Report on Artificial Intelligence[*]

Artificial Intelligence and data protection: Challenges and envisaged remedies

(11 June 2018 - Draft)

Alessandro Mantelero[**]

[*] This is a draft of the Report commissioned by the Council of Europe to the author. The aim of this draft is to provide a first outline of the Report to be discussed by the Consultative Committee of Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108) in the 36th Plenary meeting. The document is therefore the expression of the personal viewpoint of the author and discusses the main topics that will be further elaborated the final version of the Report, according to the comments that will be formulated by the Plenary.

[**] Associate Professor of Private Law at the Polytechnic University of Turin, Department of Management and Production Engineering.

Summary

Part I – The state of the art

I.1 Introduction

Defining the field of research of this Report is not an easy matter, since both data protection and AI have quite uncertain borders. On one hand, data-intensive technologies (including AI) represent a challenge to the application of some of the traditional principles of data protection, making them blurrier, less clear-cut or more difficult to apply [CoE 2017; Hildebrandt, 2016; Barocas & Nissenbaum, 2015; Citron & Pasquale, 2014; Mantelero, 2014; Rubinstein, 2013; Boyd & Crawford, 2012; Tene & Polonetsky, 2012]. On the other hand, AI is a broad field, whose boundaries are uncertain, encompassing a variety of approaches that attempt to emulate human cognitive skills [Villani, 2018, 4].

Data protection and AI are necessarily correlated. Leaving aside science fiction scenarios, the significant evolution of AI applications over recent years has some of its roots in the progressive datafication process [Mayer-Schönberger & Cukier, 2013, 78; Lycett, 2013]. Therefore, personal data are increasingly both the source and the target of AI applications (e.g. personal assistants, smart home devices etc.).

Against this background, different approaches are emerging in AI development, use and regulation. Regarding data protection regulation, the global framework offers different ways to safeguard fundamental rights and, in particular, the right to the protection of personal data. Europe's leading position in the field of data protection – recently also recognised in the ongoing debate in the US on digital propaganda – may lead to a prominent role for this region in addressing the regulatory challenge of AI development. In fact, AI is largely unregulated and often not grounded on fundamental rights, relying instead mainly on data processing.

The adoption of a European perspective may also mitigate the envisioned clash between a development of AI which is market- and technology-oriented with one which is inclusive. From the perspective of Convention 108 and, more in general, of the Council of Europe's approach to fundamental rights, although this contrast exists in practice, a solution can be provided by the regulatory framework and in the jurisprudence of the European Court of Human Rights.

In terms of policy, the foundational nature of fundamental rights has led the Parties of Convention 108 to enhance a development of technology grounded on these rights and not merely driven by market forces or high-tech companies. Moreover, the historical roots of European data protection urge policy makers to consider the potential negative consequences of data processing technologies.

This rights-based approach necessarily impacts on AI development, which should be consistent with the values expressed in Convention 108 and in the regulations of the Council of Europe. The Parties of the Convention should therefore actively drive AI developers towards a value-oriented design of products and services, in contrast with vague or overly optimistic views of AI development.

At the same time, governments should be the first to use AI in a way which is centred on safeguarding and promoting data protection and fundamental rights, thereby avoiding a growth of AI systems or technologies that may restrain individual and collective rights and freedoms.

For these reasons, it is important to extend European regulatory leadership in the field of data protection to a value-oriented regulation of AI [Villani, 2018, 7] based on the following three precepts:

- Value-based approach (encompassing social and ethical values)
- Risk assessment and management
- Participation

The Council of Europe's standpoint is broader than the EU borders and encompasses a broader variety of legal cultures and regulatory approaches. Despite this, common ground on the Council of Europe's legal framework, and Convention 108 itself, provide a uniform background in terms of common values.

The Council of Europe may be one of the best fora to combine attention to fundamental rights and flexibility in technology regulation, adopting a principle-based approach. The broader nature of principles makes it possible to give specific interpretations of them consistent with a changing world, where detailed legislative provisions seem unable to react quickly to socio-economic and technological changes.

Moreover, a principle-base regulation leaves room for the peculiarities which characterise each local context. This is even more relevant with regard to AI applications, which have a potential impact on contextual legal, ethical and social values [IEEE, 2016].

Of course, data protection per se does not cover all these issues, which require a broader viewpoint encompassing human rights and societal issues[1] [EDPS, 2018; Mantelero, 2018; Council of Europe, 2017]. However, data protection may have a boosting and complementary role in addressing these different aspects.

In this sense, the focus on individuals that characterises data protection, the awareness of the social consequences of data use and the link with personality rights may facilitate a data controller's broad approach, which goes from data protection towards fundamental rights and collective interests. Regarding its complementary role, data protection helps to disclose the way data are used and the purposes of processing, which represent a key element to better understand potential consequences on a variety of different rights and freedoms.

[1] See Consultative Committee of Convention for the protection of individuals with regard to automatic processing of personal data, 'Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data' (hereinafter Guidelines) adopted on 23 January 2017.

Finally, AI raises many different sector-specific issues concerning both the various AI fields of application (labour, justice administration, crime control, contract relationships, etc.) and the consequences of AI use (e.g. sustainability, environment impact, political impact etc.), which should of necessity be separately addressed. Given the focus of Convention 108, these issues are not discussed in this Report, which concerns the common core of all these applications, represented by data processing. In terms of potential impact, this analysis may therefore provide a contribution to the debate on the issues concerning both AI in general and its specific applications.

I.2 AI development

Over the years, many reports and scientific contributions have been published on AI and its evolution. Therefore, it is not necessary to recall the up-and-down trajectory of the scientific and social interest for AI technology that has characterised our society since the early studies in this field [McCulloch & Pitts; Turing, 1950] to the most recent contributions. Moreover, it is superfluous to mention the increasing array of AI applications and the results achieved through them.

Nevertheless, the historical perspective is important to properly understand the present and near future for AI. In this regard, two questions arise: why the policy debate of the last few years focuses on AI? Which kind of AI is reasonably expected in the next few years? The answers to these questions are crucial to address the AI regulation. In fact, it is necessary to contextualise technology development in the AI field, avoiding any confusion due to commercial or media narrative about AI.

First, AI is not a mere hype. As happened in the past for cloud computing, Big Data and IoT, there is a clear push from some vendors to magnify AI applicative scenarios and AI often becomes a buzzword used in contexts that do not properly concern this technology. However, in this attention to AI, there is a basis of truth concerning the peculiar technological environment that nowadays makes it possible to realise expectations that were merely envisaged in the past.

Over the last decade, the increasing availability of bandwidth for data transfer, data storage and computational resources – due to the new paradigm based on cloud computing – and the progressive datafication of large part of our life and environment have created a completely new context. This has led AI development to a breakthrough, enabling new forms of data management to extract further information and create new knowledge from data.

Big Data analytics and Machine Learning technologies[2] represent the most recent result of this development process [The Norwegian Data Protection Authority, 2018, 5]. The

[2] The difference between these two technologies can be summarised as follows: "patterns and connections. This is where AI can make a difference. While traditional analytical methods need to be programmed to find connections and links, AI learns from all the data it sees. Computer systems can therefore respond

concrete application of these technologies make it possible to envisage the kind of AI that is reasonably expected in the next few years and shows how we are still very far from the so-called General AI [Bostrom, 2016; Executive Office of the President, and National Science and Technology Council - Committee on Technology, 2016, p. 7; The Norwegian Data Protection Authority, 2018].

For these reasons, although "algorithms and artificial intelligence have come to represent new mythologies of our time" [CNIL, 2017], this report focuses on the existing and near future applications of AI, leaving aside challenging questions concerning human-like AI, in terms of machine liability and risks for humanity [Bostrom, 2016; Kurzweil, 2016]. In this sense, the Convention 108, both in the original text and in the modernised version, refers to "automated processing" or "automatic processing" and not to autonomous data processing systems, implicitly highlighting how autonomy is a key element of human beings [European Commission, 2018].

This brief summary of the state of the art clearly shows how AI is unavoidably based on data processing. Therefore, AI algorithms necessarily have an impact on personal data use and pose questions about the adequacy of the existing data protection regulations in addressing the issues that these new paradigms rise.

[Possible addition: brief description of the functioning of analytics and machine learning algorithms]

I.3 The adopted perspective

The major AI threats mainly concern disputable sets of values adopted by AI developers and users, where the latter encompass both consumers and decision-makers that use AI to support their decisions. In this sense, there is an emerging trend towards a technocratic and market-driven society, which pushes for personal data monetisation, forms of social control and "cheap & fast" solutions for decision-making on large (e.g. smart cities) and small (e.g. precision medicine) scale.

The expansion of this trend is challenging and progressively eroding individual self-determination, privacy-focused models, and mindful and cautions decision-making processes. Data bulimia, complexity of data processing and an extreme data-centred logic may undermine the democratic use of data, replacing individuals and collective bodies, as well as freedoms and self-determination, with a sort of dictatorship of data [O'Neil, 2017] set up by data scientists insensitive to societal issues.

continuously to new data and adjust their analyses without human intervention. Thus, AI helps to remove the technical barriers that traditional methods run into when analysing Big Data" [The Norwegian Data Protection Authority, 2018, 5].

Against this background, to prevent a negative outcome to prevail over the benefits of AI innovation [ITU, 2017], it is necessary to reaffirm and stress the centrality that the human being should occupy in the technology (and AI) development. To reach this goal, it is necessary to reaffirm the prevalence of fundamental rights over the logic of efficiency and profit.

In the light of this, the right to the protection of personal data may become the stepping stone to design a different data society, in which AI development is not driven by pure economic interests or dehumanising algorithmic efficiency.

To boost this fundamental rights-oriented paradigm a wide debate is necessity. We need to slow down the tension towards an extreme datafication of everything and to affirm the effective prevalence of induvial and collective rights. This means that governments and citizens should realise the risks of datafication and be aware of the potentially dangerous implications of data-driven solutions [Rouvroy, 2016].

As happened in the past for industrial and product development, risk awareness is not a barrier to innovation, but rather a driver. Innovation should be developed in a responsible manner, safeguarding fundamental rights by considering this safeguard the preeminent target.

This necessarily requires the development of assessment procedures, the adoption of participatory models and supervisory authorities. A human rights-oriented development of technology may therefore increase costs and force developers and business to delay the current time-to-market, given the necessary prior assessment of the impact of products and services on individual rights and society. Nevertheless, business and society are mature enough to consider responsibility towards individuals and society as the first goal in AI development.

Alternatively, if AI follows a different path – such as other previous technologies in their early stages (e.g. biotechnology) – it will risk being developed in an unregulated manner, which is merely driven by technological feasibility, market or political interests. These are all factors that do not guarantee per se a human right compliant development.

For these reasons, AI data-centric development should therefore be based on the principles of the Convention 108 as the foundations for the flourishing of digital society. The key elements of this approach are:

- Proportionality (the development of AI should be inspired by the proportionality principle, which means that efficiency should not prevail over individuals rights and freedom; individuals have the right not to be subordinated to automated AI systems; legislators can limit AI applications to safeguard individual and societal interests).
- Responsibility (which is not mere accountability, but it requires developers and decision-makers to act in a social responsible manner. It also entails the creation of specific bodies to support and monitor their actions)

- Risk management (accountable AI requires the assessment of the potential negative consequences of AI applications, as well as the adoption of adequate measures to exclude or mitigate these consequences)
- Participation (participatory models in risk assessment are essential to give voice to citizens. At the same time, citizens' participation must not diminish the accountability of decision-makers)
- Transparency (despite the current limitations affecting transparency of AI applications, a certain degree of transparency can contribute both to guarantee an effective citizens' participation and to properly assess the consequences of AI applications).

I.4 Existing framework and principles

The exiting regulatory framework applicable to AI and data processing is mainly grounded on the Convention 108, although other legal instruments concerning data protection (such as recommendations[3] and guidelines[4]) may also be relevant with regard to specific fields. In this context, the Guidelines on Big Data adopted by the Council of Europe [Council of Europe, 2017] represent the first attempt to address the issue concerning the use of data-intensive solutions for decision-making and are part of a broader wave of documents and resolutions adopted by several institution at European level on the impact of the use of algorithms on our society [Council of Europe-Committee of experts on internet intermediaries (MSI-NET), 2018; European Data Protection Supervisor - Ethics Advisory Group, 2018; European Parliament, 2017; European Union Agency for Fundamental Rights (FRA), 2018].

The scope of the Guidelines adopted on Big Data was "to contribute to the protection of data subjects regarding the processing of personal data in the Big Data context by spelling out the applicable data protection principles and corresponding practices, with a view to limiting the risks for data subjects' rights. These risks mainly concern the potential bias of data analysis, the underestimation of the legal, social and ethical implications of the use of Big Data for decision-making processes, and the marginalisation of an effective and informed involvement by individuals in these processes". In this regard, although focused on Big Data analytics, these Guidelines cover the main issues that characterise data-intensive and complicated applications for decision making. For this reason, the considerations expressed about the potential positive role of risk assessment procedures (encompassing ethical and societal issues),

[3] See, e.g. Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.

[4] See, e.g., Practical guide on the use of personal data in the police sector (2018); Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data (2017).

testing activities, data minimization, expert committees, precautionary approach[5] and freedom of the human decision-maker can be reaffirmed in the AI context.

Some of these remedies are further explained in this report (see below Part II). On the other hand, AI concrete applications urge the analysis of new issues (such as the role played by transparency and the range of values that should underpin AI applications) and suggest the adoption of new remedies (e.g. a broader data protection impact assessment, potential limitations to AI use). Finally, there are existing supervisory bodies (e.g. data protection Supervisory Authorities) whose role may be reconsidered in the light of the new challenges of AI applications and the variety of their potential consequences on society.

In this sense, AI – like Big Data[6] and in a analogous way[7] – represents a challenge for the application of some traditional principles of data processing[8] and may justify the research of new applicative solutions to safeguards personal information and fundamental rights.

I.5 Individuals' self-determination in data processing

Over the last years, privacy scholars have repeatedly emphasised the weakness surrounding data subject's consent in terms of self-determination. Long and technical notices about data processing, social and technical lock-ins, obscurity of interface design, and lack of data subject's awareness are some of the reasons of this weakness.

Moreover, AI-based profiling and hidden nudging practices challenge both the idea of freedom of choice based on contractual terms as well as the idea of data protection in terms of data subjects' control over their information. Finally, the complexity and obscurity that often characterise AI algorithms necessarily affect the chance of obtaining an actual informed consent.

[5] See also Commission - European Group on, Ethics in Science and, & New Technologies, 2018, 16 ("As the potential misuse of 'autonomous' technologies poses a major challenge, risk awareness and a precautionary approach are crucial").

[6] See Guidelines, Section II ("Given the nature of Big Data and its uses, the application of some of the traditional principles of data processing (e.g. the principle of data minimisation, purpose limitation, fairness and transparency, and free, specific and informed consent) may be challenging in this technological scenario").

[7] See also in this sense The Norwegian Data Protection Authority, 2018 ("This report elaborates on the legal opinions and the technologies described in the 2014 report «Big Data – data protection principles under pressure». In this report we will provide greater technical detail in describing artificial intelligence (AI), while also taking a closer look at four relevant AI challenges associated with the data protection principles embodied in the GDPR: Fairness and discrimination, Purpose limitation, Data minimisation, Transparency and the right to information").

[8] In this sense, for example, analytics make it difficult identify a specific purpose of data processing at moment of data collection and, on the other hand, machine learning algorithms, whose processing purposes are necessarily specified, may not predict and explain how these purposes are achieved. In both the cases the transparency about the purpose and manner of data processing remain therefore frustrated.

To address these issues, legal scholars have highlighted the potential role of transparency [Edwards & Vale, 2017; Selbst & Powles, 2017; Wachter, Mittelstadt & Floridi, 2017; Burrell, 2016; Rossi, 2016], risk assessment [Guidelines, 2017; Mantelero, 2017] and more flexible forms of consent, such as broad consent [Sheehan, 2011] or dynamic consent [Kaye et al., 2015]. Although none of these solutions provides a definitive answer to the shortcomings affecting individual consent, in different contexts these solutions, alone or combined, may boost data subjects' self-determination.

Moreover, the notion of self-determination is not circumscribed by a given data processing. It can be used in a broad sense as freedom of choice concerning the use of AI and the right to have a not-smart version of AI-equipped devices and services. This sort of "zero option" for AI goes beyond the individual dimension and it also relates to the way in which a community decides the role that AI can play in shaping social dynamics, collective behaviour, and decisions affecting entire groups of individuals [Asilomar AI Principles, 2017 ("Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives")].

I.6 Minimisation

As in the Big Data context [Guidelines, 2017], data minimisation represents a challenging issue for AI. Although the technologies adopted are different, both Big Data and machine learning AI algorithms need a large amount of data to produce useful results. This means that only a certain degree of minimisation is possible.

Moreover, as mentioned in the previous section about the "zero option", the adoption of solutions other than AI can contribute to reduce the number of data collected. They may reduce the amount of required information (e.g. carrying out a survey focused on a sample of the population instead of collecting information about a large portion of this population).

Furthermore, some of the guidelines on Big Data adopted by the Council of Europe, can be extended to AI. In this sense, the adopted Guidelines on Big Data state a principle which can be applied to AI as well: data should be collected and processed in such a way as to "minimise the presence of redundant or marginal data".[9] In the AI context this approach mainly concerns the training data. In this sense the Norwegian Data Protection Authority pointed out that "it would be natural to commence with a restricted amount of training data, and then monitor the model's accuracy as it is fed with new data" [The Norwegian Data Protection Authority, 2018].

Although machine learning necessarily requires large datasets in the training phase, it is important to adopt a design paradigm that critically assesses the nature and amount of data used, reducing redundant or marginal data and progressively increasing the

[9] See Guidelines, Section IV, para 4.2.

dimension of the training dataset.[10] Furthermore, minimisation may also be achieved using of synthetic data to train algorithms.

I.7 Bias

Potential bias represents the most critical issue in data-intensive applications, since both deterministic and machine learning AI models use input data to extract further information (analytics) or to create and train ML models. As a consequence, biased datasets may negatively affect both these kind of algorithms, with a higher impact in the case of ML where bias in datasets may affect the design and the development (training) of the algorithm.

This issue has been already partially addressed by the Council of Europe in the Guidelines on Big Data, which suggest the adoption of a by-design approach to avoid any "potential hidden data biases and the risk of discrimination or negative impact on the rights and fundamental freedoms of data subjects, in both the collection and analysis stages".[11]

Potential bias may be due to biased datasets [AI now, 2017, 4, 16-17], but may also be the consequence of intentional or unintentional decisions adopted by developers. In this sense, machine predictions and performance "are constrained by human decisions and values, and those who design, develop, and maintain AI systems will shape such systems within their own understanding of the world" [AI Now, 2017, 18]. For this reason, AI development cannot be left in the hands of AI designers alone: given their technical education, they may be less aware of the societal sequences of their decisions.

Committees of experts from different fields (social science, law, ethics, etc.) may represent the adequate context in which the various issues concerning the impact of AI on individuals can be discussed and addressed (see below Section II.3.1). In this way, expert committees could reduce the potential limitations affecting AI developers' standpoint. Moreover, multidisciplinary committees may also detect the potential bias due to the nature of AI developers, namely gender bias, ideological bias or bias due to underrepresentation of minorities.

Another manner to reduce the potential bias of AI applications is the adoption of participatory forms of risk assessment [Mantelero, 2018]. From this perspective, risk assessment is not merely focused on data security and data quality (see below Section II.3.2) but also on the effective engagement of the groups potentially affected by AI applications, which can contribute to the detection and removal of the existing bias.

[10] See also Guidelines, Section IV, para 4.3 ("When it is technically feasible, controllers and, where applicable, processors should test the adequacy of the by-design solutions adopted on a limited amount of data by means of simulations, before their use on a larger scale").

[11] See Guidelines, Section IV, para 4.2.

This approach, which is focused on responsible AI design [Guidelines, 2017],[12] aims to prevent biased conditions that may affect datasets or algorithms. In a context necessarily characterised by a certain degree of obscurity and complexity, these types of prior assessment and responsible design are more efficient than analyses carried out when a discriminatory result is discovered [Selbst, Andrew D. 'Disparate Impact in Big Data Policing'. Georgia Law Review 52, no. 1 (19 February 2018). www.georgialawreview.org. 163 ("Even if the result can be traced to a data quality problem, those problems are often quite complicated to rectify. It might be easy to determine that something is off about the data, but it can be more difficult to figure out what that something is […] Even if all the sources of bias are identified, the magnitude of each source's effect is still likely unknown")].

Attention to potential bias, since the early stages of design, also entails deeper consideration on training datasets and the training phase in general. This can reduce the negative consequences due to historical bias affecting pre-existing data sets. In this light, it has been proposed "to track the provenance, development, and use of training datasets throughout their life cycle" [AI Now, 2017].

Regarding the training phase, accurate tests before the deployment of AI algorithms on large scale, may reveal hidden bias. For this reason, the Guidelines on Big Data highlighted the role of simulations [Guidelines Big Data;[13] AI Now, 2017].

Finally, the issues concerning machine bias cannot be diminished using the fallacy of human decisions as an argument, by describing AI decision-making as a way to reduce human error rates. This comparison is often wrongly framed. Firstly, AI solutions are designed to be applied in a serial manner. Therefore, such as in product liability, a wrong design (i.e. bias) necessarily affects many people that are in the same or similar situations, while human errors affect a given case.

Secondly, although there are fields in which the error rates of AI are close or lower that human performances, such as image labelling [Artificial Intelligence Index, 2017], complicated decision-making tasks are affected by higher error rates.

Finally, there is a socio-cultural dimension in human error that differs from machine error, in terms of social acceptability and forgiveness. This necessarily influences the propensity to adopt potentially fallible AI solutions.

[12] See Guidelines, Section IV.4.2("Controllers and, where applicable, processors should carefully consider the design of their data processing, in order to minimise the presence of redundant or marginal data, avoid potential hidden data biases and the risk of discrimination or negative impact on the rights and fundamental freedoms of data subjects, in both the collection and analysis stages").

[13] See Guidelines, Section IV.4.3 ("When it is technically feasible, controllers and, where applicable, processors should test the adequacy of the by-design solutions adopted on a limited amount of data by means of simulations, before their use on a larger scale. This would make it possible to assess the potential bias of the use of different parameters in analysing data and provide evidence to minimise the use of information and mitigate the potential negative outcomes identified in the risk-assessment process described in Section IV.2").

Part II- Challenges and possible remedies

II.1 Limitations to AI use

Data protection regulations, as well as Convention 108, provides specific safeguards that can be applied to algorithms (and AI algorithms) used for automated decision-making systems. Nevethless, the red line between human and automated decisions cannot be drawn on the basis of the mere existence of a non-human decision-making process. In fact, the supposedly reliable nature of AI mathematics-based solutions leads those taking decisions on the basis of the results of algorithms to believe the picture of individuals and society that analytics suggest. Moreover, this attitude may be reinforced by the risk of potential sanctions for taking a decision that ignores the results provided by analytics. Therefore, the presence of a human decision-maker is not *per se* sufficient.

AI algorithms benefit from the allure of mathematical objectivity, which combined with the complexity of data management and the subordinate position of the decision-makers in the organization, make it difficult for a human decision-maker to take a decision different to the one suggested by the algorithm.

Against this background, the distinction is between the cases where there is an effective freedom of the human decision-maker and those cases where there is not. In this sense, the Guidelines on Big Data have already highlighted the importance of preserving the effective freedom of the human decision maker.[14]

An important role in carrying out this assessment of different situations of potential imbalance can be played by expert committees (see below Section II.3.1), which may also facilitate stakeholders' participation in the assessment (see below Section II.3.2).

When decisions can be attributed to AI-based systems, as well as when human decision makers cannot have an effective role in supervising AI decisions, a further broader question arises about the adoption of those systems instead of other human-based applications.[15] This should lead communities or groups potentially affected to a participatory discussion on the adoption of AI solutions, analysing their potential risks (see below risk assessment) and, if the solutions are adopted, monitoring these applications (see below vigilance).

[14] See Guidelines, Section IV.7.4 ("On the basis of reasonable arguments, the human decision-maker should be allowed the freedom not to rely on the result of the recommendations provided using Big Data").

[15] See, e.g., ITU, 2017, 34 ("Dr. Margaret Chan, [the now former] Director-General of WHO, observed that "medical decisions are very complex, based on many factors including care and compassion for patients. I doubt a machine can imitate – or act – with compassion. Machines can rationalize and streamline, but AI cannot replace doctors and nurses in their interactions with patients").

II.2 Transparency

Transparency has different meanings. It may consist in the mere disclosure about the AI applications in use, in the description of their logic or in the access to the structure of the AI algorithms and – when applicable – to the datasets used to train these algorithms.

Although awareness is important for a public scrutiny of automated decision-making models, a generic information on the use of AI weakly contributes to tackle the risks of unfair or illegitimate data use. On the contrary, the access to the structure of algorithms makes it possible to assess potential biases, but IP protection and competition issues may limit this access. Moreover, in some cases, transparency may conflict with the performance of the tasks carried out by public bodies (e.g. predictive policing systems).

For these reasons, the solution focused on the disclosure of the logic of algorithms may be the most appropriate. Nonetheless, this disclosure can be interpreted in a narrow or in an extensive manner. Providing information about the nature of input data and expected output,[16] disclosing the variables of algorithms and their weight, and providing access to the architecture of analytics are different possible way to be transparent regarding the logic of AI algorithms.

Complex models of analysis (e.g. *deep-learning*) challenge this notion of transparency – in terms of explanation of the logic of algorithms [Goodman & Flaxman, 2016] and of the decisions adopted using analytics[17] – and non-deterministic models make it difficult to provide specific information about the logic of data processing.

Moreover, the dynamic nature of many algorithms may contrast with a static idea of transparency. Algorithms are continuously update and changed while transparency is static per se as the disclosure concerns the algorithm used in a given moment.

Finally, access to AI algorithms is not enough to detect potential bias. Specific resources in terms of time and skills are necessary to carry out this analysis and to detect potential bias in the design or functioning of algorithms. This reduces the deterrent effect of solutions such as auditing or human decision-maker intervention.[18] For this reason, research projects are developing methods of bias detection based on algorithms, although introducing a sort of algorithmic supervisor for algorithms does not reduce the complexity of the data governance systems.

These different elements do not affect the positive effect of increasing algorithms transparency [Burrell, 2016], mainly when used in the public sector,[19] and its potential role in safeguarding data subjects' self-determination [Edwards & Vale, 2017; Selbst & Powles, 2017; Wachter, Mittelstadt & Floridi, 2017; Rossi, 2016], but transparency

[16] This information may be provided through learning by use models, giving data subjects the chance to test analytics with different input values. Nevertheless, also in this case there is the risk of a misleading identification of the relevant inputs [Diakopoulps, 2013, 18].

[17] In these cases, it may be impossible providing an explanation of the reason for the decision suggested by the algorithm [Burrell, 2016].

[18] These remedies are possible, but in many cases the auditing process requires a significant effort and the human intervention is weakened by the complexity of data processing.

[19] In this regard, public sector is characterised by a more stable nature of used algorithms, to comply with the principle of equal treatment and is committed to specific duties in terms of transparency and access rights concerning administrative processes.

represents only a part of the solution to address the challenges of AI. Moreover, algorithms are only a part of AI applications, since an important role is played by the datasets used to train them or to be analysed by AI. This means that bias datasets necessarily lead to biased results. Finally, data-intensive applications may focus on data in a decontextualized way, losing contextual information not processed by AI, which may be useful to understand and apply the solution provided by AI applications.

II.3.1 Risk assessment

Given the limits affecting transparency and individual self-determination (see above Section I.5) data protection regulations are increasingly emphasising the role of risk assessment. Risk assessment carried out by data controllers and a safe AI environment have a positive impact on individuals' trust and on their willingness to use AI applications. In this way, users' preferences can be based on an effective risk analysis and on the measures adopted to mitigate the risks [The Norwegian Data Protection Authority, 2018, 4], rather than relying on mere marketing campaign or brand reputation.

In this regard, the use of algorithms in the context of modern data processing techniques [Council of Europe-Committee of experts on internet intermediaries (MSI-NET), 2018] as well as data-intensive technological trends [EDPS, 2018] have led to the adoption of a border viewpoint in bringing into focus the potential negative outcomes of data processing [Asilomar AI Principles, 2017 ("Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact")]. This has forced groups of experts and scholars to go beyond the traditional sphere of data protection [Taylor, Floridi & van der Sloot, 2017] and consider the impact of data use on fundamental rights and collective social and ethical values [Mantelero, 2018].

This assessment concerning the compliance of data use with ethical and social values is more complicated than the traditional data protection assessment. Whereas the driving values (e.g. integrity of data) in the context of data security and data management are technologically-based and can therefore be generalised across different social contexts, with social and ethical values the situation is different. These are necessarily context-based and change from one community to another, making it difficult to identify the benchmark that should be adopted in this kind of risk assessment.

This point is clearly addressed in the first section of the Guidelines on Big Data [Council of Europe, 2017], which urges both data controllers and data processors to "adequately take into account the likely impact of the intended Big Data processing and its broader ethical and social implications", in order to safeguard human rights and fundamental freedoms, in the light of Convention 108.[20]

The innovative element in risk-assessment concerns therefore the range of the interests safeguarded and rights protected. In this sense, the assessment encompasses rights that go beyond the traditional sphere of data protection, such as the right to non-discrimination [Barocas & Selbsr, 2016], and it also takes into account the compliance of data processing with social and ethical values [European Economic and Social Committee, 2017; AI Now, 2017, 34-35 ("In order to achieve ethical AI systems in which their wider implications are addressed, there must be institutional changes to hold power accountable")].

Moreover, the Guidelines recognise the relative nature of social and ethical values and, in this sense, require that data uses are not in conflict with the "ethical values commonly accepted in the relevant community or communities and should not prejudice societal interests, values and norms".[21] Although the Guidelines recognise the difficulties of defining the values that should be taken into account in conducting a broader assessment, they nevertheless do point out some practical steps to identify these values. They suggest that "the common guiding ethical values can be found in international charters of human rights and fundamental freedoms, such as the European Convention on Human Rights", following the position of privacy scholars who have examined this issue.[22]

Given the context-dependent nature of the social and ethical assessment and the fact that international charters may only provide high-level guidance, the Guidelines combine this general suggestion with a more tailored option, represented by "ad hoc ethics committees".[23] When the assessment highlights "a high impact of the use of Big Data on ethical values", these committees, which in some cases already exist in practice, should identify the specific ethical values to be safeguarded with regard to a given use of data, providing more detailed and context-based guidance for risk assessment.[24]

In light of the above, the "architecture of values" defined by the Guidelines is based on three layers. The first general level is represented by the "common guiding ethical values" of international charters of human rights and fundamental freedoms. The second layer takes into account the context-dependent nature of the social and ethical assessment and focuses on the values and social interests of given communities. Finally,

[20] See Guidelines, Section IV, para 1.1.

[21] Guidelines, Section IV, para 1.2.

[22] See David Wright, 'A framework for the ethical impact assessment of information technology' (2011) 13 Ethics Inf. Technol. 199, 201–202.

[23] See Guidelines, Section IV, para 1.3 ("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").

[24] The same two-layer model, based on general guidelines and tailored guidance provided by *ad hoc* committee, is already adopted in clinical trials. In this field, as the big data context, the specific application of technology poses context-related questions which must necessarily be addressed on the basis of the conflicting interests that characterise each case. This results in an "in the context" assessment of the conflicting interest.

the third layer consists in a more specific set of ethical values provided by ethics committees and focused on a given use of data.

The complexity of this assessment and the continuous evolution of both the potential risks and the measures to tackle them. In this regard, data protection Supervisory Authorities may play a significant role in supporting data controllers, informing them about the state-of-the-art of data security measures and providing detailed guidelines on the risk-assessment process.[25] Therefore, the Guidelines do not leave the assessment exclusively in the hands of data controllers. In line with the approach adopted in Regulation (EU) 2016/679, when the use of big data "may significantly impact" the rights and fundamental freedoms of data subjects, controllers should consult supervisory authorities to seek advice and to mitigate the risks outlined in the impact assessment. [26]

II.3.2 Ethics committees

In respect to data intensive applications, the potential role of ethics committees is increasingly drawing attention in the current AI, although there is not a unanimous consensus on the nature and function of these committees. Theoretical analyses, policy documents and corporate initiatives offer different solutions in this regard.

The first difference emerging in these approaches concerns the level these committees should work at [Polonetsky, Tene & Jerome, 2015; Calo, Ryan. 2013; White House, 2015; IEEE, 2016]. Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings 13 Colorado Technology Law Journal 333-367]. Some proposals describe them as national committees [Villani, 2018] which provide general guidelines addressing the issues of AI development. This solution is not completely new and is similar to the existing national bioethical committees. Nevertheless, in the case of AI data-intensive applications that use personal information, the potential interplay between these national committees and the national data protection Supervisory Authorities should be carefully addressed [Mantelero, 2016], as well as the interplay with other national authorities, such as antitrust authorities or national security bodies. Moreover, many countries have independent bodies responsible for supervising specific sectors where AI applications operate or may operate. From a regulatory perspective, it is therefore important to deal with these authorities and to reconsider their role or strengthen their mutual cooperation, including data protection Supervisory Authorities [European Data Protection Supervisor, 2016, 3, 15].

Following a different approach, ethics committees may operate at company level, supporting data controllers with regard to specific data applications. In terms of data protection, this means that they focus on data controllers' operations [Mantelero, 2018]. Moreover, they may assume a broader nature and act as expert committees

[25] See Guidelines, Section IV, para 2.8.
[26] See Guidelines, Section IV, para 2.8.

considering not only the ethical issues, but also a broad array of societal issues of AI applications, as well as the contextual application of fundamental rights [Mantelero, 2018]. In this sense, several companies[27] have already set up internal or external committees that give advice about critical projects.

The second solution, based on ethics committees at company level, creates less difficulties in terms of overlap with the exiting regulatory or supervising bodies but may require defining the relationship between these committees and the Supervisory Authorities. National legislators might empower Supervisory Authorities to scrutinise these committees when shortcomings in their abilities or decisions may affect data processing [Conseil national du numérique, 2015]. Moreover, such as in other cases of advisory boards, the creation of AI ethics committees raises questions in terms of their independency. This concerns the internal or external nature of the committee and the best practices to exclude any conflict of interests.

The complexity of AI tools and applications have an influence on the possible composition of these committees. When societal issues are relevant, legal expertise, ethical or sociological background, as well as domain-specific knowledge will be required to the members of the committee.[28]

The role of these committees may be even more relevant in the context where transparency and stakeholders' engagement is difficult, such as predictive justice, crime detection or predictive policing.

The review carried out by ethics committees can provide a significative support to AI developers to design rights- and socially-oriented algorithms. Moreover, the necessary dialogue between the developers and the committee's experts can positively boost the

[27] See, in this sense, the increasing propensity of the big data-intensive and high-tech companies to set up their own ethics committees or advisory boards. See, e.g., Natasha Lomas, 'DeepMind now has an AI ethics research unit. We have a few questions for it…' *TechCrunch* (4 October 2017) < http://social.techcrunch.com/2017/10/04/deepmind-now-has-an-ai-ethics-research-unit-we-have-a-few-questions-for-it/> accessed; Axon AI Ethics Board <https://it.axon.com/info/ai-ethics> accessed 9 May 2018; DNA Web Team, 'Google drafting ethical guidelines to guide use of tech after employees protest defence project' *DNA India* (15 April 2018) <http://www.dnaindia.com/technology/report-google-drafting-ethical-guidelines-to-guide-use-of-tech-after-employees-protest-defence-project-2605149> accessed 7 May 2018. See also United Nations, 2011. Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. United Nations Human Rights Council (UN Doc. HR/PUB/11/04).

[28] When there is a lower level of complexity in terms of consequences of AI applications, the committee could be replaced by an expert in ethics and society. This role is similar to DPO's role for data protection and also for the members of the ethics committee there should be a mandatory requirement concerning this appointment and the quality of the appointed members. Regarding the reasons leading companies to this appointment, the key objective should be in relation to the nature of data use in terms of significant potential impact on fundamental rights, ethical and societal issues should be the main criteria. See in this sense IEEE, 2016, 41-42 which recommends "to create roles for senior level marketers, ethicists or lawyers who can pragmatically implement ethically aligned design […] A precedent for this new type of leader can be found in the idea of a Chief Values Officer created by Kay Firth-Butterfield" (CSER Cambridge. Kay Firth-Butterfield: Lucid AI's Ethics Advisory Panel., 2016. https://www.youtube.com/watch?v=w3-wYGbNZU4).

creation of more transparent data processing procedures as well as facilitate a better definition of their rationale (see also Section II.2).

II.3.3 Participatory assessment and vigilance

Experts (e.g. ethics committees) may play an important role in detecting the potential negative consequences or the critical issues of AI applications and may support data controllers in addressing them. Nevethless, there are cases in which this analysis cannot be carried out without engaging the target communities or groups.

Like in the Social Impact Assessment, societal consequences of AI may therefore arouse an interest in public participation, individual and group empowerment through the assessment process, non-discrimination and equal participation in the assessment. Moreover, a participatory approach[29] in impact assessment can be useful to get a better understanding of the different competing interests and ethical and social values.[30]

Stakeholder engagement also represents a development goal for the assessment [United Nations Office of the High Commissioner for Human Rights, 2006], since it reduces the risk of under-representing certain groups and may also flag up critical issues that have been underestimated or ignored by data controller [Wright & Mordini, 2012, 402].

[29] The role of participatory approaches and stakeholders' engagement is specifically recognised in the context of fundamental rights [The Danish Institute for Human Rights, 2016, 24; Paul De Hert, 'A Human Rights Perspective on Privacy and Data Protection Impact Assessments. In David Wright and Paul De Hert (eds) Privacy Impact Assessment (Springer Dordrecht) 72 ("Further case law is required to clarify the scope of the duty to study the impact of certain technologies and initiatives, also outside the context of environmental health. Regardless of the terms used, one can safely adduce that the current human rights framework requires States to organise solid decision-making procedures that involve the persons affected by technologies")].

[30] Participation of the different stakeholders (e.g. engagement of civil society and the business community in defining sectoral guidelines on values) can achieve a more effective result than mere transparency, although the latter has been emphasized in the recent debate on data processing [The Danish Institute for Human Rights, 2016, 10 ("Engagement with rights-holders and other stakeholders are essential in HRIA […] Stakeholder engagement has therefore been situated as the core cross-cutting component")]. See also Walker, 2009, 41 ("participation is not only an end – a right – in itself, it is also a means of empowering communities to influence the policies and projects that affect them, as well as building the capacity of decision-makers to take into account the rights of individuals and communities when formulating and implementing projects and policies"). A more limited level of engagement, focused on awareness, was suggested by Council of Europe-Committee of experts on internet intermediaries [Council of Europe-Committee of experts on internet intermediaries (MSI-NET), 2018, 45 ("Public awareness and discourse are crucially important. All available means should be used to inform and engage the general public so that users are empowered to critically understand and deal with the logic and operation of algorithms. This can include but is not limited to information and media literacy campaigns. Institutions using algorithmic processes should be encouraged to provide easily accessible explanations with respect to the procedures followed by the algorithms and to how decisions are made. Industries that develop the analytical systems used in algorithmic decision-making and data collection processes have a particular responsibility to create awareness and understanding, including with respect to the possible biases that may be induced by the design and use of algorithms")].

However, stakeholder engagement should not become a way for decision makers (data controllers, in this case) to avoid their responsibilities as leaders of the entire process [Palm & Hansson, 2006]. Decision makers remain committed to achieving the best results in terms of minimising the potential negative impacts of data processing on individuals and society.

Finally, a participatory assessment to tackle the risks of large scale effects of algorithmic decision-making systems [CNIL, 2017, 30] may also lead data controllers to adopt co-designing solutions in developing AI applications, actively engaging the potential groups of people affected by them.

II.4 Liability

Liability concerning AI applications remains an open issue for different reasons. As demonstrated by experience in the field of product liability, whose principles focused on risk management and uncertainty may be largely extended to AI, different regulatory models (strict liability, liability based on fault etc.) and strategies (state intervention, mandatory insurance etc.) are possible.

The extension of the product liability logic to algorithms, with its canalization on the producer of the liability, seems to provide a valuable solution. This seems more efficient than the different option based on a sort of data protection officer for algorithms [CNIL, 2017, 56 “identifying within each company or authority a team that is responsible for an algorithm’s operation the moment this processes the data of humans”], where the pervasive nature of AI applications, the plurality of parts involved and the role played by the users make it difficult to disentangle the different aspects of AI liability.

Moreover, liability is a sort of closing rule of the system, which is useful when the different ex ante remedies (such as transparency) did not work [Asilomar AI Principles, 2017 (“Failure Transparency: If an AI system causes harm, it should be possible to ascertain why”)]. However, since tort liability is a matter largely referred to national legislators, this report does not investigate the different possible solutions that can be adopted in this regard.

Nevethless, it is worth pointing out how risk management, transparency and liability can be combined not only with regard to the development phase of AI applications, but also in the following stage, when the algorithms are in use. This can lead supervisory authorities and data controllers to adopt forms of algorithm vigilance based on the model of pharmacovigilance to quickly react in case of unexpected and dangerous performance of algorithms (e.g. Microsoft's chatbot Tay[31]) [Commission Nationale de l'Informatique et des Libertés - LINC, 2017].

[31] See Vincent, James. 'Twitter Taught Microsoft's Friendly AI Chatbot to Be a Racist Asshole in Less than a Day'. The Verge, 24 March 2016. https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

Annex I

Summary of further topics that can be discussed in the final version of the Report

Sector-specific analysis: Public sector and workplaces

Public sector

AI applications rise specific issues when they are used in the public sector. This is due to the imbalance of power existing between citizen and administration and the essential nature of the services provided. This suggests the adoption of higher safeguards, not merely circumscribed to the potential role of ad hoc committees or auditing. Safeguards should also relate to the adoption of an evaluation process that critically considers the necessity and opportunity of the proposed AI solutions concerning the services provided by public agencies or private companies acting on their behalf.

In this regard, the serious due process concerns about the use of AI solutions by public agencies require that "at a minimum they should be available for public auditing, testing, and review, and subject to accountability standards" [AI Now, 2017].

Workplaces

Leaving aside the AI impact on the labour market, AI solutions may affect the relationships within the workplace. They may increase the potential control of the employer over employees, in a context that is often characterised by an imbalance of power.

Moreover, the adoption of hidden and unassessed forms of data processing may transform workplace in an *in vivo* social experiment. This rises addtional important questions about the role of transparency, ethics committees and voluntary participation in data processing.

Finally, some devices provided by employers to employees may have a duale use. For instance, wearable well-being devices can be used in the workplace to detect biological parameters useful to safeguard employers' health but can also be used by employees outside the workplace, for example, to track their sport performances. This dual use, if not supported by an analysis of its consequences on data protection and adequate design solutions, may infringe the barrier between private and working life [AI Now, 2017, 10], raising questions about pervasive forms of control and the right to disconnect.

References (provisional list)

- AI Now. 2017. AI Now 2017 Report. Accessed 26 October 2017. https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf
- Artificial Intelligence Index - Annual Report 2017. http://aiindex.org/2017-report.pdf accessed 5 December 2017.
- Asilomar AI Principles. 2017 https://futureoflife.org/ai-principles/ accessed 27 March 2018.
- Axon AI Ethics Board https://it.axon.com/info/ai-ethics accessed 9 May 2018
- Barocas, Solon & Nissenbaum. 2015. Big Data's End Run around Anonymity and Consent. In Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum (eds), *Privacy, big data, and the public good : frameworks for engagement* (Cambridge University Press).
- Bostrom, Nick. 2016. Superintelligence paths, dangers, strategies. Oxford, Oxford University Press.
- Boyd, Danah & Crawford, Kate. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. 15(5) Information, Communication, & Society 662–679
- Burrell, Jenna. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms 3(1) Big Data & Society https://doi.org/10.1177/2053951715622512 accessed 03 March 2018.
- Calo, Ryan. 2013. Consumer Subject Review Boards: A Thought Experiment. 66 Stan. L. Rev. Online 97 http://www.stanfordlawreview.org/online/privacy-and-big-data/consumer-subject-review-boards accessed 23 February 2018.
- Citron, Danielle K. & Pasquale, Frank. 2014. The Scored Society: Due Process For Automated Predictions. 89 Wash. L. Rev. 1–33
- CNIL. 2017. How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence. Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill', December 2017, p. 14. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf
  .
- CNIL. 2017. How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence. Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf accessed 23 May 2018.
- Commission - European Group on, Ethics in Science and, & New Technologies. 2018. Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Retrieved from https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

- Commission Nationale de l’Informatique et des Libertés - LINC. 2017. La Plateforme d’une Ville Les Données Personnelles Au Coeur de La Fabrique de La Smart City. https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip5.pdf accessed 4 March 2018.
- Conseil national du numérique. 2015. Ambition numérique : Pour une politique francaise et europeéenne de la transition numérique http://www.cil.cnrs.fr/CIL/IMG/pdf/CNNum--rapport-ambition-numerique.pdf accessed 03 March 2018.
- Council of Europe. 2017. Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a.
- Council of Europe-Committee of experts on internet intermediaries (MSI-NET). 2018. Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications. https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.
- Diakopoulps Nicholas, 2013. Algorithmic Accountability Reporting: on the Investigation of Black Boxes (Tow Center for Digital Journalism).
- DNA Web Team, ‘Google drafting ethical guidelines to guide use of tech after employees protest defence project’ DNA India (15 April 2018) http://www.dnaindia.com/technology/report-google-drafting-ethical-guidelines-to-guide-use-of-tech-after-employees-protest-defence-project-2605149 accessed 7 May 2018.
- Edwards, Lilian and Vale, Michael. Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. (2017) 16(1) Duke Law and Technology Review 18-84.
- European Commission - European Group on, Ethics in Science and, & New Technologies. 2018. Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems, https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf accessed 4 May 2018.
- European Data Protection Supervisor - Ethics Advisory Group, 2018. Towards a digital ethics https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf accessed 4 March 2018.
- European Data Protection Supervisor. 2016. Opinion 8/2016. EDPS Opinion on coherent enforcement of fundamental rights in the age of big data.
- European Economic and Social Committee. 2017. The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context. https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data accessed 4 June 2018.
- European Parliament. 2017. European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI))

http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//EN&language=EN accessed 4 March 2018.

- European Union Agency for Fundamental Rights (FRA). 2018. Artificial Intelligence, Big Data and Fundamental Rights http://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights accessed 4 May 2018.
- Executive Office of the President, and National Science and Technology Council - Committee on Technology. 2016. Preparing for the Future of Artificial Intelligence (Washington D.C.) https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf accessed 4 May 2018.
- Goodman, Bryce & Flaxman, Seth. 2016. EU Regulations on Algorithmic Decision-Making and a “right to Explanation” arXiv:1606.08813 [cs, stat]http://arxiv.org/abs/1606.08813 accessed 03 March 2018.
- Goodman, Bryce and Flaxman, Seth. 2016. EU Regulations on Algorithmic Decision-Making and a “right to Explanation”. [2016] arXiv:1606.08813 [cs, stat] http://arxiv.org/abs/1606.08813 accessed 03 March 2018.S.
- IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. 2016. Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1. IEEE, 2016. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.IEEE
- ITU, 2017. AI for Good Global Summit Report 2017 https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf accessed 27 March 2018.
- Kaye, Jane et al. 2015. Dynamic consent: a patient interface for twenty-first century research networks. 23 (2) European Journal of Human Genetics 141
- Kurzweil, Ray. 2016. The singularity is near : when humans transcend biology (London : Duckworth, 2016).
- Lycett, Mark. 2013. Datafication’: making sense of (big) data in a complex world (2013) 22 (4) European Journal of Information Systems 381–386
- Mantelero Alessandro. 2018. AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. Assessment. Computer Law & Security Review (2018), https://doi.org/10.1016/j.clsr.2018.05.017 accessed 12 May. 2018.
- Mantelero, Alessandro. 2014. The future of consumer data protection in the E.U. Rethinking the “notice and consent” paradigm in the new era of predictive analytics. Computer Law and Security Review, 30 (6): 643-660.
- Mantelero, Alessandro. 2017. Regulating Big Data. The guidelines of the Council of Europe in the Context of the European Data Protection Framework’ (2017) 33(5) Computer Law & Sec. Rev. 584-602.
- Mayer-Schönberger, V. and Cukier, K. 2013. Big Data. A Revolution That Will Transform How We Live, Work and Think (London : John Murray).

- McCulloch, Warren S. and Walter H. Pitts. 1943. A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5:115-133, 1943.
- Mireille Hildebrandt. 2016. SmartTechnologies and the End(s) of Law : Novel Entanglements of Law and Technology (Edward Elgar Publishing).
- Natasha Lomas, 'DeepMind now has an AI ethics research unit. We have a few questions for it…' TechCrunch (4 October 2017) http://social.techcrunch.com/2017/10/04/deepmind-now-has-an-ai-ethics-research-unit-we-have-a-few-questions-for-it/ accessed 3 May 2018.
- OmerTene & Jules Polonetsky. 2012. Privacy in the Age of Big Data. A Time for Big Decisions. 64 Stan. L. Rev. Online 63–69.
- O'Neil, Cathy. 2017. Weapons of math destruction (London : Penguin Books, 2017).
- Palm, Elin and Hansson, Sven Ove. 2006. The case for ethical technology assessment (eTA). 73(5) Technological Forecasting & Social Change 543, 550–551.
- Philip Bray et al. 2015. International differences in ethical standards and in the interpretation of legal frameworks SATORI Deliverable D3.2 http://satoriproject.eu/work_packages/legal-aspects-and-impacts-of-globalization/ accessed 20 February 2017.
- Rossi, Francesca. 2016. Artificial Intelligence: Potential Benefits d Ethical Considerations' (European Parliament: Policy Department C: Citizens' Rights and Constitutional Affairs 2016) Briefing PE 571.380 http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_EN.pdf accessed 03 May 2018.
- Rouvroy, Antoinette. 2016. "Of Data and Men": Fundamental Rights and Liberties in a World of Big Data' https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020.
- Rubinstein, Ira S. 2013. Big Data: The End of Privacy or a New Beginning?. 3 (2) International Data Privacy Law 74–87.
- Selbst, Andrew D. and Powles, Julia. 2017. Meaningful Information and the Right to Explanation. 7(4) International Data Privacy Law 233–242
- Sheehan, Mark. 2011. Can Broad Consent be Informed Consent? (3) Public Health Ethics 226–235.
- Solon Barocas, Andrew D. Selbsr, 2016. Big Data's Disparate Impact. 104 (3) California Law Review 671-732
- Taylor, Linnet, Floridi, Luciano and van der Sloot, Bart (eds). 2017. Group Privacy: New Challenges of Data Technologies (Springer International Publishing).
- The Danish Institute for Human Rights. 2016. Human rights impact assessment guidance and toolbox (The Danish Institute for Human Rights, 2016) https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-and-toolbox accessed 20 December 2017.

- The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. 2016. Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1. IEEE, 2016. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html accessed 21 February 2018.
- The Norwegian Data Protection Authority. 2018. Artificial Intelligence and Privacy Report. https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf accessed 28 May 2018.
- The Norwegian Data Protection Authority. 2018. Artificial Intelligence and Privacy Report. https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf accessed 27 February 2018.
- Turing, Alan. M. 1950. Computing Machinery and Intelligence. 49 Mind 433–460.
- United Nations Office of the High Commissioner for Human Rights. 2006. Frequently asked questions on a human rights-based approach to development cooperation' (New York and Geneva: United Nations).
- United Nations, 2011. Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. United Nations Human Rights Council (UN Doc. HR/PUB/11/04).
- Villani, Cédric. 2018. For a Meaningful Artificial Intelligence towards a French and European Strategy https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf accessed 29 March 2018.
- Wachter, Sandra, Mittelstadt, Sandra and Floridi, Luciano. 2017. Why a right to explanation of automated decision - making does not exist in the General Data Protection Regulation. 7(2) International Data Privacy Law 76–99.
- Walker, Simon Mark. 2009. The Future of Human Rights Impact Assessments of Trade Agreements. (Utrecht: G.J. Wiarda Institute for Legal Research) https://dspace.library.uu.nl/bitstream/handle/1874/36620/walker.pdf?sequence=2 accessed 26 April 2018.
- White House. 2015. Consumer Privacy Bill of Rights. §103(c) (Administration Discussion Draft 2015. https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf accessed 12 March 2018.
- Wight, David and Mordini, Emilio. 2012. Privacy and Ethical Impact Assessment' in David Wright and Paul De Hert (eds) Privacy Impact Assessment (Springer Netherlands 2012) 397–418.
- Wight, David and Mordini, Emilio. 2012. Privacy and Ethical Impact Assessment. In David Wright and Paul De Hert (eds). Privacy Impact Assessment (Springer Dordrecht) 397–418.
- Wright, David and De Hert, Paul (eds). 2012. Privacy Impact Assessment (Springer Dordrecht).
- Wright, David. 2011. A framework for the ethical impact assessment of information technology. 13 Ethics Inf. Technol. 199, 201–202.