaethics. Philos. Psychol. 27 (6), 789–806.

Fricker, M., 2007. Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press, NY.

Han, B.-C., 2018. The Expulsion of the Other: Society, Perception and Communication Today. Polity Press, Cambridge, UK.

Hesse, A., Glenna, L., Hinrichs, C., Chiles, R., Sachs, C., 2019. Qualitative research ethics in the Big Data era. Am. Behav. Sci. 63 (5), 560–583.

Hiriyannaiah, S., Akanksh, B.S., Koushik, A.S., Siddesh, G.M., Srinivasa, K.G., 2020. Deep learning for multimedia data in Internet of Things. In: Tanwar, S., Tyagi, S., Kumar, N. (Eds.), Multimedia Big Data Computing for Intenet of Things: Concepts, Paradigms and Solutions. Springer Nature, Singapore, pp. 101–129.

Holloway, K., 2020. Big Data and learning analytics in higher education: legal and ethical considerations. J. Electron. Resour. Librariansh. 32 (4), 276–285.

Hongladarom, S., 2021. Machine hermeneutics, postphenomenology, and facial recognition technology. AI Soc. J. Knowl. Cult. Commun.

Hu, J., Zhang, Y., 2018. Measuring the interdisciplinarity of Big Data research: a longitudinal study. Online Inf. Rev. 42 (5), 681–696.

Ihde, D., 2002. Bodies in Technology. University of Minnesota Press, Minneapolis.

Jurkiewicz, C.L., 2018. Big Data, big concerns: ethics in the digital age. Public Intregr. J. Am. Soc. Public Adm. 20 (suppl. 1), 46–59.

Kuyumdzhieva, A., 2019. Ethics challenges in the digital era: focus on medical research. In: Koporc, Z. (Ed.), Ethics and Integrity in Health and Life Sciences Research. Emerald Publishing Limited, Bingley, UK, pp. 45–62.

Miller, A., 2013. Contemporary Metaethics: an Introduction, second ed. Polity Press, Cambridge, UK.

Mittelstadt, B.D., Floridi, L., 2016. The ethics of Big Data: current and foreseeable issues in biomedical contexts. Sci. Eng. Eth. 22 (2), 303–341.

Morán-Reyes, A.A., 2019. Eticidad y filosofía de la información: Fundamentación ética para la bibliotecología. Alfagrama, Buenos Aires.

Ricœur, P., 1994. Oneself as Another. University of Chicago Press, IL.

Romele, A., 2020. Digital Hermeneutics: Philosophical Investigations in New media and Technologies. Routledge, NY.

Ryan, M., Stahl, B.C., 2021. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. J. Inf. Commun. Eth. Soc. 19 (1), 61–86.

Strasser, B.J., Edwards, P.N., 2017. Big Data is the answer... But what is the question? Osiris 32 (1), 328–345.

Strawson, G., 1994. The impossibility of moral responsibility. Philos. Stud. Int. J. Philos. Anal. Tradit. 75 (1-2), 5–24.

Taddeo, M., Floridi, L., 2017. The moral responsibilities of online service providers. In: Taddeo, M., Floridi, L. (Eds.), The Responsibilities of Online Service Providers. Springer, Cham, Switzerland, pp. 13–42.

Tang, Y., Xiong, J., Becerril-Arreola, R., Iyer, L., 2020. Ethics of blockchain: a framework of technology, applications, impacts, and research directions. Inf. Technol. People. 33 (2), 602–632.

Taylor, L., 2016. The ethics of Big Data as a public good: which public? Whose good? Philos. Trans. A Math. Phys. Eng. Sci. 374 (2083), 1–13.

Tiefensee, C., 2016. Inferentialist metaethics, bifurcations and ontological commitment. Philos. Stud. Int. J. Philos. Anal. Tradit. 173 (9), 2437–2459.

Van Roojen, M.S., 2015. Metaethics: A Contemporary Introduction. Routledge, NY.

Williams, B., 2012. Morality: an Introduction to Ethics, 13nd ed. Cambridge University Press, Cambridge, UK.

Zuk, P.D., 2015. A third version of constructivism: rethinking Spinoza's metaethics. Philos. Stud. Int. J. Philos. Anal. Tradit. 172 (10), 2565–2574.