

Computer Ethics - Philosophical Enquiry (CEPE) Proceedings

Manuscript 1005

The Right to Human Intervention: Law, Ethics and Artificial Intelligence


Maria Kanellopoulou - Botti

Fereniki Panagopoulou

Maria Nikita

Anastasia Michailaki

Follow this and additional works at: https://digitalcommons.odu.edu/cepe_proceedings

 Part of the [Applied Ethics Commons](#), [Communication Technology and New Media Commons](#), [Digital Humanities Commons](#), [Ethics and Political Philosophy Commons](#), [Information Literacy Commons](#), [Information Security Commons](#), [Science and Technology Studies Commons](#), and the [Social Media Commons](#)

The right to human intervention: law, ethics and artificial intelligence¹

Maria Kanellopoulou-Botti
Ionian University

Fereniki Panagopoulou
Panteion University

Anastasia Michailaki
Ionian University

Maria Nikita
University of Macedonia

Abstract

The paper analyses the new right of human intervention in use of information technology, automatization processes and advanced algorithms in individual decision-making activities. Art. 22 of the new General Data Protection Regulation (GDPR) provides that the data subject has the right not to be subject to a fully automated decision on matters of legal importance to her interests, hence the data subject has a right to human intervention in this kind of decisions

Keywords: human intervention, automatization, algorithms, Art. 22, GDPR

1. Introduction

The right to human intervention is a renewed right under Art. 22 of the General Data Protection Regulation,. The purpose of this right is to address the fear of the 'instrumentalization' of the person, when important decisions affecting the person are taken solely by automatic decision making. Such important decisions that can be taken by automatic decision making include, for example, hiring personnel, the granting of a loan by a bank or entering into a health insurance contract. Citizens have been generally supported as benefiting from automatic decision making, as this method has been shown to constitute an efficient way of cost containment, as well as being possibly fairer than the taking of exclusively human decisions. The present analysis deals with the right to human intervention from a legal, ethical/philosophical and technological point of view. Law and ethics will be examined in connection to the technological neutrality of algorithms as a way to ascertain the lack of discrimination and bias and the

¹ This paper is composed within the framework of a research project titled "The New Data Protection Law in Europe - a Regulation for the future" within the framework of the Operational Programme "Human Resources Development, Education and Lifelong Learning" of NSRF - Partnership Agreement 2014-2020 and is co-funded by Greece and the European Union – European Social Fund. (Law 4314/2014 in accordance with the requirements of European Regulation (EC) 1303/2013)

protection of the human rights of citizens. The paper will deal with: (a) the true and correct interpretation of the existing law on the right to human intervention; (b) new protective provisions, both for citizens affected by automatic decision making but also with the provisions for organizations/other entities for whom the use of automatic decision making is a crucial part of their operations; (c) an analysis of the philosophical/ethical foundations of the above rights; (d) ethical rules on automated decision making; and (e) the fundamental principles of the making of algorithms, which will primarily serve the principles of justice, transparency, accountability and equality for citizens.

Furthermore, the paper will analyze, along with the legal, ethical/philosophical and technological implications of the right to human intervention, the technical way of automated decision making towards serving the principles of justice, transparency, accountability and equality.

2. State of the art

The international literature and press have presented us with innumerable cases and stories, as well as research projects, which show algorithmic inequalities, discrimination and bias. Search engines, internet applications of many kinds, such as e-dating, Airbnb, Uber, etc., are being constantly accused of racism, misogyny and class discrimination. The critics demand the purification of the algorithms and human intervention in decision making, towards the total ascertainment of due process for consumers. An attempt to provide a partial answer to these criticisms in Europe was made by virtue of Art. 22 GDPR

2.1. The Law

The right to human intervention, as enshrined in Art. 22 of the GDPR, is not exactly a new right. It has evolved from the older, similar right under Art. 15 of the Data Protection Directive 95/46/EC. Art. 22 GDPR provides that the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning the data subject or which affects the data subject significantly, in a similar way. Crucially, profiling is explicitly included in the provision and its meaning is defined as the automated personal data processing towards the evaluation of certain traits on the person's performance in her work, the financial situation of the person, her health, her personal preferences, interests, reliability or behavior, as well as the position of movements of a person. Art. 22 is all the more important due to the possibilities of the analysis of Big Data. As an exception, the paper provides that the right shall not apply (a) where the decision is necessary for entering into, or for the performance of, a contract between the data subject and a data controller; (b) it is authorized by European Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights, freedoms and legitimate interests; or (c) where the decision is based on the data subject's explicit consent.

These provisions, as clear as they may appear, have already given rise to serious doubts, internationally, as to their interpretation. It is supported that Art. 22

provides an absolute prohibition of automated decision making without human intervention, unless an exception (Art. 22, par. 2 a, b or c) applies. Indeed, the Article 29 Data Protection Working Party issued guidelines on automated decision making and profiling, supporting precisely this interpretation. It has also been argued, however, that the right under Art. 22 provides only a sui generis right to object to automated decision making, without, at the same time, granting a 'pure' right to human intervention. Yet another interpretation suggests that it introduces a right to explanation regarding the decision taken, with arguments based on Arts. 13, 14 and 15 of the GDPR and the Recital 71. This construction has also been heavily criticized, stating that no such rights has been instituted, but a much more limited right to information of the data subject. A third view suggests that both previous constructions are mistaken and that we much stick to what the Regulation really states, namely a right to meaningful information, for the data subject, who can also ask for the logic of the automated decision making in question. In some texts we see that the most trivial intervention by a natural person means that the requirements of Art. 22 have been met, whereas other researchers claim that this intervention must be substantial, and also, by a competent to decide person. We are confronted, therefore, with a field of important disagreements, where, however, it is crucial to arrive soon at a complete resolution of these questions.

The numerous official objections and comments filed on the draft guidelines of the Art. 29 Data Protection Working Party which took a clear position for the absolute prohibition of automated decision making under Art. 22 show us that this matter presents us, already before the application of the Regulation, serious difficulties of interpretation. They also show us that the issue affects a series of important sectors such as banking, commerce, education, public security, the workplace, insurance and many more.

2.2 Ethics

The right to human intervention gives rise to a series of philosophical/ethical questions. The designer of an automated decision system, of the algorithms, bears an important moral responsibility to avoid as much as possible any bias which may be inserted in the system, a responsibility not fully explored in the literature. Another novel ethical question relates to the autonomy of the ethical person and is twofold: it touches, on the one hand, the controller and the manager of the algorithms, and on the other hand, the data subject who is being affected by these algorithms. The autonomy of the former is being limited, as there is no room for an evaluation under his own free will; parameterization and the initial preparation of the algorithmic system conclude his involvement and the next choices are made by the automated system itself. The controller/manager, therefore is unable to exercise a main facet of his freedom as a moral person, a freedom which would allow him to change his mind, in line with a potential future change of circumstances. To cite Kant, this limitation of freedom of will in relation to the initial stages of the construction of the system also limits his freedom and his dignity. As regards the data subject being evaluated, he cannot project elements of his personality, which are not capable of being evaluation under the system used, as they are not able to be parameterized.

If we use as an example the case of the automatic hiring in the public or private sector with an automated decision making system, this could lead is, initially, to state

that this way is the best answer to a series of 'pathological' instances ; neither favoritism nor nepotism can survive, in theory at least, this system. The role of human judgment is minimized or even, extinguished and with it, also the risk of fallacies or negligence for which human judgment is notorious. The application of algorithmic standards and automated decision making has been already applied for many years in sectors such as the economy and has been perfected by the use of hedge funds, banking institutions and state organizations, promises equality before the law, meritocracy, justice and equal opportunity to anyone enjoying the same qualifications to gain a place in the public or private sector. However, as nothing is clear of all negative or potentially harmful effects, the total ban of the human intervention due to automated decision making is coupled with important ethical implications.

Will we really achieve our goals of meritocracy, justice, equality? Even if data management takes place without human intervention, by algorithms, the choice of the data sets and their parameters to be taken, or, more importantly, the choice of data that is to be omitted from the calculation, remains a task for a human, who will always parameterize the systems of algorithms. In other words, it seems that human intervention does not exactly 'vanish' in reality, but rather that it is assigned at an earlier stage to said automated decision-making system, before the process of selection. This, however, negates, or at least severely weakens, the possible legal contest of the results of a public sector competition by a particular person, since the results are not related to a particular case, but to a whole class.

The third, and possibly more crucial ethical issue, deals exactly with the above inability to parameterize particular traits of human personality, which are essential to our humanness. One can never parameterize a character and a personality, therefore these remain outwith the scope of perception of any automated decision-making system. Nevertheless, this inability despises our most important personal and professional traits, our morality and our virtues, precisely because morality and virtues cannot be parameterized. Consequently, the choices made entirely by algorithms neglect the only trait of our personality that, in reality, guarantees our normal inclusion and functionality within the workplace and, in general, within any social environment.

2.3. Artificial Intelligence

As far as automated decisions that take place by artificial intelligence are concerned, what we initially look for is technological neutrality, so as to prevent, as much as possible, bias and discrimination from this very first initial stage of setting up the systems. As it is people who make algorithms, they are also subject to the chance that bias and discrimination are instituted within the system when it is being made, whether willfully or by negligence, either consciously or unconsciously. Algorithms are never totally neutral, as they have been imbued with values, given that ontologies are not born out of nothing, but are formed by active, open and everyday practices of the world ontologies that they are destined to describe and classify. Algorithms may themselves be classified into policy-neutral and policy-directed, this (to make the matter even more complicated) being not a clear dichotomy, but a spectrum. Even policy-directed algorithms can work towards suppressing discrimination, which is a goal not usually connected to them, but to the policy-neutral ones. What is, therefore, paramount in this context, is both transparency of purposes, as well as the relevant accountability.

3. Conclusion

As may be clarified, human intervention does not always lessen the danger of discrimination and that technology can prevent bias, proposing not only privacy, but also fairness by design. This can be achieved through the application of the principle of justice when it comes to algorithms, which will prevent discrimination. We not only need human intervention, but also algorithmic neutrality, or 'correct' policy-directed algorithms, as with human intervention, unfair factors may inappropriately affect decisions. For example, an objective evaluation of candidates for a job tends to avoid any human intervention. This may show a 'lack of trust' in human intervention, as this was the cause of favoritism, especially in relation to positions in the public sector, which caused a lot of distortion and damage in the past.

A purely objective evaluation may not take into account human factors such as personality traits, which cannot normally be evaluated by technical means. Interviews are usually necessary, so human intervention remains important, as an actual person is certainly far more important than a mere collection of data related to him. Human intervention must not confirm the lack of trust, but lead to a deeper evaluation of all possible traits of human personality.

Most importantly, in order to better protect human rights, the application of the principles of transparency and accountability is paramount. In this case, transparency does not mean the satisfaction of the right to an explanation, or to information, depending on the proper construction of Art. 22, as set out above, by simply delivering the source code of the software used, as the data subject will very rarely be in a position to decode it: instead, it relates to the disclosure of the principles, policies and purposes that these algorithms serve, thus giving effect to the substance of the right to human intervention under Art. 22.

References

Alexandropoulou-Egyptiadou E. (2016). *Personal Data*, Nomiki Vivliothiki (in Greek).

Armamentos, P., Sotiropoulos, V. (2005). *Personal Data, Interpretation of L. 2472/1997*, Sakkoulas (in Greek).

Bosker, B. (2013, February 13). Google's Online Ad Results Guilty Of Racial Profiling, According To New Study. *The Huffington*.

Bottis, M. (2009). Civil Liability for the illegal processing of personal data, *Applications of Civil Law*, 7, 784-797

Bottis, M., Bouchayar, G. (2018). Personal Data v. Big Data: Control Lost, Discrimination Found, under publication with the *International Journal of Technology, Policy and Law*, 1-12.

- Bottis, M., Bouchayar G. (2018), Personal Data v. Big Data II: Challenges of Commodification, under publication with *the International Journal of Technology, Policy and Law*, 1-8.
- Bottis, M. (2004). *Information Law*, Nomiki Vivliothiki, 1-330 (in Greek)
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15(3), 209-227.
- Cavoukian, A. (2010). Privacy by Design: The Definitive Workshop. *Identity in the Information Society*, 3(2), 247-251.
- Christodoulou, K. (2013). *Data Protection Law*, Athens, Sakkoulas (in Greek).
- Christou, V. (2016). *The right to the protection from the processing of personal data, Foundations, Interpretation and Perspectives*, Sakkoulas (in Greek)
- Clarke, R. (2009). Privacy Impact Assessment: its origins and development. *Computer Law and Security*, 25(2), 123-135.
- Costa, L. (2016). *Virtuality and capabilities in a world of ambient intelligence*. Springer International Publishing, Switzerland.
- Crawford, K., & Schultz, J. (2014). Big data and due process: toward a framework to redress predictive privacy harm. *BCL Rev.*, 55, 93.
- Davies, S. (2016). The Data Protection Regulation: A Triumph of Pragmatism over Principle. *European Data Protection Law Review*, 2, 290.
- De Hert, P., & Papakonstantinou, V. (2016). The new General Data Protection Regulation: Still a Sound system for the protection of individuals?. *Computer Law and Security Review*, 32(2), 179-194.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. (2011). Fairness through awareness, <http://arxiv.org/pdf/1104.3913.pdf>
- Dynarski, S. (2016, April 6). Why Talented Black and Hispanic Students Can Go Undiscovered. *The New York Times*, <https://www.nytimes.com/2016/04/10/upshot/why-talented-black-and-hispanic-students-can-go-undiscovered.html>
- Finley, S. (2014). *Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods*. Palgrave Macmillan.

- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on information systems (TOIS)*, 14(3), 330-347.
- Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation', ICML, Workshop on Human Interpretability in Machine Learning, arXiv:1606.08813(v3). *AI Magazine*, 38(3), 50-57.
- Haber, E. (2016). Privatization of the Judiciary. *Seattle University Law Review*, 40, 115.
- Haber, E. (2015). The cyber civil war. *Hofstra Law Review*, 44, 41.
- Hoffman, A. L. (2016, July 16). Science will not save us: Medicine, research ethics and my transgender body, Autostraddle.
- Jernigan, C., & Mistree, B. F. (2009). Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10), <https://firstmonday.org/article/view/2611/2302>
- Kakar, V., Franco, J., Voelz, J., & Wu, J. (2016). Effects of Host Race Information on Airbnb Listing Prices in San Francisco. *MPRA* paper no. 69974.
- Citron, D. K. (2008). Technological Due Process, *Wash. UL Rev.*, 85.
- Knight W. (2017). The Dark Secret at the Heart of AI. *MIT Technology Review*, www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the debate. *Big Data and Society*, 3(2).
- Mitrou L. (2017). *The General Data Protection Regulation, New Law, New Obligations, New Rights*, Sakkoulas, 1-199 (in Greek).
- Mitrou L. (2013). The technological dimension of the protection of personal data. *DiMME*, 1/2013, 14-25 (in Greek).
- Nissenbaum H. (2001). How computer systems embody values. *IEEE Computer*, 34, 118-120.
- O' Neil, C. (2016). *Weapons of math destruction. How Big Data increases Inequality and Threatens Democracy*, Allen Lane.
- Panagopoulou-Koutnatzi F. (2017). New rights for citizens under the General Data Protection Regulation, a first estimate and constitutional evaluation, *Applications of Administrative Law*, 1/2017, 81-98 (in Greek).

- Panagopoulou-Koutnatzi F. (2017). *The General Data Protection Regulation 679/2016/EC, Introduction and protection of rights*, Sakkoulas, 1-135 (in Greek).
- Papaconstantinou E. (2006). *Law and Informatics*, Sakkoulas (in Greek).
- Pariser E. (2011). *The Filter Bubble, what the internet is hiding from you*, Penguin UK.
- Pasquale F. (2015). *The Black Box Society: the Secret Algorithms that Control Money and Information*, Harvard University Press.
- Schmer B.W. (2011). The limits of privacy in automated profiling and data mining, *Computer Law and Security Review*, 27(1), 45-52.
- Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, 7(4), 233-242.
- Schrage, M. (2014). Big data's dangerous new era of discrimination, *Harvard Business Review*.
- Shultz S., Art. 22, in: Peter Gola (ed.) *Datenschutz-Grundverordnung Kommentar*, CH Beck, Munchen 2017, 1.9.
- Tene, O., & Polonetsky, J. (2013). Judged by the Tin Man: Individual Rights in the age of Big Data, *Journal of Telecommunications and High Technology Law*, 11, 351.
- Tene, O. (2017). Taming the Golem: Challenges of Ethical Algorithmic Decision Making. *North Carolina Journal of Law and Technology*, 19, 125. Available at SSRN: <https://ssrn.com/abstract=2981466>
- Tene, O., & Polonetsky, J. (2013). A Theory of Creepy: Technology, Privacy and Shifting Social Norms. *Yale LJ and Tech.*, 16, 59.
- Tene, O., & Polonetsky, J. (2012). Big Data for all: Privacy and user control in the age of analytics. *NW JTech. and Intell. Prop*, 11, 239.
- The Economist (2012, March 31). Don't Hate Me Because I Am Beautiful, <http://www.economist.com/node/21551535>
- Thierer, A. (2012). The perils of classifying social media platforms as public utilities. *CommLaw Conspectus*, 21, 249.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76-99.

Zarsky, T. Z. (2016). Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review*, 47, 995.