Special Issue of Philosophy & Technology

Marcello D'Agostino – Massimo Durante

Introduction
The Governance of Algorithms

Academic affiliations
Marcello D'Agostino, Professor of Logic, Department of Philosophy, University of Milan, marcello.dagostino@unimi.it
Massimo Durante, Professor of Philosophy of Law and Legal Informatics, Department of Law, University of Turin, massimo.durante@unito.it

Abstract: In our information societies, tasks and decisions are increasingly outsourced to automated systems, machines and artificial agents that mediate human relationships, by taking decisions and acting on the basis of algorithms. This raises a critical issue: how are algorithmic procedures and applications to be appraised and governed? This question needs to be investigated, if one wishes to avoid the traps of ICTs ending up in isolating humans behind their screens and digital delegates, or harnessing them in a passive role, by curtailing their freedom and autonomy.

Keywords: algorithms; governance; delegation; standard; knowledge basis; discrimination; black box; rationality; technological normativity.

The development and deployment of digital technologies and services induce pervasive and radical changes in our lives and societies. The growth in the number of devices and artificial agents, their increased intelligence, autonomous behavior and connectivity are changing significantly living standards as well as altering traditional conceptions and ways of understanding the reality. Hence, beyond the benefits determined by technological innovations, there are also theoretical and practical challenges, and sometimes threats, that need to be addressed to ensure that technological innovation goes hand in hand with societal needs, beliefs and expectations.

Against this backdrop, social sciences and humanities involved in research about Information and Communication Technologies (ICTs) need to proactively address the manifold impact of the taking-up of novel, specific technologies that are based on the functioning of algorithms. In fact, we

increasingly delegate tasks and decisions to automated systems, machines and artificial agents that mediate human relationships, by taking decisions and acting on the basis of more or less sophisticated algorithms. Considered in a broad sense, algorithms are "encoded procedures for transforming input data into a desired output, based on specified calculations" (Gillespie 2013) or, to put it in more general terms, they are sets of rules for solving a problem in a finite number of steps.

In our information societies, characterized by cloud computing, the web of everything, autonomous computing, the internet of things, networked platforms, social networks and so forth, an ever-rising number of entities (devices, systems, machines, agents etc.) are governed by algorithmic procedures in their relentless collection, sharing, aggregation and production of data and corresponding decisional procedures. Social sciences and humanities are expected to deal with this urging "governance of algorithms", if we wish to avoid the traps of ICTs ending up in isolating humans behind their screens and digital delegates, or harnessing them in a passive role, by curtailing their freedom and autonomy.

The issue is even more pressing in light of the technological dependence of our information societies. As Luciano Floridi has properly remarked, "advanced information societies are more and more heavily dependent on ICTs for the normal lives and their growth" (Floridi 2014). From a technological standpoint, we live in societies where human progress and welfare are not "just related to, but mostly dependent on, the successful and efficient management of the life-cycle of information" (Floridi 2014). This raises further concern, since the map of our technological dependencies may also draw the map of our social fragility, i.e. the chart of the fragmentation and vulnerabilities that cut across and imperil our lives and the societal system.

There is actually a puzzling tension at the core of the governance of algorithms. "On the one hand, algorithms are invoked as powerful entities that govern, judge, sort, regulate, classify, influence, or otherwise discipline the world. On the other hand, algorithms are portrayed as strangely elusive and inscrutable, or in fact as virtually unstudiable" (Barocas, Hood, Ziewitz 2013). More plainly, algorithms claim to govern the reality and, at the same time, constitute a reality that needs to be governed. Algorithms are directed to solve problems that are not always detectable in their own relevance and timeliness, whose solutions cannot always be evaluated by clear and agreed standards. Furthermore, algorithms are meant to solve such problems through procedures that are not always visible and assessable in their own. As Frank Pasquale has argued and illustrated, there is a "knowledge problem" in "The Black Box Society", as to the openness, transparency and fairness of algorithmic procedures and applications (Pasquale 2015).

As algorithms play a crucial role in our information societies, there are therefore two further main aspects of this puzzling tension that need to be addressed. Let us refer to these aspects, respectively, in terms of:

1) A "delegation problem": how are algorithmic procedures and applications to be governed once we delegate to them the solution of problems?

Delegation involves trust and therefore risk that is always related to all fiduciary relationships (Luhmann 1979). However, we need to examine what trust may be when it is referred to technological devices and artificial agents guided by algorithmic procedures and applications. Is it still possible to talk of trust? Is trust passed onto engineers and programmers? On the ICTs environment as a whole? On our capability to detect, investigate and evaluate the functioning of algorithmic procedures and applications? What are the risks and threats involved in this process of delegation? Delegation of tasks and decisions to algorithmic procedures and applications on a systemic scale may freeze human relationships, curtail human freedom and autonomy, and exclude sections of society from the labour market and the role of decision-making. Moreover, Stephen Hawking, Elon Musk, Bill Gates and other A.I. experts have already warned us about the risks of delegating too much to the increased artificial intelligence of devices and agents: "The ethical dilemma of bestowing moral responsibilities on robots calls for rigorous safety and preventative measures that are fail-safe, or the threats are too significant to risk" (Hawking, Musk, Gates et al., 2015). From a moral and legal standpoint, the delegation of tasks and decisions to automated systems, devices and agents, may create a dangerous gap between the resolution to act and the legal and/or moral consequences of automated actions.

2) A "standard problem": by which standards or indicators are the relevance and the timeliness of problems as well as the efficiency and the fairness of solutions brought about by algorithmic procedures and applications to be measured?

As far as tasks and decisions are delegated to automated systems, devices and agents, guided by algorithmic procedures and applications, all these automated entities end up incorporating the norms or criteria that guide their actions and decisions. Once embedded in the automated functioning of systems, devices or agents, norms and criteria become less detectable and measurable, so that they are taken away from public scrutiny and appraisal. Automation and algorithmic procedures and applications promote technological normativity (i.e., the embodiment of norms in automated systems, devices and agents) and thus tend to curtail the sphere of social reflexivity (i.e., the public interpretation, discussion and evaluation of norms), which is crucial for the accurate understanding and application of norms. The impoverishment of the sphere of reflexivity may disable us to judge adequately the relevance and the timeliness of problems as well

as the efficiency and the fairness of solutions brought about by algorithmic procedures and applications.

Thus, the functioning of algorithmic procedures and applications is to be examined and discussed in light of the above-mentioned problematic aspects (i.e. the limits of delegation and the assessment of standards) that are likely to affect the different areas in which algorithms are largely implemented. Let us cast some more light on how much ubiquitous and profound are the impacts of algorithms on our lives and societal system.

Algorithms and their governance affect different research areas and hence need to be examined from a multi-disciplinary perspective that includes, in and beyond social sciences and humanities, the following main fields of investigation: law, ethics, knowledge representation, technology, economics, and social studies. Such a multi-disciplinary perspective may also include studies in computer science, epistemic and interpersonal trust, ethics and theory of information, ethics of algorithms and big data, business theory, robotics, and so forth. Let us raise some critical and challenging issues concerning the main fields of investigation.

From the standpoint of decision theory, artificial intelligence or robotic agency, and notably military strategy, the delegation of decisions to automated systems, devices or agents, may deeply affect important dimensions of the law. It does so in different ways: by introducing forms of technological normativity: legal norms are increasingly embodied by design in artificial systems, devices or agents; by eliminating the interface between the law's terms and its application: this curtail the public scrutiny reserved to judicial interpretation and review; by imperiling the rule of law in democratic societies: what legal automation gains in terms of efficacy may be lost in terms of legitimacy. This is also likely to raise critical issues and hard cases as to the assessment of legal responsibility for delegated decisions, tasks and actions. Furthermore, algorithms can also raise crucial issues in discrimination law as a result of statistical correlations that have predictive (yet not explicative) value, which tend to orientate individual or collective human decisions and behaviours.

From this perspective, technologies based on algorithmic procedures (notably, through big data analysis) more and more often infer personal information from aggregated data, thereby profiling human beings and anticipating their expectations, views and behaviours. This has important normative consequences, since profiling and anticipation are meant to mold and even discriminate human behaviours, by surreptitiously classifying them: e.g., many major auto insurers charge much higher rates to drivers with less education, even if with perfect driving records, based on statistical correlations between school rates and car incidents, which are inferred by big data automatically collected and analysed, although these factors bear no logical and direct relation to insurers risk. This statistical correlation applies to many fields: from marketing strategy to public or private policies; from reputation profiling to health insurances or the banking system. Needless to say, this

process of profiling and anticipation does not only affect the field of law but, first and foremost, that of ethics. Human agents tend in fact to develop adaptive strategies by conforming their behaviour to the expected output of the algorithmic procedures, with serious distortive effects.

The algorithmic rules of rationality may actually replace or displace "the self-critical judgement of reason" (Daston 2013), thus impoverishing or limiting the moral sphere of autonomy and reflexivity. The systematic implementation of algorithms also carries the risk of over-confidence in those choices suggested by calculations based on potentially false or approximate assumptions, but guiding the choices of individuals without them being completely aware of such preferences. In this perspective, there is the further risk of a process encouraging the systematic use of algorithmic solutions masking the complexity of socio-economic issues that require other types of interventions. Often, little or no room is left for a process of rational argumentation that should be able to challenge the results of the algorithmic procedures by putting in questions some of their hidden assumptions or by taking into account some neglected aspect of the problem under consideration; at the same time, it is widely recognized that scientific and social advances crucially depend on the possibility of such an open and free critical discussion.

Furthermore, understood as particular modes of agency or novel forms of social ordering, algorithms may challenge deep-rooted philosophical categories, while raising new ethical dilemmas, as for instance by incorporating moral values in the design of automated systems, devices and agents, by dislocating the traditional fora of moral responsibility, or by robbing us of our freedom and autonomy.

At present, algorithms play a crucial role in the overall process of mining, collection and aggregation of data, creating new methods and forms of expertise in business and scientific research but also raising crucial issues and problems concerning data analytics and/or the ethics of big data. Against this backdrop, few big players – Apple, Google, Facebook, Amazon, Twitter etc. – increasingly give us access to a greater and greater amount of information through their own algorithms, thus raising key questions of freedom of information, reputational profiling and epistemic trust.

These are only few examples of how pervasive and profound may be the impact of algorithms on our societies as well as on our lives. This impact requires us to face all the raised issues and questions from a multidisciplinary standpoint that includes law, politics, computer science, epistemology, the ethics of algorithms, the ethics and theory of information, economics and business theory, social studies and so forth.

No single special issue can at present compound and account for all the envisaged and listed issues and questions, precisely because of the all pervading dimension and in-built functioning of the algorithmic operations and applications in our information societies. In this perspective, the

papers selected for our special issue seek to focus on and deal with at least some of these main issues and questions from such an alleged multi-disciplinary standpoint. Let us briefly introduce them.

From a legal standpoint, which touches upon sensitive issues of policy regulation and democracy, Ugo Pagallo's paper focuses on concerns and legal challenges brought on by the increasing use of algorithms. Notably, he analyses a particular class of algorithms that augment or replace analysis and decision-making by humans in the legal field, i.e. data analytics and machine learning. Taking into account Jack Balkin's work on "the laws of an algorithmic society", his paper draws attention to the obligations of transparency, matters of due process, and accountability, while shedding light on some crucial differences between the US and EU law on the regulation of algorithmic operators, both public and private. In this regulatory framework, Pagallo shows us that, in the US, scholars debate whether and to what extent new duties and responsibilities of algorithmic operators as, e.g., information fiduciaries, have to amend the current framework of self-regulation and light government, whereas in EU law much of the new duties and responsibilities of algorithmic operators have been attributed to them, since they are considered data controllers. His paper efficaciously explains why and to what extent the normative challenges brought on by the balance between delegation of decisions to algorithms and non-delegation is an open and disruptive issue that will likely represent the main topic of debate over the next years of the algorithmic society.

Paul de Laat's paper addresses a significant and debated question in the field of decision-making procedures assisted by algorithms as developed by machine learning: i.e., whether transparency can contribute to restoring accountability for such systems. In this perspective, this paper complements Pagallo's legal investigation from an ethical standpoint. De Laat analyses arguments pro and contra transparency. Contrary to mainstream views, he believes transparency can be useful to restore accountability only up to a point. Therefore, he fully examines and sets the ground for the most relevant objections to transparency: the loss of privacy when data sets become public; the perverse effects of disclosure of the algorithms themselves; the potential loss of competitive edge; the limited gains in answerability due to the inherent opaqueness of algorithms. His paper suggests some alternative approaches to the understanding of algorithmic decisions, which require models of machine learning that should either be interpreted ex post or be interpretable by design ex ante.

Reuben Binns also deals with the want of accountability in a range of social contexts governed by algorithmic decision-making procedures. Against this social backdrop, accountable decision-makers must provide their decisions with justifications for their automated system's outputs. Binns wonders what kinds of broader principles we should expect such justifications to appeal to. In this perspective, he correctly identifies and emphasizes not only the private but also, and above all, the public dimension of the justification process. Thus, drawing from political philosophy, he suggests an account of algorithmic accountability in terms of the democratic ideal of 'public reason'. In his

paper, Binns compellingly argues that situating demands for algorithmic accountability within the proper and publicly shared justificatory framework may enable us to better articulate their purpose and assess the adequacy of efforts towards them, while setting up a fundamental public forum for discussing and critically examining the governance of algorithms.

Hykel Hosni and Angelo Vulpiani deal with the issue of the governance of algorithms from a key and critical standpoint, which is destined to receive increasing attention in the next future: i.e., the epistemological standpoint. The authors consider that the availability of unprecedented amounts of data and increasingly sophisticated algorithmic analytic techniques has led to a revival of the old inductivist view. The main purpose of Hosni and Vulpiani's paper is thus to assess critically the role of "big data" in reshaping the key aspects of forecasting and in particular the claim that bigger data leads to more accurate predictions. The authors provocatively argue that this is not generally the case by discussing the representative example of weather forecasts and suggesting that a clever and context-dependent compromise between modelling and quantitative analysis stands out as the best forecasting strategy.

Erik Dahl discusses the issue of the governance of algorithms from the standpoint of epistemic responsibility. He considers that today's web devices not only retrieve information from the web in response to our queries but they even purport to answer queries directly without requiring us to comb through search results in order to find the information we want. In this scenario, how does one know whether a web device is trustworthy? One way to know is to inspect its inner workings. However, Dahl is quite skeptical about this answer, since ordinary users cannot really inspect the inner workings of web devices because of their scale, complexity, and corporate secrecy. Hence, he suggests that Individual Understanding, Expert Testimony, testing through Experience, and Social Vetting are four viable methods of appraising black-boxed technology, without inspecting its inner workings. By deploying these methods, Dahl contends we can remain responsible inquirers while benefitting from today's epistemic resources on the web.

The technological standpoint is picked up by Erik Olsson, whose paper discusses a highly disputed and timely critical technical feature in the governance of algorithms: notably, the alleged wisdom-of-crowds justification for PageRank and for other similar inlink-based ranking algorithms. In his contribution, Olsson fully examines and discusses this sort of justification and argues that neither the influential preferential attachment model – elaborated by Barabási and Albert (1999) – nor the more recent model – that has been introduced by Masterton, Olsson and Angere (2016) – allows for a satisfactory wisdom-of-crowds justification of PageRank and of similar ranking algorithms. As a remedy, Olsson suggests that future work should explore "dual models" of linking on the web: i.e., models that are able to fruitfully combine the two above-mentioned approaches, by viewing links as being attracted to both popularity and importance.

The social impact of algorithmic decision-making is faced and widely examined in Lepri, Oliver, Letouzé, Vinck and Pentland's paper. Bruno Lepri et al. consider that algorithmic decision-making processes may lead to more objective and therefore potentially fairer decisions than those made by humans influenced by several factors and biases (notably, prejudice, fatigue, or even hunger). However, algorithmic decision-making may be criticized for its potential to enhance information and power asymmetry, discrimination, and opacity. The paper provides an overview of technical solutions to enhance fairness, accountability, and transparency. It also highlights the urgency to engage multi-disciplinary teams of researchers, practitioners, policy makers and citizens to co-develop, deploy and evaluate in the real-word algorithmic decision-making processes. Notably, the paper describes the Open Algorithms (OPAL) project as a step towards a world where data and algorithms are used as lenses and levers in support of democracy and development.

Finally, John Danaher examines the ethical impact of algorithmic delegation of tasks by means of the growing usage of personal AI assistants. Their employ is effectively a form of algorithmic outsourcing: getting a smart machine learning algorithm to do something in our behalf. The paper draws an initial ethical framework to analyze the main ethical objections to their usage. Some scholars claim that algorithmic outsourcing is dehumanizing, leads to cognitive degeneration, and robs us of our freedom and autonomy. Some others argue that it is problematic in those cases where its use may degrade important interpersonal virtues. The paper argues that there are no general, conclusive objections that should prevent the employ of personal AI assistants wholesale, although some objections remain contextually relevant and require redress or at least serious consideration prior to usage.

References

Barabási L. and Albert R. (1999), "Emergence of Scaling in Random Networks", in Science, Vol. 286, Issue 5439, pp. 509-512.

Barocas S., Hood S., and Ziewitz M. (2013), "Governing Algorithms: A Provocation Piece" (March 29, 2013), available at SSRN: https://ssrn.com/abstract=2245322.

Daston L.-J. (2013); "How Reason Became Rationality", Max Planck Institute for the History of Science, available at http://www.mpiwg-berlin.mpg.de/en/research/projects/DeptII_Daston_Reason/index_html.

Floridi L. (2014), The Fourth Revolution. How the Infosphere is Reshaping Human Reality, Oxford: Oxford University Press.

Gillespie T. (2013), "The Relevance of Algorithms", in Gillespie T., Boczkowski P., and Foot K. (eds.), Media Technologies: Essays on Communication, Materiality, and Society, Cambridge, MA: MIT Press, pp. 167-194.

Hawking S., Musk E., Gates B. et al. (2015), "Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter", available at: https://futureoflife.org/ai-open-letter.

Luhmann N. (1999), "Trust: a mechanism for the reduction of social complexity", in N. Luhmann (ed.), Trust and power: Two works, pp. 1–103, New York: Wiley.

Masterton G., Olsson E., and Angere S. (2016), "Linking as Voting: How the Condorcet Jury Theorem in Political Science is Relevant to Webometrics", in Scientometrics, Volume 106, Issue 3, pp. 945–966.

Pasquale F. (2015), The Black Box Society: The Secret Algorithms That Control Money and Information, Boston: Harvard University Press.